IPU WORKING PAPER

MICHIGAN STATE UNIVERSITY ■ **INSTITUTE OF PUBLIC UTILITIES** REGULATORY RESEARCH AND EDUCATION

AN EXAMINATION OF RTO CAPACITY MARKETS

IPU WORKING PAPER NO. 2011-4

© KENNETH ROSE, PH.D.

SEPTEMBER 1, 2011 ▪ ipu.msu.edu

The Institute of Public Utilities at Michigan State University supports informed, effective, and efficient regulation of utility network industries providing electricity, natural gas, water, and telecommunications.

In keeping with this mission, the IPU Working Paper series is designed to disseminate relevant applied research and works in progress to the regulatory policy community.

Institute of Public Utilities Regulatory Research and Education
MICHIGAN STATE UNIVERSITY
W157 Owen Graduate Hall
East Lansing, MI 48835
517.355.1876 ■ ipu@msu.edu
ipu.msu.edu

Recommended citation
Rose, Kenneth. "An Examination of RTO Capacity Markets," IPU Working Paper No. 2011-4. Institute of Public Utilities, Michigan State University, September 1, 2011.

Acknowledgments
This paper was completed with funding from Citizens Against Rate Excess (CARE), Lansing, Michigan. The opinions expressed here are the author's and not necessarily those of CARE or any other person or organization.

About the Author
Dr. Kenneth Rose is an Independent Consultant and Senior Fellow at the Institute of Public Utilities, as well as a lecturer for the John Glenn School of Public Affairs at The Ohio State University. Dr. Rose is a nationally recognized expert with more than twenty-seven years of research experience in the structure, economics, and regulation of U.S. electricity markets. Dr. Rose's areas of expertise include electricity wholesale and retail market restructuring, market power, and market monitoring. Other research topics have included competitive bidding for power supply, regulatory treatment of uneconomic costs, Clean Air Act implementation, environmental externalities associated with electricity production, and other issues associated with the electricity industry and its regulation. Dr. Rose received his B.S. (1981), M.A. (1983), and Ph.D. (1988) in Economics from the University of Illinois at Chicago. Contact: ken@kenrose.us

MICHIGAN STATE
UNIVERSITY

AN EXAMINATION OF RTO CAPACITY MARKETS

KENNETH ROSE, PH.D.

SEPTEMBER 1, 2011

Introduction

Regional Transmission Organizations and Independent System Operators (both referred to here as RTOs[1]) have developed a mix of complex wholesale market mechanisms designed to simulate the operations of a competitive market. These mechanisms include energy markets (real-time and day-ahead markets), ancillary services markets (for example, frequency response service and spinning or synchronized reserve), and transmission congestion-based transmission rights. These market mechanisms operate within a complex framework of RTO operating rules and are overseen primarily by the Federal Energy Regulatory Commission (FERC) through FERC's rulemakings and regulations.

Three RTOs--PJM, ISO New England, and New York ISO--have also developed "locational" (that is, sub-regional) capacity markets intended to encourage building new capacity, retaining existing capacity, and permitting other resources, such as demand-response programs, also to participate in the market. The argument for creating a capacity market, in addition to the existing market mechanisms, is that markets for items such as energy and ancillary services do not provide sufficient revenues to recover the power suppliers' fixed costs. Also, for RTO system reliability a sufficient reserve margin is needed beyond what is necessary most hours of the year. A supplier that operates a "peaking" facility that runs only a small percentage of the year may not expect to recover its investment within a reasonable time to make the investment worthwhile. Capacity markets are intended to provide that revenue by creating an additional market mechanism.

This paper first looks at the economics of how a firm chooses a quantity to supply to a market and how that determines a regional market's supply curve. Next, it describes how the capacity markets work using PJM as an example. Finally, it examines how the economic theory and the RTO practice fit together within the context of recent FERC and capacity market developments.

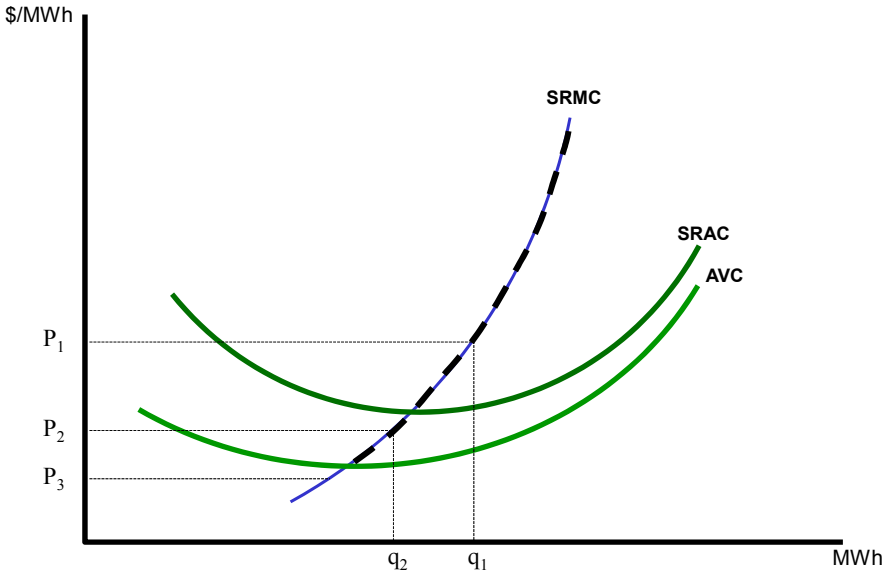
Economic Analysis of a Firm and Regional Supply

The argument for creating a separate capacity market is that energy market revenues are insufficient to induce an adequate amount of new capacity or to keep existing capacity from leaving the market. An additional capacity payment will, in this view, induce new entry and encourage existing facilities to remain or expand. This supposition is examined here using the tools of standard microeconomic analysis.

[1] The terms Regional Transmission Organizations (RTOs) and Independent System Operators (ISOs) have specific meaning and distinctions under Federal Energy Regulatory Commission (FERC) rules. For purposes of this paper, the term RTO is used generically to refer to an organization that operates an integrated wholesale electricity transmission system and an array of wholesale markets within the system.

Figure 1 shows the typical U-shaped short-run cost curves that illustrate a firm's decisionmaking on how much to produce given its cost structure.[2] The horizontal axis depicts units per output, or in this case, megawatthours (MWh). The vertical units are expressed in dollars per MWh. Average cost initially declines as the average per unit cost decreases as output increases. Average cost eventually increases at higher levels of output. The bottom curve is the average variable cost (AVC), or those costs that vary with output. The short-run average cost curve (SRAC) includes both variable and fixed costs; the difference between the two curves is the average fixed costs. The short-run marginal cost curve (SRMC) is the additional per unit total cost for each additional unit of output.

Figure 1. Firm short-run cost curves



The segment of the SRMC curve that slopes upward and lies above the AVC is the firm's short-run supply curve (depicted with a dashed line over the SRMC curve). This can be used to identify the quantity the firm is willing to supply for a given price. For example, at price P_1 the firm is able to earn a short-run profit; that is, revenue in excess of average costs. At P_2 the firm will incur a short-run loss, but continue to operate. This is because the firm can still recover all variable costs and at least some fixed costs. At P_3 the firm would be better off shutting down, since it cannot recover the variable cost from continuing operation.

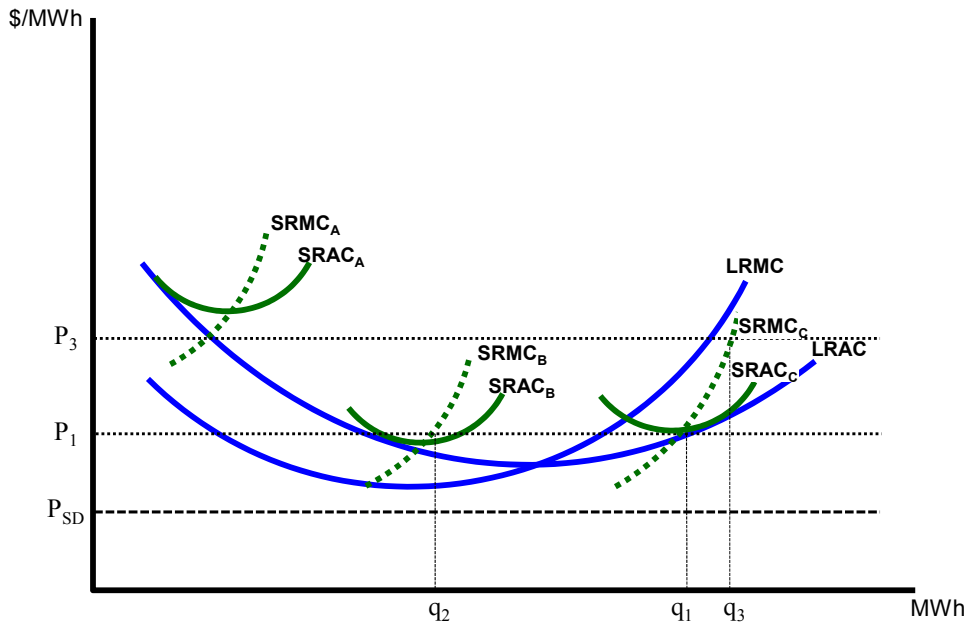
As the firm adds capacity (either capacity at existing plants or new units and entire plants), its potential output expands as well. Figure 2 shows the short-run cost curves for three capacity scales as a hypothetical firm expands capacity (A, B, and C).[3] The dashed lines are the short-run marginal cost curves

[2] The analysis used here is similar to what is used in standard intermediate microeconomic textbooks. This has a long pedigree in regulatory economics as well as it was used, for example, by Clemens, *Economics and Public Utilities* (1950) and Kahn, *The Economics of Regulation: Principles and Institutions* (1970).

[3] An actual firm would have many variations between these short run curves; only three are shown here for clarity of exposition.

for each level of capacity, and the short-run supply curves (again, the portion above AVC; in Figure 2 AVC is not shown). The short-run average cost curves trace the long-run average cost curve (LRAC), which represents the lowest average cost obtainable by the firm when all inputs are variable (since all costs, including capital costs, are variable in the long run).

Figure 2. Firm short-run and long-run cost curves



At price P_1, if a firm had expanded to capacity level "C" shown in the figure, the firm would choose to supply q_1 level of output. At this price and output level the firm would recover all its costs, both fixed and variable (since price is just equal to SRAC). If the firm had chosen a plant scale represented by the cost curves "B," output would be considerably lower, at q_2, but the firm would earn a modest (short-run) profit (since price is greater than SRAC). Note that at a price below P_{SD} the firm will not supply any output at any of the capacity scales shown.[4] Finally, at price P_3, the firm would earn a considerable profit at the capacity scale of "B" or "C," since this is considerably above average cost; at this price output would increase to q_3 at scale "C". However, if the firm was at capacity scale "A," the firm may produce no output at all; if it chose to it may recover only a portion of its fixed costs, depending on the location of AVC. Depending on the price and scale of operation, over the long run the same firm could vary output considerably and face considerable profit or loss.[5]

[4] There may be a capacity scale between "B" and "C" that would allow the firm to produce a positive output. In other words, the firm could expand capacity beyond "B" and move down the LRAC curve to where long-run and short-run average costs would be equal and is at minimum LRAC. Only three possible capacity scales for the firm are considered here out of many along the LRAC curve.

[5] The role of long-run marginal cost (LRMC in Figure 2) will be discussed later in this paper.

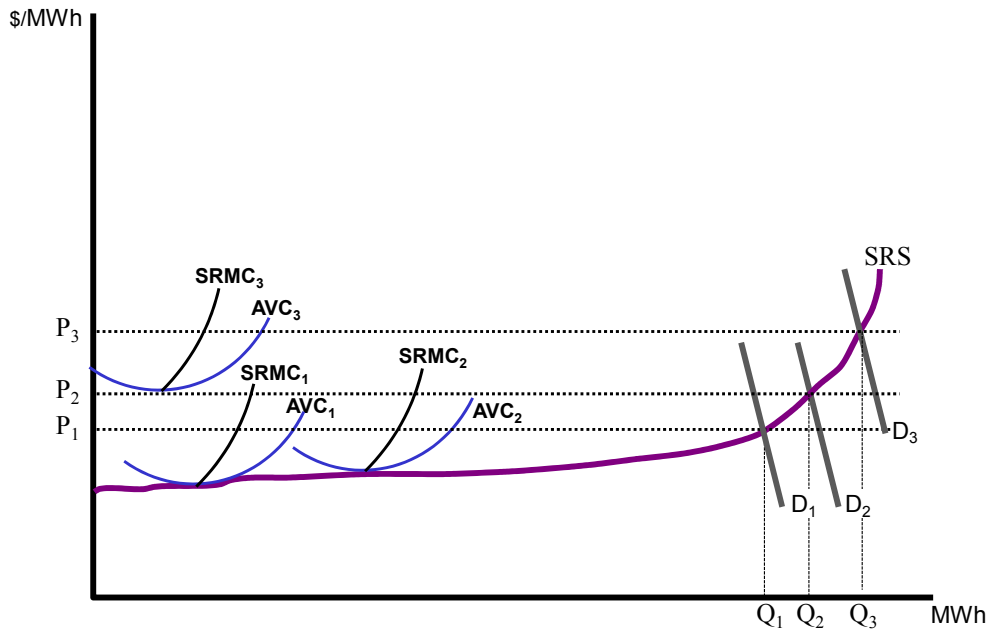
The downward-sloping portion of the firm's long-run average cost curve represents a well-known concept in electricity supply: the region of output where a firm has economies of scale—that is, where output can be increased by more than the proportional increase in total cost.[6] Where long-run average cost is increasing as output increases, there are diseconomies of scale. An output region may exist in-between where long-run average cost is flat, or where there are neither economies nor diseconomies of scale. Figure 2 shows the usual U-shaped average cost curve, but it is typically assumed that electricity production exhibits a wide output range where there are economies of scale and then flat long-run average cost for a range before average cost begins to increase (at least for very large systems). The implication is that one firm can supply a considerable amount of output (within some limited range of output) at a lower cost than two or more smaller firms.[7]

This analysis assumes the firm is a "price taker;" that is, the firm has no market power and cannot raise the price above a competitive level. The price is, therefore, determined by the interaction of many buyers and sellers in the market, of which this hypothetical firm is only one and too small to have a significant effect on the price. For this reason, the price level is shown as a horizontal line and also represents the firm's marginal revenue.[8] Given the conditions that exist in most RTOs, such as the high concentration of capacity ownership within RTO sub-regions, this is not a likely assumption. However, the primary task of this paper is to examine the argument that additional revenue is needed for capacity, beyond energy and other revenue sources.[9]

In a regional market with many sellers at different scales of operation, all the sellers together would determine the short-run supply curve. This would be the horizontal sum of all the firm SRMC curves (again, just the portion above AVC). Three AVC and SRMC (above AVC) curves are shown in Figure 3 for firms indicated with subscripts 1, 2, and 3 (the SRAC curves are not shown).[10] All the short-run supply curves for firms in the region taken together would determine the aggregate short-run supply curve (SRS). This can then be used to determine the price and quantity in the market for different levels of demand. Figure 3 illustrates this with an aggregate short-run supply curve and three different market demand curves. The price P_1 corresponds with the total quantity supplied for the region of Q_1, and P_2 and P_3 with Q_2 and Q_3, respectively.

[6] Another way of stating it is that this is a case where a doubling of the output results in a less than doubling of costs.
[7] If one firm can produce the *entire* output of a market at a lower cost than two or more firms, that firm is defined as a natural monopoly. This provides the foundation for cost-based regulation.
[8] A firm with market power would face downward-sloping *firm* demand and marginal revenue curves. The profit maximizing output level would still be where marginal cost equaled marginal revenue; however, the firm would be able to obtain a higher price based on its demand curve. The steeper the firm demand curve, the more the firm has pricing ability or market power.
[9] How market power in capacity markets is viewed by FERC is discussed later in this paper.
[10] In actual regional markets, there are many suppliers that form the short-run supply curve. For clarity, only three firms are shown to serve as examples.

Figure 3. Short-run firm cost curves and regional supply curve



In this scenario, no single firm is of sufficient scale to provide the entire region's output to meet demand and even the largest firms are relatively small compared to the output needed to meet regional demand.[11] At price P_1, both firms 1 and 2 would supply output (together providing about half the region's output with D_1 demand), but firm 3 would provide no output. For demand level D_2 with P_2 and Q_2, price and quantity, firms 1 and 2 again would supply output. But since the price is still slightly below firm 3's AVC, it still will supply no output. At price P_3, all three firms would supply output and, depending on where the firms' respective average cost curves are, all would likely earn a profit.

Applying this analysis to the electric supply industry requires several supply and demand distinctions to be made. For electricity, demand fluctuates considerably by season and by time of day, so the output levels needed to reach all three demand curves shown in Figure 3 *could occur within the same day*. In addition, to maintain reliable service demand must be met immediately ("on demand") by ramping generating units up or down as demand changes. How do these distinctions change the economics just described? In theory, it might be assumed that the firm simply adds plants when needed and shuts them down when demand falls; increasing and decreasing output by adjusting its capacity scale, effortlessly sliding along the long-run average cost curve. In practice, however, it is not easy to build new plants (and here the challenges of a "lumpy" capacity technology come into consideration). Depending on the type of facility, new plants generally require local permits, state regulatory approval, transmission and fuel access, and substantial financial capital. Even under the best circumstances, the process takes at least one and one-half to two years to complete.

[11] Although for some RTO sub-regions or transmission zones one supplier can provide nearly all the area's demand.

An electric utility serving a single service territory mostly with its own generation solves the problem of changing demand by having different kinds of generation resources on hand: base-load plants or units to run most hours because they have low operating costs (marginal costs) but cannot be ramped up and down quickly; smaller intermediate units that have higher operating costs but that can be up and running faster than base-load plants; and "peaking" units that can meet load quickly but have high operating costs and may run only during periods of high demand (for example, during the summer months for a summer-peaking utility). All of these units may be kept on hand year-round, even though some peaking units may not operate for months at a time. Customers typically pay a rate that is close to the average cost of the entire system, which includes all generating unit types.

An RTO combines (or at least attempts to combine) a number of utility systems into one regional system that faces the same problem of fluctuating demand, but on a broader scale and in a market context. Rather than relying on one utility to adjust its scale to meet demand, an RTO has a number of suppliers with a wide variety of operational scale. Suppliers use different generation technologies and fuels, and vary considerably by size (capacity of less than one megawatt to several thousand megawatts). A claimed benefit of an RTO system is that the RTO can dispatch from a wide array of supply and demand resources.[12] However, unlike a utility providing most of the electric supply for an area, an RTO may have suppliers of wider-ranging scale, including some that are relatively small in scale and that have high average and marginal costs. (One such firm is depicted by firm A in Figure 2.)

It is in the long run that firms change the scale of their operation since all inputs, including capital, are variable. Firms can be at different points on their long-run average cost curve, and long-run average cost is different across firms due to differences in technology and input costs. Typically, an average cost curve is drawn as a smooth continuous function (as shown in Figure 2). However, it is actually a dynamic function of changing technology, highly variable input costs such as fuel, and future investment decisions of the firm. For this reason, it is difficult to pinpoint what the price *should* be to encourage sufficient capacity expansion and retention to ensure reliability. As will be discussed, several RTOs have created separate capacity markets to provide additional revenues for suppliers. They have based the price of that capacity, in part, on an estimate of "new-entry" supply.

Several inferences can be drawn from this analysis. First, while there is a price below which a firm will shut down its operations and not produce any output, this depends on where the firm chooses to be in terms of capacity scale. It may be that by increasing capacity a firm may lower costs and provide an output. Second, smaller-scale operations may require a higher price, which may provide considerable economic profit (that is, a price greater that average cost) to other lower-cost suppliers. In the short-run, a firm does not include investment or capital cost when deciding how much output to produce (using the SRMC to decide), since it cannot quickly change its capital stock, but considers only whether it is able to recover its variable costs. In the long run, and looking at possible future capital decisions, a firm will compare expected average (total) cost with current average variable cost. If the prospective average total cost is lower than AVC, it makes sense to make the new investment; past investment costs, or sunk costs, should not be a decision factor. However, if the AVC is lower than expected average total cost, then

[12] The focus here is on supply resources. As will be discussed later, demand resources are often able to bid in capacity markets similar to supply resources. However, the intent of this is to reduce demand, often (but not always) during peak hours. This does not change the cost curves, since it is assumed that the firm will use the best available (most efficient) technology.

capacity investment may not make sense and the firm should stay with the current capacity. In other words, capital expenditures are considered only on a forward, not backward, basis.

RTO Capacity Markets in Practice

Several RTOs, including New York ISO, ISO New England, and PJM, have developed some type of capacity obligation and resource procuring mechanism. The names, designs, and terminologies are different but they share several basic elements, including:

1) an obligation on those responsible for serving end-use customers (load) to have sufficient capacity to reliability serve that load;
2) a methodology to determine a capacity reserve margin and future capacity needs for sub-regions within the RTO and for the entire RTO;
3) a process for soliciting qualified supply (and demand) resources to meet future capacity needs (for constructing an offer or supply curve);
4) some type of benchmark to judge the cost of new capacity;
5) a methodology or approach for creating a "demand curve"; and
6) a process (such as an auction) to select resources and determine a capacity "price."

PJM has a mechanism called the "Reliability Pricing Model" (RPM) which uses all these elements and will be described in more detail.[13] The RPM mechanism replaced the "Capacity Credit Market" (CCM) that required Load-Serving Entities (LSEs)[14] to own or acquire capacity resources equal to peak load served plus a reserve margin. Under CCM, LSEs could use their existing capacity, buy, or build new capacity; acquire capacity through bilateral arrangement; or use the CCM to meet the obligation. RPM replaced this approach with a capacity mechanism that is "locational" (sub-regional) and uses a three-year forward obligation for capacity. Auctions began in April 2007 with capacity prices determined by using an offer-based supply curve and a simulated downward-sloping demand curve.

PJM runs "Base Residual Auctions" (BRA) that procure forward capacity resource commitments for a delivery year three years in the future and three "Incremental Auctions" that may be held 20 months, 10 months, and 3 months before each delivery year. The first and third incremental auctions allow suppliers to procure replacement capacity for commitments they can no longer fulfill. The second incremental auction allows PJM to procure more capacity if the delivery year peak load forecast has increased since the base auction was conducted. Demand-side resources and new transmission projects can (and do) participate in the auctions.

LSEs must participate in the RPM for load served in a PJM control zone. LSEs pay a locational reliability charge equal to the daily "unforced capacity obligation"[15] in its zone multiplied by the final zonal

[13] PJM's approach is complex and will only be summarized here. This discussion is not intended to provide sufficient detail for market participants and others that require more exhaustive detail. PJM provides training and training materials and also has extensive user documentation available on its web site (www.pjm.com).
[14] A Load Serving Entity is basically any entity that serves end-use customers within PJM and has been given the authority or obligation to sell electricity to end-use customers. This includes local distribution (or utility) companies and can be a qualified end-use customer (such as a large industrial customer).
[15] Unforced Capacity is defined by PJM as the installed capacity rated at summer conditions that are not on average experiencing either a forced outage or forced derating.

capacity price. LSEs in PJM can "self-supply" resources to meet their capacity obligations by designating resources they own or purchase bilaterally (but resources must be offered in base auctions). The base auction provides an opportunity to purchase capacity requirements beyond self-supplied resources. The "Fixed Resource Requirement" (FRR) allows LSEs to meet a fixed capacity obligation. The market clearing price is paid to all resources committed in the auction and may be offset by performance-based penalties.

The RPM process and results can be summarized using a diagram such as the one shown in Figure 4. The numbers used are from the Base Residual Auction for the delivery year 2014/2015 that was held in May 2011.[16] The vertical axis is in dollars per megawatt-day and the horizontal axis is in megawatts. The upward-sloping curve represents supply offers from the resources and is drawn as a close approximation, not the actual bids. The downward-sloping line (in blue) is the "Variable Resource Requirement" (VRR) or the "downward-sloping demand curve." The VRR "curve" (actually constructed of four straight line segments) is found by connecting three points that are the intersecting points (on the horizontal axis) representing adjustments to the reliability requirement with points (on the vertical axis) representing adjustments to the net "cost of new entry" or net CONE (that is, the estimated cost of new entry—CONE—minus energy and ancillary service revenue offset[17]).

The top point on the VRR curve is the upper bound on the price or 150 percent of net CONE ($513.35) and the minimum target reserve margin, which is set 3 percent below the adjusted target reserve margin (140,755.8 MW). The middle point is the intersection of net CONE (set at $342.23 for this auction) and the adjusted target reserve margin plus 1 percent (145,901.4 MW). The lower point is the intersection of 20 percent of net CONE ($68.45) and the adjusted target reserve plus 5 percent (151,047.1 MW). The VRR curve depicts the PJM RTO for this particular auction. The intersecting points are recalculated for each auction and a VRR curve is found for the "locational deliverability areas" (LDA) that are part of the auction.[18] Note that while the VRR "curve" is often referred to as a "downward-sloping demand curve," this does not represent actual demand for capacity resources. Rather, this is an engineering construct designed to find a "clearing price" based on estimated resource needs, estimated new entry costs, and offer prices.[19]

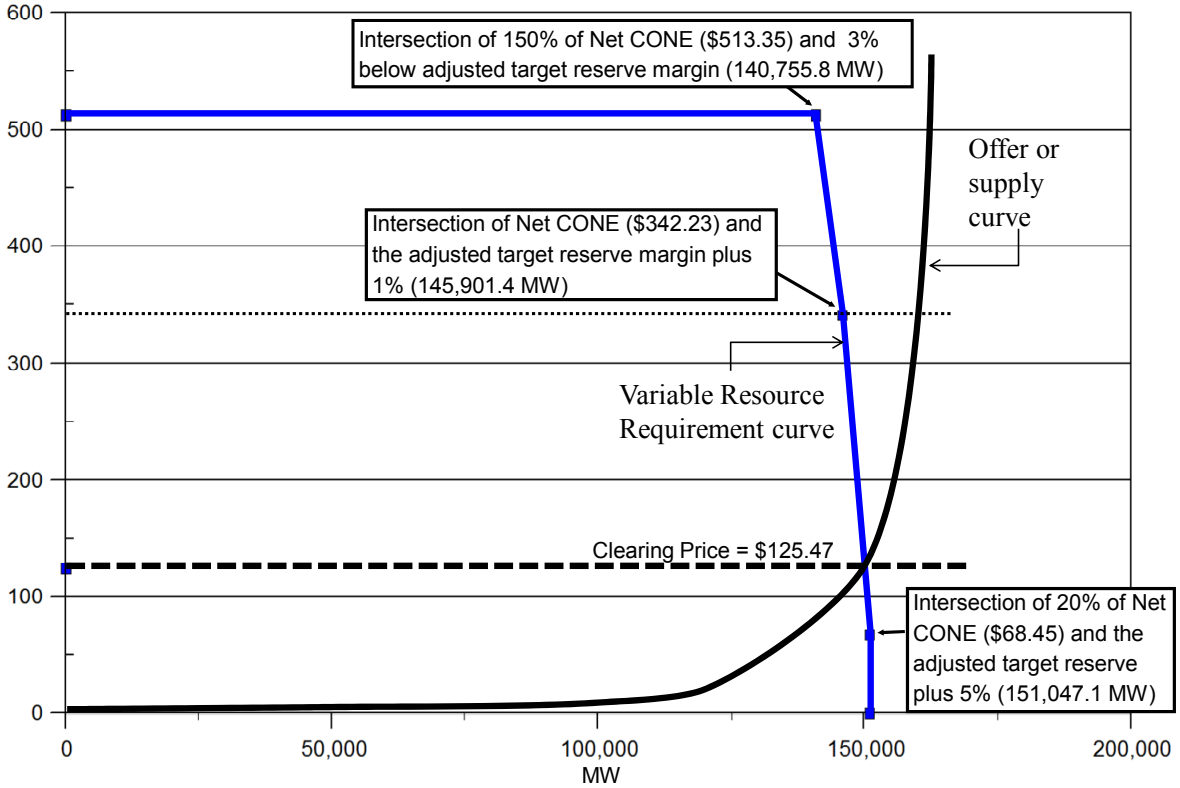
[16] The figure only shows, to present a clear example, the "limited DR [demand response] supply curve," and does not include the "annual resource and extended summer DR supply curve" that was also used for this auction. The resource clearing price shown of $125.47 was found by the intersection of the limited DR supply curve and the "VRR curve," shown in blue in Figure 4. The clearing price to meet the "minimum extended summer resource requirement" was $125.99 in the same auction.

[17] "Net CONE" is calculated by beginning with an estimated levelized revenue requirement (in $/MW-year; for the 2014/2015 BRA held in 2011, the previous year's estimate was used and adjusted using a regional Handy Whitman Index). The process then deducts the estimated annual energy revenue based on 2008-2010 zonal or average LMPs and the estimated annual ancillary services revenues. This annual number is then divided by 365 (arriving at a number that is in $/MW-day) and there is an "unforced capacity" adjustment (based on probable unit availability). Net CONE was calculated for five zones in eastern PJM and for the entire RTO for the 2011 BRA.

[18] Currently, PJM has identified 25 sub-regions as Locational Deliverability Areas (LDAs). Not all LDAs are included in each auction.

[19] This construct is based on a view that capacity can be separated from other input factors and from the final product (electricity). Why this is likely mistaken is discussed in the last section of this paper.

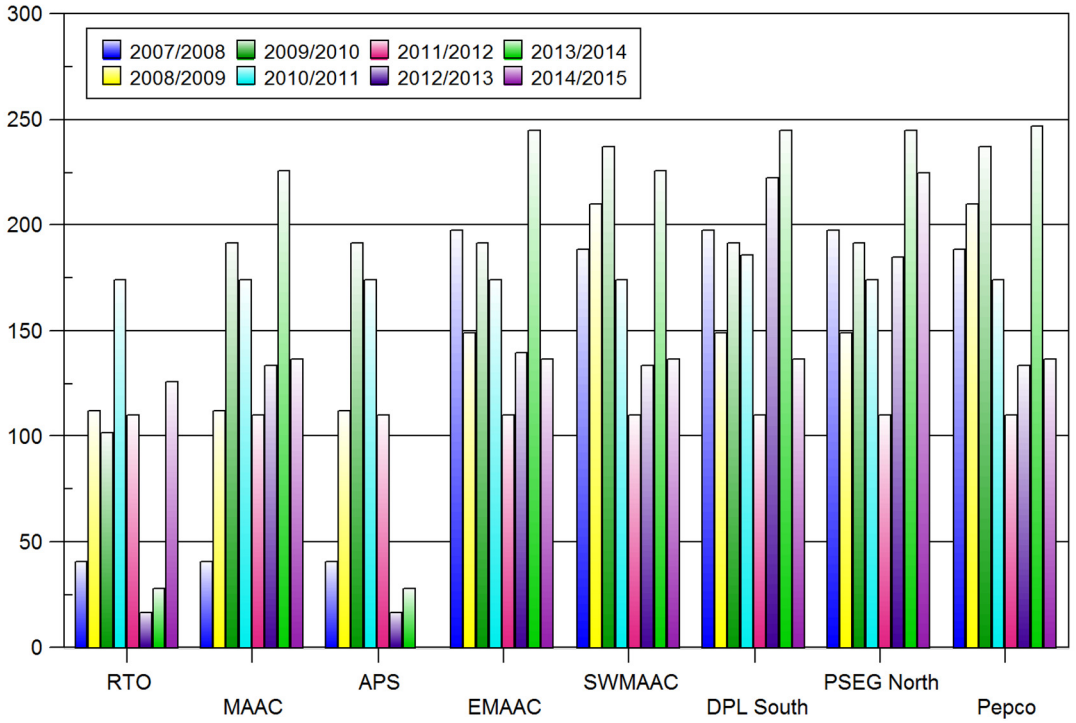
Figure 4. PJM 2014/2015 base residual auction ($/MW-day)



Source: Based on information from PJM, http://www.pjm.com/markets-and-operations/rpm/rpm-auction-user-info.aspx.

Figure 5 graphs BRA clearing prices for delivery years 2007/2008 through 2013/2014 (auctions held in 2007 through 2010).[20] The RTO price increased significantly until the 2010/2011 delivery year (the auction held in 2008), then dropped considerably until the 2011 auction shown in Figure 4 where the RTO price reached nearly $126. The drop in prices after 2008 is likely due to the recession's impact on the need for future capacity. Prices in the APS (Allegheny Power System) area followed a pattern similar to the RTO. However, the MAAC area (which encompasses most of Pennsylvania, New Jersey, Delaware, central and eastern Maryland, and the Delmarva Peninsula of Virginia[21]) has experienced significantly higher prices since the RPM mechanism began. For the 2014/2015 delivery-year auction held in 2011, the MAAC area capacity price fell to $136.50.[22] The higher prices in the eastern part of PJM are generally attributed to transmission constraints and higher capacity costs.

Figure 5. BRA clearing prices 2007/2008 through 2014/2015 ($/MW-day)



Data Source: 2010 State of the Market Report for PJM, Monitoring Analytics, LLC, 2011 and "Base Residual Auction Results," PJM, 2011.

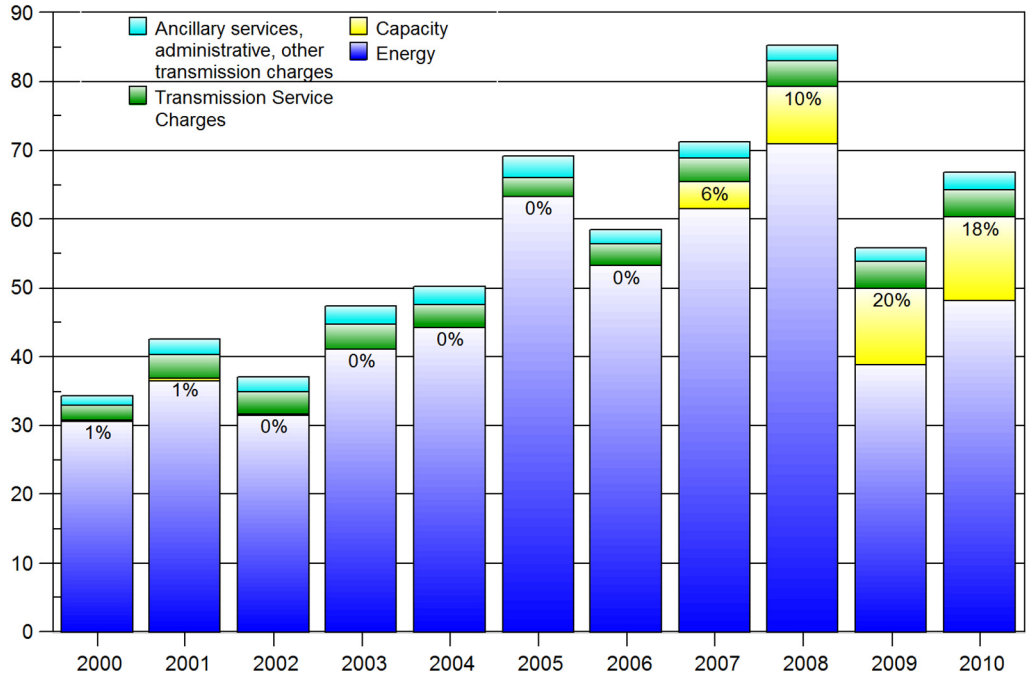
[20] A delivery year is June 1st through May 31st.

[21] This is the transmission zones of Baltimore Gas and Electric Co., Metropolitan Edison Co., Pennsylvania Electric Co., PPL Electric Utilities, Atlantic City Electric, Delmarva Power, Jersey Central Power and Light Co., PECO, Public Service Electric and Gas Co., and Rockland Electric Co.

[22] The capacity price the area pays is the specific area's price, otherwise the RTO price applies everywhere else.

The capacity prices are specified in dollars per MW-day and are difficult to translate into what customers pay for power. Figure 6 provides a wholesale price breakdown by category over the 2000 through 2010 period for PJM. Energy is the single largest component of the wholesale price, with transmission service charges a distant second through 2006. The capacity component, while insignificant before 2007, became the second largest category after the RPM mechanism began in 2007 (the percentage figures are for the capacity portion of the price). The capacity component grew to 20 percent and 18 percent of the wholesale price in 2009 and 2010, respectively. The fourth price component includes ancillary services, administrative, and other transmission charges and remained a relatively small share of the wholesale price.

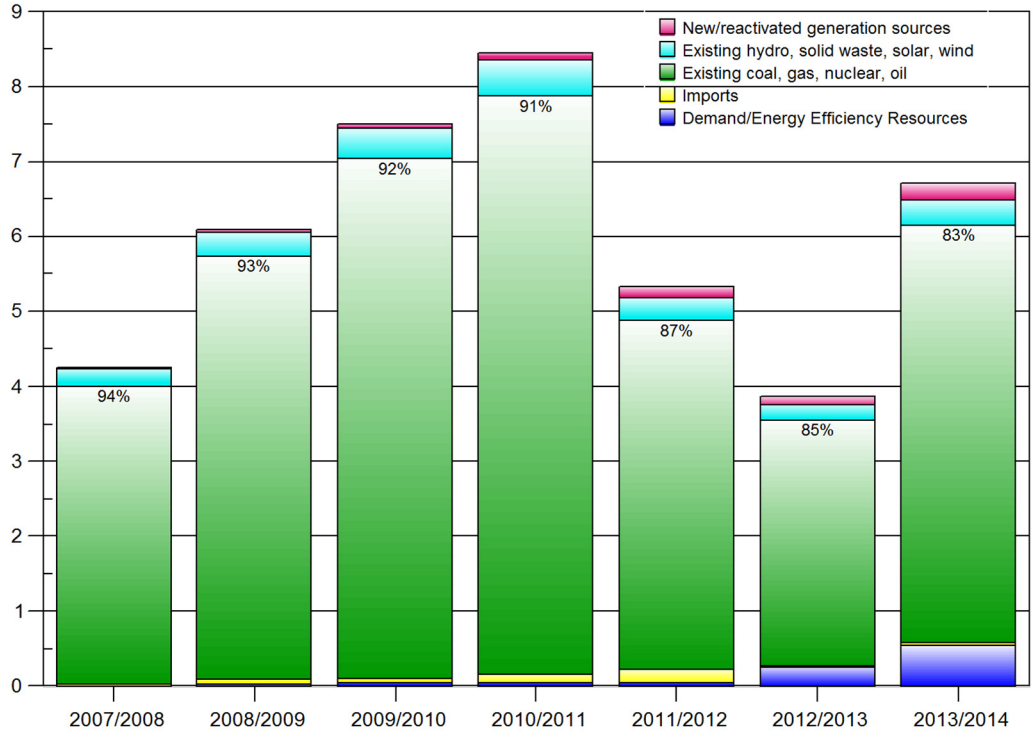
Figure 6. Total price per MWh by category, 2000 through 2010 ($/MWh)



Data Source: 2010 State of the Market Report for PJM, Monitoring Analytics, LLC, 2011.

Figure 7 shows trends for the revenue generated by the RPM mechanism by type of resource for each delivery year. Revenues have overwhelmingly been allocated to existing resources and, in particular, to existing coal, natural gas, nuclear, and oil generation. Taking into account all resources (which includes hydroelectric, solid waste, solar, and wind), 90 percent to 99 percent of revenues went to existing resources in the first six years of RPM. Only in the last delivery year did the percentage drop below 90 percent, to 88 percent in the 2010 auction (2013/2014 delivery year). New and reactivated generation sources received a small portion of the RPM revenues. Much is made also of the impact of demand management and energy-efficiency resources on the auction, but these resources, while significant and growing, are still (at least so far) a relatively small portion of annual revenues.

Figure 7. RPM revenue by type of resource ($ billions)



Data Source: 2010 State of the Market Report for PJM, Monitoring Analytics, LLC, 2011.

As part of the RPM, PJM also established a Minimum Offer Price Rule (MOPR) that, as FERC noted in a recent order, addresses the concern that "some market participants might have an incentive to depress market clearing prices [in the capacity market auctions] by offering *supply at less than a competitive level*" [footnote omitted, emphasis added].[23] PJM developed a "conduct screen" to use as a benchmark to determine if an offer is "uncompetitively low" and should be subject to mitigation. The conduct screen for combined-cycle (CC) and combustion-turbine (CT) technologies is 90 percent of the net CONE for both types of plants and 70 percent for unspecified plant technologies. Capacity resources that fail the conduct screen are re-priced at the same threshold levels; that is, 90 percent of the CC and CT asset class Net CONE, and 70 percent for other unspecified plant types.[24]

A number of resources are exempt from the MOPR and can offer a price below these thresholds (allowing zero-price offers), including nuclear, coal, integrated gasification combined cycle (IGCC), hydroelectric, wind, and solar.

FERC states that the MOPR apparatus is needed "to protect against *the exercise of buyer market power.*" FERC also requires that a sell offer failing the conduct screen must be subject to mitigation, reasoning that "*the uneconomic offer is increased to a competitive level*."[25] On the other side of the capacity market, the Independent Market Monitor (IMM) also monitors for *supplier* market power. The IMM "found serious market structure issues, measured by the three pivotal supplier test results, by market shares and by Herfindahl-Hirschman Index (HHI), but no exercise of market power in the PJM Capacity Market in the first three months of calendar-year 2011."[26] This is consistent with earlier IMM findings of serious structural problems but no exercise of market power.[27] The IMM found HHIs well above 3000 in five of seven RPM markets. Two markets were above 8000, considered extremely high concentration levels. These five areas all had market shares between 49 percent and 94 percent.[28]

Importantly, FERC only addressed "buyer market power" in its April 2011 MOPR Order, not seller or supplier market power. Of course, effective competition requires checks on both forms of potential market distortion. As will be discussed in the next section, a fundamental problem appears to exist with how the capacity market is defined that could lead to an erroneous conclusion that "buyer market power" is a problem.

---

[23] Federal Energy Regulatory Commission, 135 FERC ¶61,022, pp. 6-7, issued April 12, 2011. Two states, Maryland and New Jersey, concerned that they are paying unreasonably high prices for capacity through the RPM mechanism, have sought other means to bring additional capacity into the state. For further explanation, see *Public Power* magazine of the American Public Power Association, "A Tale of Two RTOs," vol. 69, no. 4, June 2011, at http://www.publicpower.org/Media/magazine/ArticleDetail.cfm?ItemNumber=32205.

[24] The information on the MOPR is from the April 2011 FERC MOPR Order, which authorized changes proposed by PJM. PJM has sought further clarification on the Order and FERC held a technical conference on July 28, 2011 (see http://www.ferc.gov/EventCalendar/Files/20110722160032-ER11-2875-001.pdf).

[25] These quotes are from paragraph 6 of the April 2011 FERC MOPR Order summarizing the 2006 RPM Settlement establishing the MOPR.

[26] 2011 Quarterly State of the Market Report for PJM: January through March, Monitoring Analytics, LLC, p. 134; www.monitoringanalytics.com.

[27] The IMM rarely finds market power being exercised in any market in the RTO. This is due to high thresholds being used to make a determination of when market power being exercised; that is, when comparing an estimated theoretical competitive market outcome with actual prices.

[28] Table 5-2 of the IMM Quarterly Report.

Analysis and Policy Implications

The rather elaborate construct or apparatus that PJM and other RTOs have created is primarily intended to maintain the reliable operation of the transmission system by paying suppliers an additional amount for capacity beyond what they receive in the energy and other markets.[29] However, there are at least two major conceptual problems with this approach that are at odds with the fundamental economics of electricity production.

Economic Efficiency

The first problem can be broken down into short-run and long-run cost considerations. The argument for creating a separate capacity market is that energy market and other revenues are insufficient to induce an adequate amount of new capacity and to keep existing capacity from leaving the market. Based on the analysis in the first section of this paper, this would mean the firm in the short run was operating below average variable cost (the "shutdown point") or was operating where the price was between average variable cost and short-run average cost (in other words, operating at a loss, but able to continue operating and recovering some fixed costs). If this situation were to occur, the additional capacity payment in theory would be equal to the difference between the firm's short-run average cost and the price. This would, again in theory, induce new entry and encourage existing facilities to remain or expand. However, the capacity mechanisms clearly do not attempt to do this. Rather, they rely on an artificial demand curve (based on cost of new entry, or CONE, and reserve requirements as described above) and the offers made by the suppliers to determine a "market clearing price."

The estimate of CONE is not an estimate of short-run marginal cost, but represents the cost of a new stand-alone CC or CT unit (in PJM's case) minus expected revenues from energy and ancillary services sales. This is not an actual firm's cost (average or marginal), and certainly is not a competitive price that would result from the interaction of firms with robust competition. FERC states that "Net CONE serves as a reasonable estimate for a competitive offer price,"[30] but is merely, at best, a short-term static estimate of a new firm's stand-alone entry cost.

As also shown in the first section, firm costs can vary considerably depending on the firm's operating scale. For example, at price P_3 in Figure 2 the firm operating at either scale B or C would earn a considerable profit. However, the firm operating at the scale shown by the figure's cost curves "A" would earn no profit at all and would consider shutting down, depending on expected future prices.[31] In other words, the same price can have very different outcomes for firms operating at different scales of operation. In reality, actual costs are dynamic and idiosyncratic to a particular firm's situation, and cannot be estimated by policymakers in advance and with a reasonable degree of precision.

In the long run, where all costs are variable, the firm in theory would choose an output where the price is equal to long-run marginal cost (LRMC in Figure 2, above).[32] In the output range where long-run

[29] These schemes are part of the RTOs' resource adequacy requirement for RTO planning and expansion, one of several FERC-mandated RTO functions.
[30] FERC MOPR Order, paragraph 43.
[31] Note that the firm could expand from that point and reduce cost.
[32] In theory, this is a common assumption. However, in practice it is difficult or impossible to pin down what the long-run marginal cost is, as noted some time ago by William Vickrey: "Another objection to the policy of setting price at marginal cost is based on the fact that reality does not conform to the regular and perfectly defined curves of the theorist, so that, in practice, marginal cost

average cost is increasing, the firm would earn a profit, since LRMC is greater than LRAC. While a firm would prefer to operate in the long run where price equals its long-run marginal costs, in an RTO market where other firms can expand or enter, the market price would be determined by a long-run supply curve, not the individual firm's or group of firms' long-run marginal cost.[33] The long-run supply curve would be determined by the interaction of all suppliers entering, expanding, or exiting the market, and by the changing conditions that affect the firms' costs, such as input costs and technology changes.

It is difficult to generalize about the shape of the industry's long-run supply curve without substantial further analysis. However, it is interesting to note that given the considerable expansion in U.S. output over the last half-century, real consumer prices for electricity (adjusted for overall price changes using gross domestic product implicit price deflator) have remained nearly constant. From 1960 to 2009, electricity generation increased by 420 percent, while the real price of electricity for all consumers fell by a comparatively small 7 percent.[34] This would imply a relatively flat long-run supply curve for the industry. Several important factors, of course, must be taken into account before any conclusion can be drawn about the supply curve's character or slope.

First, that 49-year time period between 1960 and 2009 saw considerable price variation, with a 59 percent difference between the lowest and highest real price.[35] Fluctuating fuel prices[36] and capital costs explain much of the variation. Also, the U.S. average price trend masks the substantial variation in price by region and state. Second, during this time period, the mix of fuels used for electric generation changed, including a drop in the use of fuel oil in the 1970s, a rise in the use of nuclear power in the 1970s and early 1980s, and increased use of natural gas beginning in the late 1990s and continuing to the present. Third, the increased use of natural gas reflects technology changes that reduced the efficient operating scale for power plants. The fuel costs and technological changes that occurred during the last half century served to effectively shift the long-run supply curve up (for example, from higher fuel prices) and down (for example, from technological improvements).

A flat long-run supply curve implies a constant-cost industry; that is, one where input prices remain the same as output expands.[37] Volatile fuel prices and technological changes would have the effect of shifting the curve up or down, but the long-run curve would be relatively flat. As noted, an analysis beyond the scope of this paper would be required to sort out the dynamic market changes that occurred during this period and to reach a conclusion about the nature of the long-run supply curve. If it is assumed that, at least for recent decades, the electricity industry has (or at least had) a flat long-run supply curve, these dynamics would suggest that as demand increased, new entrants and expanding output from existing firms

---

may be not only quite difficult to determine with any approach to accuracy but also subject to extreme and erratic fluctuation, depending on the precise circumstances of the moment." William Vickrey, "Objections to Marginal-Cost Pricing," *The Journal of Political Economy*, Vol. 56, No. 3, (Jun., 1948), pp. 218-238.

[33] The industry long-run supply curve is not the sum of the long-run marginal cost curves of firms in the industry, as is the case with the short-run marginal cost curves as discussed in the first section (and shown in Figure 3).

[34] Based on U.S. Department of Energy, Energy Information Administration data, Table 8.2a "Electricity Net Generation" and Table 8.10 "Average Retail Prices of Electricity."

[35] The real price of electricity for all customers fell or was flat from 1960 until 1973, then increased considerably until 1982, fell again until 2000, and has been rising again since 2004.

[36] The cost of fuel used to generate electricity is discussed in "The Impact of Fuel Costs on Electric Power Prices," June 2007, prepared for the American Public Power Association (APPA). Posted at: http://www.appanet.org/files/PDFs/ImpactofFuelCostsonElectricPowerPrices.pdf

[37] In other words, the additional inputs of fuel, capital, and other inputs needed to produce higher output are obtained without an increase in the per-unit price of the inputs.

would drive the price back down to the long-run equilibrium price. It would do this either through regulatory or market processes. Fluctuating fuel costs would shift the curve both up and down,[38] but new technologies would most likely have shifted the curve down. The long-run average cost curve would also be relatively flat. This implies that creating additional compensation for capital based on short-run new-entrant costs will likely overcompensate suppliers with lower long-run average costs beyond what is necessary for them remain or expand in the market.[39] This would amount to a subsidy that has the effect of supporting higher-cost suppliers, while providing unnecessary compensation to lower-cost ones.

In summary, this analysis suggests that in the short run, paying separately for capacity would potentially help some small-scale new entrants (if the payment was high enough), but only overcompensate and add to the economic profit of many (perhaps most) existing suppliers. In the long run, if the market for electricity is operating efficiently, then the capacity payment is unnecessary since all costs, including capital costs, would already be recovered in the market price for electricity since that price would be at or above long-run average cost.

Inputs v. Outputs

The second conceptual problem with the RTO capacity market approach is more fundamental from an economic theory standpoint. This approach attempts to construct a "market" to sell capacity to customers as if it were a final product that can be separated from other products that firms produce. This form of unbundling or separation may work for some outputs or products, such as byproducts or externalities where a separate market can be set up by regulators[40] or the ancillary services market that RTOs have created with FERC approval. The capital investment that capacity markets intend to induce is different; it is actually an *input* in the production process of the firm, not an *output*. The cost curves described above are a function of the inputs firms use to produce output: capital, land, labor, energy, materials, and so on. Each of these inputs has separate "factor" markets in which electricity suppliers operate. But these markets are not what the RTO capacity markets are trying to supplement or correct. Rather, they are attempting to create a final product market for something that is merely one input of many that are needed to generate electricity.[41]

This may explain why the capacity construct that the RTOs are using has become so complex. Every aspect of the capacity market design has to be redesigned and readjusted to fit changing conditions, rather than allowing the market participants to adjust to market information over time, as happens generally in competitive markets. The "demand curves" are artificial constructs based on estimated required capacity and cost of new capacity. There is no actual "demand" from end-use consumers, since capacity alone is of little use to them; they require delivery in the form of usable electrical energy. Suppliers that make the offers that constitute the supply or offer curve know how other suppliers will behave and will bid accordingly. These are not the ingredients of a healthy competitive market. PJM's MOPR is a case in point. The MOPR requires significant intervention in the market created by PJM because it actually does

[38] The constant-cost assumption here means that input prices remain constant as output expands, not that there are no exogenous price increases or decreases for the inputs, such as fuel. As noted, fuel price changes shift the curve.
[39] This would be the case even if it was assumed that there was an increasing input costs, or an upward-sloping long-run supply curve; only the price would not decline back to the original equilibrium price, but at a price higher than the original and below the short-run price.
[40] For example, the sulfur dioxide allowance trading market created by the Clean Air Act Amendments of 1990 and administered by the Environmental Protection Agency.
[41] It might be fair to ask: why not have a market construct for every input used, not just one for capacity?

not operate as a self-sustaining competitive market. This shift of focus away from the primary problem of how to secure capacity for reliability purposes may explain why "buyer market power" is regarded as a "market" problem by the RTOs and FERC.

The complex mechanism of capacity markets is not self-sustaining, since the RTOs and regulators will need to continuously update and fix the apparatus as conditions change. But no adjustment can fix the underlying conceptual problem of trying to artificially create a market for a production input, as compared to a final product. A truly competitive market, in contrast, changes as circumstances change, without the stakeholders having to agree on changes and without the regulator having to insert its judgment by choosing and approving what it thinks will work.

Conclusion

Proponents have argued that the capacity markets construct works and that it has helped attract new entry and retain existing capacity.[42] There is little doubt that additional revenues for suppliers would increase and retain capacity—in effect these revenues raise the horizontal price line (or marginal revenue) from where it would be without the capacity revenue (as shown in Figures 1 though 3).[43] The consequence, as Figure 7 shows, is that the overwhelming amount of RPM revenue is going to existing resources and little is going to new or reactivated resources. This likely means the total price received from all sources of revenue exceed average cost. This means economic profit—perhaps substantial economic profit—for existing suppliers. This amounts to a subsidy or transfer payment to suppliers over and above what is necessary to earn a normal profit and to continue operating or expanding. Distorted incentives will tend to perpetuate supply-side inefficiencies in resource allocation and operations. This uneconomic subsidy is non-trivial for customers. In the PJM case, it amounts to almost $800 for each person who lived in the service area since RPM began in 2007 (and through 2010).[44] Given recent economic conditions, the region can ill-afford such generous subsidies.

The essential point of federal regulatory policy in the electricity sector is to ensure the operation of a reliable transmission and supply system that is responsive to changing market conditions and public policies. Policymakers may need to refocus their efforts on how to accomplish these goals in a sustainable way that is consistent with fundamental economic principles. The current approach falls short in these basic respects.

---

[42] In a press release, PJM stated "PJM's analysis shows that, since the first auction in 2007, the RPM has retained and attracted 40,787 megawatts (MW) of power capacity resources compared to what would have happened without the RPM." From news release "Demand Resources and Energy Efficiency Continue to Grow in PJM's RPM Auction," May 13, 2011. Obviously, determining how much was retained due to RPM would be challenging.

[43] Figures 1 through 3 show costs as a function of electric energy produced or MWh; capacity payments are made in terms of capacity or MW. For this reason the marginal revenue curve would increase as output and capacity expanded in a nonlinear manner.

[44] According to PJM's independent market monitor, the total RPM revenue from 2007 through 2010 was $42,196,737,603, or just over $42 billion. PJM states that 54 million people live in their region. That works out to about $781 per person over the four-year period.

[blank page]