Ensuring the Future Construction of Electricity Generation Plants: The Challenge of Maintaining Reliability in New U.S. Wholesale Electricity Markets

*By A. Joseph Cavicchi and Andrew Kolesnikov**

Introduction

Several independently-operated, federally regulated, hourly wholesale electricity markets have been established in the U.S. during the last several years. Driven by the U.S. Federal Energy Regulatory Commission's (FERC) landmark 1996 regulatory order providing open access to the U.S. high voltage transmission network, various regions embraced the opportunity to form sophisticated, internet-based trading platforms that produce transparent hourly spot prices for wholesale electricity supplies. Concomitantly in most regions where these markets were introduced, significant investments in new, high-efficiency, low-emission electricity generators have occurred. These investments flooded the marketplace with excess supply of electricity generating capacity, quickly revealing weaknesses in the underlying market structures and resulting in documented under-compensation of generating capacity clearly required to maintain system reliability. The recognition that market modifications must be considered has resulted in numerous FERC proceedings focused on resolving the problem before a crisis ensues.

At the time restructuring was initiated it was understood that future investment was an important issue, but energy markets were expected to produce accurate price signals, and simply formulated capacity markets were expected to value facilities that were infrequently operated. Although much investment occurred at the onset of restructuring in many parts of the U.S., expectations associated with how the markets would function were not realized. This has become a pronounced problem during the current period of excess supply in many regions, but the time when more generation capacity will be required is rapidly approaching, driving the urgency to modify existing wholesale market structures.

Without delving into the myriad details associated with short-term wholesale electricity market design in the U.S., it is well understood that the combination of bid mitigation systems, designed to thwart the potential exercise of market power, and so-called reliability must-run contracts results in electricity market-clearing prices that undervalue electricity generation capacity in certain geographic regions. Usually these particular geographic areas are sub-regions of larger areas encompassing the operational footprint of a wholesale market.[1] It is within these sub-regions that the under-compensation, price signaling problem is most pronounced. Where we would expect the market system to reveal the value of generating capacity to investors, it does not, requiring the market operator to scramble to either support aged resources or acquire new resources in order to maintain system security and reliability. This observed approach to maintaining short-term system security, and ensuring long-term generating capacity adequacy, was not envisioned when these markets were put in place.[2]

At the same time energy prices have been suppressed, the initially constituted capacity markets[3] have been based on vertical demand curves that have proven to be a poor approach to pricing capacity. These initial market structures have been developed using the classic approach for defining a reliability standard: the amount of generation capacity available to the system should be adequate to ensure that only one major outage occurs every ten years.[4] Because there is limited ability for consumers to reduce demand in response to high prices (not to mention poor price revelation to consumers overall), the one-day-in-ten-year standard currently sets the establishment of generation capacity level throughout the U.S. regional electricity markets. Thus, capacity market minimum quantities have been established using this reliability standard. Simply stated, a generation quantity is set at some percentage above measured or forecasted peak demand (typically 12-18% above), and this amount is defined as the total amount of generation capacity required throughout a region to ensure reliable operation of the electricity system (resulting in the quantity which defines the vertical demand curve). System buyers responsible for serving consumers are required to purchase an amount of capacity based on peak obligations and face financial penalties if they do not purchase enough; generators either sell capacity bilaterally or receive revenues from auctions administered by system operators that ensure system buyers meet their obligations.

The vertical demand curve has been characterized as having two distinct undesirable characteristics. First, auction prices are volatile: whenever system capacity is above or below the set quantity, prices either shoot up to penalty levels, or decline to nearly zero. And second, when capacity is in, or near to being in, short supply, there can be opportunities for sellers to withhold supply and drive up prices. Moreover, the combination of total system excess supply and sub-regions where capacity is in short supply creates opportunities for buyers in some instances to realize preferential pricing by free-riding on the system.[5] Thus, suppressed energy pricing and unworkable capacity markets have resulted in observable inadequate remuneration for various generation facilities.

The resolution of these problems will not be simple. The market operator cannot force the construction of generating capacity when needed,[6] and buyers of generating capacity will employ all means possible of limiting expenditures for reliability, given its costs are not always easy to allocate equitably across system users. Moreover, generating capacity can often provide reliability and security services over fairly wide geographic regions, while consumers are in many instances represented by several utilities (load serving entities (LSE)) that are not subject to consistent regulatory frameworks, further complicating cost allocation issues. The

*Joseph Cavicchi and Andrew Kolesnikov are Vice President and Analyst, respectively, at Lexecon, an FTI Company, working in its Harvard Square office in Cambridge, MA. Cavicchi may be reached at jcavicchi@lexecon.com

[1] See footnoes at end of text.

existence and urgency surrounding implementing solutions to these problems cannot be underemphasized.

Theoretical Considerations for Ensuring that New and Existing Capacity Receive Remunerative Compensation

Electricity market pricing theory offers two possible methods of correcting the current pricing problems: value-of-lost-load (VOLL) pricing, or setting out minimum acceptable quantities (as described above). In a market modeled after the classic value-of-lost-load design, spot energy prices during times of tight supply are designed to mimic the consumer's marginal willingness to pay for electricity by allowing him to make the optimal trade-off between reliability and cost. As supply and demand edge closer together, prices spike, reflecting the high willingness to pay in order to avoid having to shed load. In reality, the absence of real-time metering precludes load from self-adjusting to the current prices; therefore, whenever power shortages are imminent, the market operator must artificially set the spot price to either an arbitrarily defined cap or an offer limitation, usually unrelated to the value of lost load. This value (often considerably less than VOLL estimates) can exceed average prices significantly and is a way of providing additional inframarginal rents to cover fixed costs, but has clearly been insufficient to compensate a generator that is typically marginal. Although, ignoring risk and market-power considerations, ideal VOLL pricing should induce a level of investment in generating capacity, which ensures a socially optimal level of reliability.[7]

However, serious flaws hamper a VOLL market design. First, since the market is not capable of determining the value of lost load by itself, VOLL must be set administratively. The difficulty of estimating the value of lost load leaves significant room for error, resulting in over- or suboptimal investment in capacity as well as either more violent or more frequent price spikes in the short run.[8] Second, setting spot price to VOLL levels whenever capacity drops below an amount necessary to ensure peak demand is satisfied produces a virtually vertical energy demand curve. Such market structure augments investors' risk premiums, which are, in turn, passed on to end users in the form of higher rates. Thus, VOLL pricing exposes consumers to unpredictable and costly price swings, making it a highly unattractive choice from a political standpoint. Third, since peaking units must rely on being paid the value of lost load during periods of shortage in order to recover their fixed costs, and since shortage hours are few and far between, fixed-cost recovery is highly uncertain. In addition, the number of shortage hours may fluctuate from year to year, depending on many random factors such as weather, availability of generating resources, and the status of the transmission network, which will cause under-recoveries in some years and over-recoveries in others. Such unpredictability with respect to cash flows will surely prompt investors to demand higher risk premiums, which will ultimately be passed down to consumers through higher prices. Finally, the inherent price volatility is further exacerbated by incentives to exercise market power. The lack of real-time metering prevents consumers from shedding load voluntarily whenever spot prices rise, rendering the short-run demand curve very inelastic. Therefore, as peak load approaches the level of operational installed capacity, generators have an increased incentive to withhold their resources and push the prices up even further. Together these shortcomings make VOLL pricing unattractive to regulators, and as we describe above, anything remotely resembling it has been eliminated due to concerns associated with the exercise of market power.

Thus, given the unattractiveness of VOLL pricing, an emphasis has been placed on setting an amount of generation quantity deemed sufficient to ensure reliability. By electing to set quantity, market designers and system operators then face the problem of how to ensure that the set quantity is available in the marketplace. As we describe above, the initial approach has been to use a vertical demand curve for capacity, as opposed to, say, a uniform price paid to all capacity, or instituting a system of individual payments made to certain generators required to maintain reliability in subregions. As the vertical demand curve for capacity has been unworkable, there has been a move underway to introduce an administrative downward-sloping demand curve to price generation capacity. Currently, this approach is in favor, although there is limited experience with the proposed market structure and considerable debate surrounding the potential success of the new approach.[9] When considered more generally, the problems associated with trying to create a regulated administrative market, such as this, have often been faced by policymakers.

Before examining in more detail the extant solutions being embraced to resolve the capacity payment problems, it is instructive to consider a theoretical paradigm developed to inform the process of deciding whether the control of price or quantity will create the most efficient outcome in those situations where an isolated economic variable (in this instance, reliability via capacity amount specification) needs to be regulated. A seminal work on this topic is Martin L. Weitzman's "Prices vs. Quantities."[10] The motivation of this work was the evaluation of the question of whether the control of pollution was better achieved by establishing pollution emission standards, or by setting pollution taxes. Over the past 30 years, we have seen the U.S. often elect the former approach, although it has not been a simple proposition to determine the most efficient method.[11] Thus, considering a framework within which the reliability assurance question can be considered is useful.

In the case outlined by Weitzman, he considers explicitly the difficult decision of determining whether quantity or prices should be used as planning instruments. He suggests a modeling framework wherein the decision is cast in the context of a trade-off between the social benefits and costs of one policy approach over another. He envisions a downward-sloping marginal benefit curve (analogous to the capacity demand curve) and an upward-sloping marginal cost curve (analogous to the capacity supply curve). He then proceeds to derive a so-called coefficient of comparative advantage that can be used to draw inferences on whether setting quantity or price is a better planning approach.[12] His results provide interesting insights applicable to the capacity-planning dilemma facing wholesale electricity market designers.

In particular, Weitzman shows that the slopes of the demand and supply curves will significantly affect the ability of the chosen policy instrument to perform efficiently. For example, he explains that, depending upon the magnitude and sign of the coefficient as determined by the slopes of the demand and supply curves, it is possible to establish whether price or quantity control will be a better planning approach. His primary findings tell us that when demand is steeply sloped (the benefit function is sharply curved), or the supply curve is nearly flat, it is better to control quantity. Conversely, when demand is elastic (the benefit function is near to being linear), the price control mode is relatively more attractive. As he explains, this is because the marginal social benefit is approximately constant in some range such that naming a price is more optimal, assuming limited cost uncertainty. Finally, if marginal costs are very steeply rising around the optimum—i.e., the supply curve is steep, as can be the case with fixed capacity—there is not much difference between controlling price and quantity. He suggests that in this situation, "non-economic" factors should play a prominent role in determining whether to control price or quantity. Generally, he finds that quantity control tends to be the less damaging approach to resolving this problem when facing uncertainty. Nonetheless, given that the capacity demand curves described herein are developed purely on the basis of expert opinion, the question of the appropriate shape certainly arises.[13]

Finally, when we examine the approaches taken to resolve this problem on a world-wide basis, we see that price is often the planning instrument of choice. For example, both Argentina and Colombia employ a fixed-capacity price paid to all capacity on the system that meets certain standards. Additionally, the U.K. has experimented with setting price, as opposed to quantity. Thus, although we limit our review herein to quantity-based planning standards where the shape of the demand curve is established administratively to achieve certain objectives, it may be the case that we are only beginning to develop an understanding of how to most efficiently approach the resolution of this problem.

Current U.S. Solutions to the Capacity Payment Dilemma: Locational Installed Capacity (LICAP) Markets

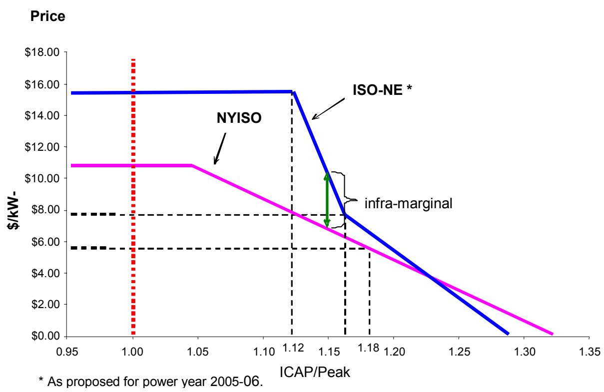
Given the problems that resulted when relying on capacity markets characterized by single vertical demand curves, there has been a significant effort placed on introducing price-quantity pairings—downward sloping demand curves—as a means of resolving the originally experienced problems. As we describe above, this is akin to making the policy decision to set quantity, and then proceeding to define the benefits function so as to achieve additional pricing objectives found to be desirable. Notwithstanding the limited experience that currently exists through the use of this approach, much effort has been expended by New York and New England to develop capacity demand curves that can be applied locationally (i.e., to sub-regions) as a means of setting prices based on desired quantities, and then stepping back and observing if investment is forthcoming in sufficient amounts to meet the desired reliability standard.

Generally, these newly constituted LICAP markets are designed to allow all generators, in particular peaking units, to recover their fixed (carrying) costs through the combination of energy-market rents and capacity payments. Additionally, they are formulated to place greater value on marginal capacity, leading to higher levels of reliability, which in turn would reduce the incidence of price spikes and lower the overall cost to consumers. Capacity prices are determined by the intersection of the short-run supply curve and the LICAP demand curve. In general, the LICAP demand curves are characterized by a flat high-price portion at low levels of installed capacity, designed to spur investment and bring installed capacity levels back up, and a downward sloping portion until prices hit zero at considerably higher levels of capacity, aimed at sending a retirement signal to the least efficient generators and reducing capacity down to the level of optimal reliability. LICAP demand curve designers define the price-quantity pairings for the curves such that an optimal level of investment in generating capacity occurs, while providing long-run prices that allow the marginal generator to recover its fixed costs.

The construction of a LICAP demand curve proceeds by first estimating two key inputs—the benchmark cost of capacity (BCC), previously called the cost of new entry, and the objective capability (OC) for the sub-region or zone in question. (OC is the amount of capacity necessary to meet forecasted demand.) BCC represents the annual fixed cost of the benchmark generator (either a frame or aero-derivative gas turbine peaking unit), which has the lowest fixed cost per megawatt of capacity and the highest variable costs. It is therefore typically the marginal generator that, during times of peak loads, sets the energy price in a market and earns the lowest infra-marginal rents. Thus, absent capacity markets, the benchmark generator systematically under-recovers its fixed costs and, assuming the decision rests solely with the generator, is driven out of the market. Under a LICAP market design, the price of capacity will hopefully hover close to the estimated BCC (EBCC) when the level of installed capacity provides adequate reliability, allowing the benchmark generator to recover its fixed costs and preventing it from exiting the market. Since any additional installed capacity would depress the LICAP price to below the EBCC, new generating units are discouraged from entry, thereby leading to an optimal long-run equilibrium level of installed capacity and a price equivalent to EBCC. OC, on the other hand, is the minimum acceptable level of installed capacity, the determination of which has been practiced by engineers for decades.[14] LICAP demand curves are defined by combining the appropriate EBCC and OC values, and then shaping the curve using expert opinion.

A LICAP market design using the demand curve described above has already been implemented in New York (by the New York ISO (NYISO)), while a similar LICAP market design proposed by ISO-NE is being reviewed by the FERC with the expectation that it will be put in place by January 1, 2006. At the same time, the PJM Interconnect is actively developing a reliability pricing model that also incorporates a capacity demand curve. Thus far, New York and New England provide excellent examples of two different approaches for drawing the capacity demand curve. Figure 1 portrays both New York's and New England's demand curves for power year 2005-06.[15]

Figure 1
LICAP Demand Curves, NYISO vs ISO-NE



The two LICAP demand curves display some similarities and some differences. In contrast to the pre-LICAP and VOLL designs' abrupt drop-off in the marginal value of capacity whenever it exceeded the required minimum, both New York's and New England's curves provide for a gradual decline in prices at above-optimal levels of LICAP, resulting in a less volatile and potentially more predictable stream of payments to generation owners, as well as hopefully more stable retail prices and sustained reliability. Both curves are designed to allow the marginal generators to recover their fixed costs, though the recovery mechanisms differ, and thus the NYISO's and ISO-NE's demand curves are not directly comparable. In New York, the price of capacity, as determined by the height of the demand curve at each particular value of LICAP, coincides with the actual capacity payment, and is calculated as the difference between the estimate of annual carrying costs of a new gas-fired combustion turbine and the estimate of the expected net revenues that a new combustion turbine would earn per year by selling into the energy and ancillary services markets.[16] While NYISO's demand curve determines the monthly capacity payments, ISO-NE's curve intersects supply at a conceptually different level. In New England, capacity payments are calculated as the difference between the LICAP price, as determined by the demand curve, and peak energy-market rents (PER),[17] and as proposed are distributed to eligible generators who made themselves available during shortage hours.[18] Because LICAP payments are reduced by price increases in the energy market, suppliers would lose as much in capacity payments as they would gain in energy rents should they choose to withhold their generation plants. Therefore, despite the fact that its demand curve incorporates energy rents, ISO-NE's proposal preserves the market-power mitigating characteristics of LICAP markets.[19] Moreover, it circumvents the difficulty and inevitable imprecision of estimating future infra-marginal rents by netting them out from the demand curve after the fact. This conceptual difference in the construction of NYISO's and ISO-NE's demand curves accounts for the significant price level discrepancy apparent upon initial comparison.

Table 1 presents a comparison of various aspects of the LICAP markets in operation in New York, and as proposed for New England. First, demand curve parameters vary as a function of LICAP zones. The definition of the zones is primarily based on system transmission limitations which require the ISO to use distinct operational guidelines to maintain reliability. Thus, the intention is that each LICAP zone be a geographic area where an incremental change in installed capacity would have a significantly different impact on reliability when compared to another area and, consequently, should be compensated differently. In New England, for instance, LICAP zones were initially designated according to the "currently-defined load zones in the NEPOOL Control Area."[20] As outlined in Table 1, ISO-NE has proposed five LICAP zones, compared to NYISO's three. There are suggestions that the loss-of-load-expectation (LOLE) should be the sole basis for the establishment of LICAP zones.[21] As such, separate zones should be created only if the installed capacity located there drops to levels insufficient to ensure a LOLE of one day in ten years, increasing capacity payments and thereby incenting new investments, and promptly eliminated as soon as new capacity brings reliability back to the required standard. However, this approach fails to recognize the fact that, in addition to transmission limits as well as other historical factors, new plant construction costs also tend to differ (in some instances significantly) among the currently established zones, a fact that is reflected by the zone-specific EBCC estimates.[22] Without accounting for these cost differences, there would constantly be an imbalance, as LICAP markets would always be over-compensating some generators and under-compensating others.

Table 1 also shows that another key difference between New York and New England LICAP markets lies within the market-clearing methodology. New York uses a nesting approach to clearing its markets. The NYISO administers monthly sequential locational installed capacity auctions, with the Long Island and New York City zones clearing an amount equal to locational sourcing requirements prior to running a larger regional market which then determines capacity prices for rest-of-state (the third New York zone) as well as that capacity that will be considered imported into New York City and Long Island. On the other hand, ISO-NE's proposed LICAP markets will clear all five zones simultaneously. This means that in ISO-NE, the amount of capacity that will be considered as imports into the various zones from rest-of-pool is determined by an optimization model and is limited to pre-defined capacity transfer limits between zones. Practically this means that in New England intra-regional supplies offering to deliver into constrained zones face a vertical supply curve, while in New York they face a sloping demand curve. Although these market clearing system approaches differ, and can result in different short-run prices, it is not expected that the revealed pricing should vary considerably over the long run.

Table 1
Key Characteristics of LICAP Markets in New York and New England

	New York[a]	New England
LICAP zones	NYC, Long Island, Rest of State	Maine, NEMA, SWCT, Rest-of-CT, Rest-of-Pool
Market-clearing methodology	Nested [b]	Simultaneous
Objective Capability (% above peak load)	18% [c]	12%
Break-even level of LICAP (% above OC)	0%	3.8% [d]
Zero-price level of LICAP (% above OC)	12% [e]	15%
Infra-marginal rent adjustment	Ex-ante	Ex-post

Note: LICAP and OC refer to locational installed capacity, and objective capability, respectively.
[a] The curves in NY are phased in, starting in 2003, in order to ameliorate rate impacts.
[b] NYISO administers sequential centralized monthly spot market auctions, whereby capacity in NYC and Long Island clears prior to Rest-of-State.
[c] Locational ICAP requirements in NYC and Long Island for power year 2004-05 are 80% and 99% of objective capability, respectively.
[d] The "target" level of ICAP in New England, the historical average level of capacity relative to OC, is numerically identical to the minimum requirement and break-even levels of ICAP in New York (1.054*1.12 is approximately 1.18).
[e] 18% in NYC and Long Island.

Although both New York and New England define objective capability as the level of installed capacity that ensures a LOLE of no higher than one day in ten years, the minimum required amount of capacity above forecast peak load differs between the two markets.[23] As Table 1 shows, the New York State Reliability Council (NYSRC) requires that installed capacity in the state exceed its peak load by 18%, while New England sets the region's minimum requirement at 12% above peak load. Since the installed capacity values on both demand curves were originally measured as a multiple of objective capability, the comparison of the two designs becomes more complicated. Therefore, in order to find a common denominator for the measure of capacity on the x-axis, the objective capabilities were converted back to peak loads in Figure 1. However, in order to preserve the curves' key parameters as they were originally defined by the ISOs, Table 1 lists them in terms of the minimum requirements.

Table 1 also shows that the New York and New England LICAP markets offer conceptually different levels of compensation to the owners of installed capacity. Whereas the marginal (benchmark) generator in New York breaks even whenever it brings the overall level of capacity to the required minimum, New England allows an additional 3.8 percent above objective capability. In effect, the break-even points in New York and New England then lie at 18 and 16 percent above peak load, respectively.[24] Similarly, the zero-price levels of LICAP, defined as 12 percent above OC in New York and falling at 15 percent above OC in New England,[25] correspond to 32 and 29 percent margins above peak load, respectively.[26] These differences can be seen in Figure 1.

Table 1 also shows the source of an obvious difference between the two curves plotted in Figure 1; the kink in ISO-NE's proposed demand curve. The kink occurs at the break-even level of installed capacity, as described above, and divides the downward-sloping portion of the curve into two segments. The left segment, by design, has a slope that is three times steeper than the slope of the right segment. Dr. Steven Stoft, who is responsible for the design of ISO-NE's proposed LICAP demand curve, argues that a steeper slope at close-to-deficient levels of installed capacity is necessary in order to send a stronger signal to investors and avoid shortages. Because "[t]he cost of too much installed capacity is considerably less than the cost of too little,"[27] the 3:1 slope ratio is justified. The value of capacity at the kink is calculated such that, assuming that the distribution of installed capacity levels maintains its historical standard deviation around the "target" level,[28] actual installed capacity falls below objective capability in only about 15 percent of years.[29] Thus, we see clearly how expert opinion leads to different demand curve parameters as well as differently shaped demand curves.

Lastly, as already discussed above, the price of locational installed capacity in New York, unlike in New England, has already been adjusted for infra-marginal rents. In general, as available capacity resources dwindle, energy and ancillary services' markets tighten, causing the prices to rise and, consequently, the rents that the generators earn by selling into these markets to increase. Conversely, as available capacity becomes more abundant, energy and ancillary services' markets loosen, leading to lower energy prices and lower infra-marginal rents. Recognizing this link between the energy and the capacity markets and the fact that generators must recover their carrying costs through a combination of rents from both markets, ISO-NE's demand curve is steeper than NYISO's at low levels of installed capacity and flatter at high levels. Thus, ISO-NE avoids the difficult estimation of infra-marginal rents by subtracting the actual rents from the LICAP price ex-post.[30] Thus, again expert opinion results in a significant difference in how capacity payments will account for inframarginal rents.

Conclusion

We clearly continue to be in a transitional mode characterized by a general lack of consensus on the appropriate policy choices to make to ensure reliability. Current policy on how to ensure future electricity system reliability in some regions of the U.S. is focused on establishing administrative LICAP market structures to value and hopefully cause, a predefined amount of generation capacity to be constructed. The ability of the new wholesale electricity markets in the Northeastern and Mid-Atlantic U.S. to signal the need for this next wave of generation investment is largely dependent on how well these new LICAP markets perform. Those investors that made past decisions based on expectations that markets would provide certain revenues will not be so willing to invest without assurances that capacity will be valued appropriately going forward. Although we expect that over the long run capacity will be compensated primarily via mid-term contracts, capacity markets will provide important signals as to the long-run marginal price of capacity. If we are to rely on administratively determined demand curves, we must be satisfied that they are shaped properly, and that there is true competition among those suppliers that offer capacity in the auctions. Currently it is clear that expert opinions differ on how to define and shape the demand curve in order to fulfill the reliability objective. As we can see from the curves, these differences will have an impact on capacity payments and thus expectations on the value of capacity in the future. In the near term it is imperative that LICAP markets send good price signals, as we cannot afford delays in needed future investments.

Footnotes

[1] For example, in New England the independent system operator (ISO-NE) has identified Southwest Connecticut as a problematic sub-region. In New York, both New York City and Long Island require separate consideration to ensure adequate capacity is available to meet demand. All these sub-regions are characterized by limited import capability, and in some instances are areas where siting new generation or transmission facilities is complicated both environmentally and technically.

[2] Although we understand that in some instances transmission system additions may resolve these observed problems, there nonetheless continues to be a fundamental problem with the current market structures when capacity shortages do not result in increased compensation to generating facility owners.

[3] At the time when independent system operators began administering wholesale electricity markets in the U.S., New York's, New England's, and Pennsylvania/New Jersey/Maryland's (PJM) ISOs each included capacity markets that were based on vertical demand curves. New York replaced its initial capacity market in 2003, New England is in the process of replacing its capacity market, and PJM is actively debating a so-called reliability pricing model to replace its capacity market.

[4] This refers to the bulk transmission and generation system as opposed to the lower voltage distribution system that will often experience weather induced outages.

[5] Initially constituted capacity markets had attributes that resulted in capacity being akin to a public good when it was in excess supply. Thus, consistent with the classic characteristic of a public good—non-exclusivity—buyers in all locations were able to take advantage of excess supply wherever it was located on the system. This problem was able to arise because generators have been required to offer their capacity in order to be eligible for capacity payments. Seriously limiting generators' ability to remove supply from the market led to capacity often resembling a public good.

[6] Of course, it is possible for the market operator to solicit supplies and make contractual obligations to buy such supplies, although taking a position in the market is completely contrary to the idea that market operators shall be independent and only provide a means for buyers and sellers to meet and transact anonymously.

[7] Stoft, Steven, *Power System Economics: Designing Markets for Electricity*, Piscataway, NJ: IEEE Press, 2002, at 159.

[8] The assumption is that regulators will eventually arrive at the correct level of VOLL by trial and error, spawning the desired level of investment in the long run.

[9] ISO-NE's proposed system is currently undergoing an extensive, more than year-long review at the FERC, while New York's system is still being reviewed by the court system to ensure that the FERC did not exceed its authority when approving the New York's new capacity market in 2003.

[10] Weitzman, Martin L., "Prices vs. Quantities," *The Review of Economic Studies*, Vol. XLI (4), No. 128, October 1974, at 477-491.

[11] For example, in Europe, pollution taxes are often favored over standards.

[12] Weitzman, *op. cit.*, at 85.

[13] It may be the case that using demand curves shaped based on consumers' willingness to forego consumption of electricity when facing high prices would be the most appropriate basis for the administrative curves currently in use. At a minimum this would allow thoughtful consideration of the resolution of the price versus quantity control question.

[14] As we mentioned earlier, minimum reliability is defined as the level of installed capacity which ensures a loss-of-load-expectation (LOLE) of 0.1, or one day in ten years, which is the second key input into the LICAP demand curve.

[15] The New England demand curve is that proposed by ISO-NE in its March 1 and August 31, 2004, filings for power year 2005-06. Nevertheless, considerable debate on the appropriate parameters for the curve is ongoing at the FERC.

[16] *United States of America, Before the Federal Energy Regulatory Commission, New York System Independent* [sic] *Operator, Inc., Docket No. ER03-647-000*, Affidavit of Dr. Thomas S. Paynter, at 20.

[17] PER are the "revenues, net of variable costs, that the Benchmark Generator would earn if it were always available." (*United States of America, Before the Federal Energy Regulatory Commission, Devon Power LLC, et al., Docket No. ER03-563-030*, Direct Testimony of Steven E. Stoft, at 20.)

[18] Capacity payments cannot be negative—i.e., if a generator's PERs exceed the LICAP price, no capacity payments are awarded to that generator.

[19] While generators would still be motivated to manipulate spot energy prices by withholding generation capacity, the concomitant reduction in capacity payments would result in market-power abuse having no effect on the overall rents. This may not hold for some low-variable-cost base load generators, whose increases in infra-marginal rents can, in theory, surpass the decreases in capacity payments as a consequence of withholding power. This is a case when infra-marginal rents exceed LICAP price and, since the associated capacity payment, calculated as the difference between LICAP price and infra-marginal rents, is always non-negative, the generator receives zero capacity payment.

[20] *United States of America, Before the Federal Energy Regulatory Commission, Devon Power LLC, et al., Docket No. ER03-563-030, Compliance Filing of ISO-NE, Inc.*, March 1, 2004, transmittal letter, at 5. Although, Southwest Connecticut was added as a zone following a FERC inquiry related to ISO-NE's March 1 filing.

[21] *United States of America, Before the Federal Regulatory Energy Commission, Devon Power LLC, et al., Docket No. ER03-563-030*, Prepared Direct Testimony of James G. Daly on Behalf of the Attorney General of Massachusetts, *et al*., at 42.

(continued on page 28)