McCullough Research

Analysis of the Balancing Energy Market

February 20, 2009

6123 Reed College Place • Portland • Oregon • 97202 • 503-777-4616 • Robert@MResearch.com

Executive Summary

Balancing energy is the energy used to keep the system in balance. In a perfect world the need for balancing energy is slight – only occurring when some unusual event occurs. In recent years ISO/RTOs from California to New England have established highly structured balancing energy markets with complex bidding rules and difficult to follow (and often secret) computations to produce prices.[1]

Balancing energy markets frequently become the benchmark markets – setting prices in the larger markets at their very high levels.

Strange bidding often occurs in these markets. The term "hockey stick bids" has been coined to describe the non-economic bidding never before seen elsewhere in our economy. These hockey stick bids start reasonably enough, but soon transition to prices a hundred times higher than common sense would expect.

This white paper attempts to explain why these bids are so frequent as well as the economics behind their submission to the nation's ISO/RTOs.

[1] ISO stands for "Independent System Operator." The ISO (also called a Regional Transmission Organization) is responsible for transmission system operations in a given geographic area.

Contents

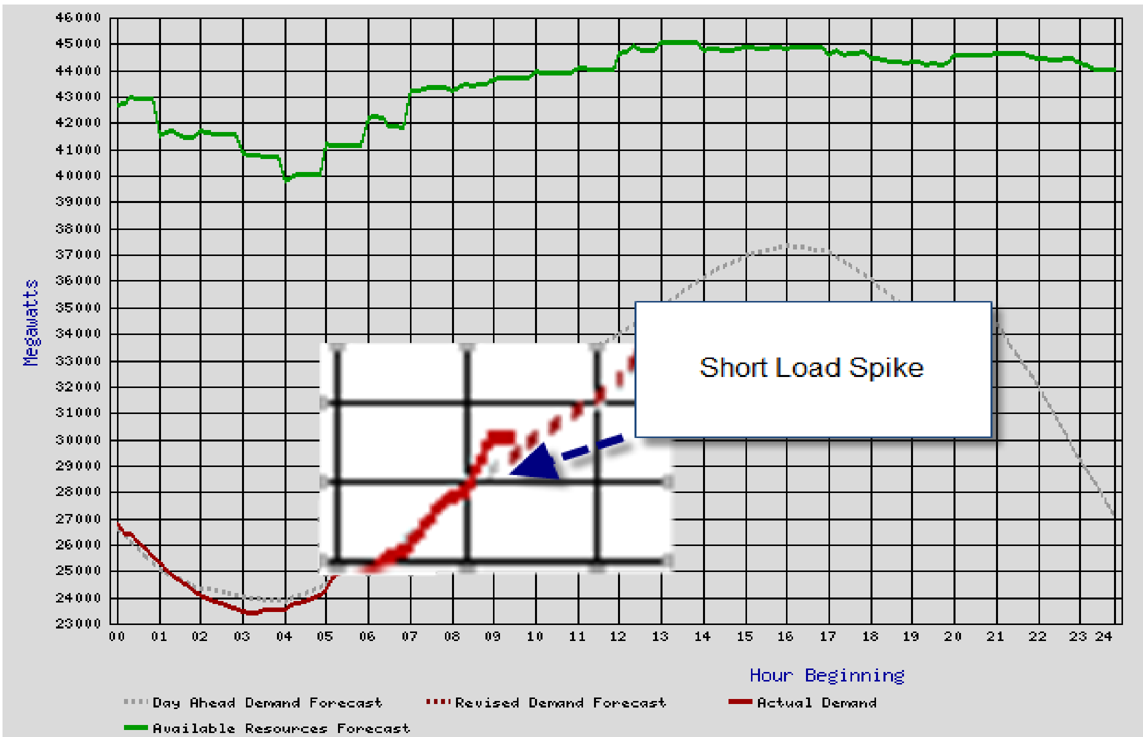
A Short History of Dispatching Resource to Meet Real Time Loads

"Balancing Energy" has a central role in each of the six existing Regional Transmission Organizations. While each RTO has implemented the concept in a different manner, the basic concept is familiar to the industry. Like the need for inventory in a store, an electric system requires the ability to adjust the system in real time. Since plant and transmission outages are generally regarded as a greater concern than inadvertent over-generation, all electric systems need generation on call to avoid service interruptions.

The balancing energy problem dates to the spirited debate between Nikola Tesla and Thomas Edison concerning the technology proposed to deliver electricity. Tesla argued for alternating current (AC) and Edison for direct current (DC). At the close of the nineteenth century, AC offered more advantages and so became the basis for today's modern system, while DC was adopted for high voltage (HV) transmission. However, an AC system's major drawback is that it cannot be dispatched in real time. DC transmission lines, on the other hand, are fully dispatchable, because an engineer can actually move electricity from point to point over a DC line.

To avoid the dispatch problem with the AC system, electricity scheduling is planned well in advance. Since the vast majority of electricity loads are quite predictable, many units can be easily scheduled a week, a day, or even an hour in advance. In a perfect world no balancing energy would be required. As we know, however, the world which we inhabit is several steps below perfection.

Even in an ideal hour, generation schedules and loads will change with weather. These changes are largely predictable, but there is always an irreducible difference between forecasts and actuals. The following chart shows a system load and the system forecast from the California ISO. The dispatchers needed 500 megawatts for 15 minutes at 6:00 A.M.



The illustration shows the relative inflexibility of real time changes, unlike the schedules planned days or hours in advance. The ISO's operators did not have the time to increase operations at a base-load unit (any base-load units can change their generating levels gradually: in the language of electricity, they can ramp up or down as required). For the short time period displayed here, the operator's best solution was to identify a base-load unit already operating in the relevant geographical area and ask the utility owning it to ramp up the unit's generation.

In traditional systems, the utility would have dispatched a unit or units from its spinning reserve. In the portion of California served by the California ISO, a unit was chosen from the "BEEP" stack (BEEP stands for Balancing Energy Ex-Post). The choice of the unit is primarily economic, although if transmission is constrained within the system, geographical constraints may also figure into the selection.

In real time the system is "stiff" – generation choices are limited and load responses are infrequently sufficient to meet requirements. Generally, the RTOs have back-stopped their balancing energy purchases with Reliability Must Run units (RMR) that can be dispatched directly without reference to economic dispatch.

Defining Competitive Real Time Markets

While there is no theoretical reason why meeting balancing energy requirements with a market should not work as well as traditional dispatch, the record of real time markets is best characterized as mixed. The central problem is market structure since real time markets tend not to meet the definition of perfect competition:

1. Many buyers
2. Many sellers
3. Transparency
4. Freedom of entry
5. Freedom of exit

Taken one at a time, it is easy to see why RTO real time markets experience frequent market failures.

Many buyers: As a general rule, only the RTO itself is the buyer. This would naturally allow the RTO to exercise monopsonistic purchasing power, but the RTO is limited to accepting the supply curve regardless of distortion. The problem is accentuated by the enormous cost of disruption should the RTO be unable to buy additional energy, for example, to offset a power outage.

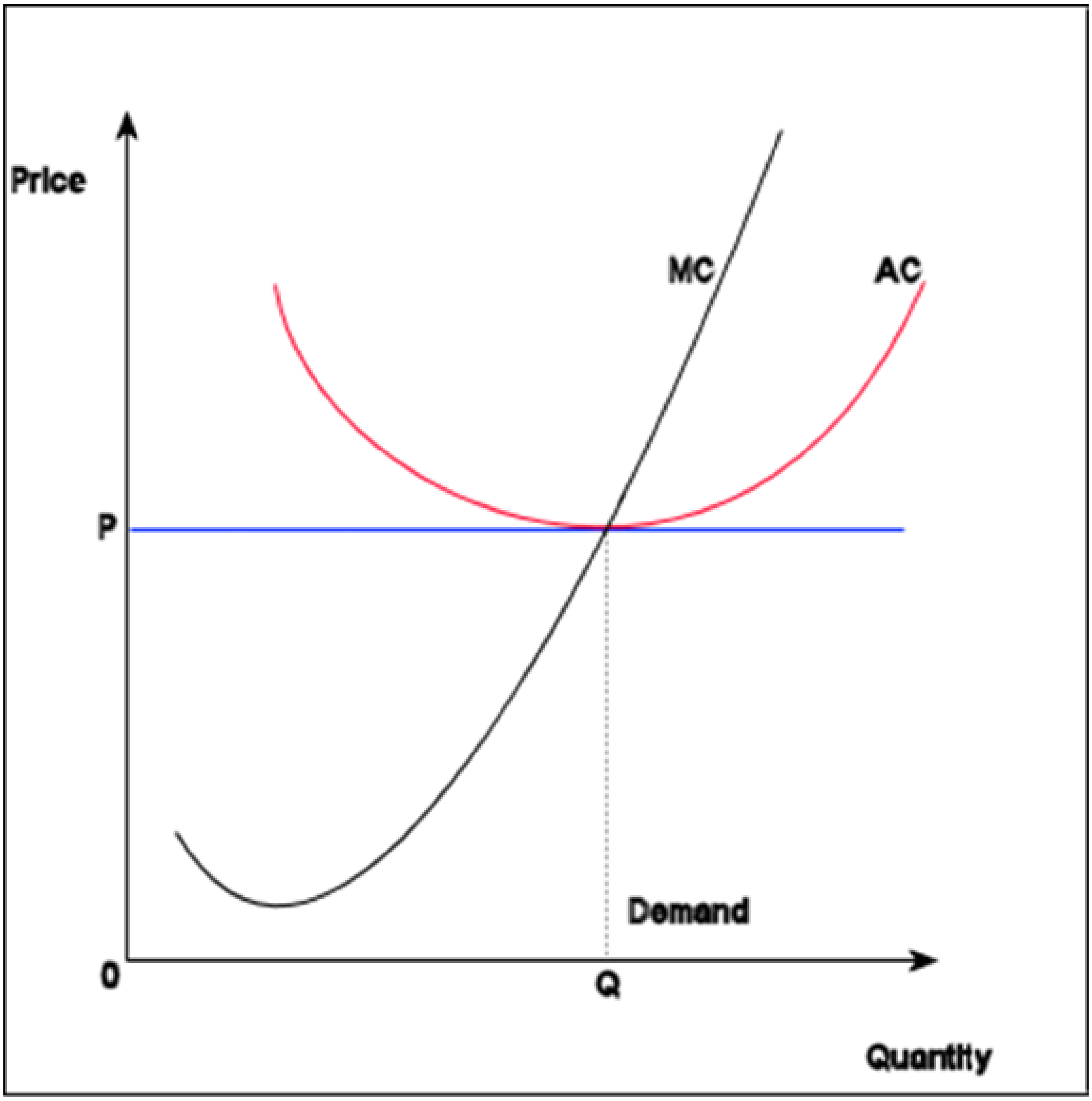
Many sellers: Also as a general rule, the number of sellers is small and is able to exercise substantial market power. Each represents a minor part of the market. Hence, it is optimal for the seller to price the electricity for sale at marginal cost. In most ISOs there are a small number of bidders and each bidder plays a major part in the market. In practice, prices vary greatly from marginal cost. In different RTOs, market manipulation gambits have led to oddities in pricing. In fact, in every American RTO, at least some non-economic bidding occurs every day.

Transparency: To varying degrees, characteristics of transparency are largely missing from the nation's electricity markets. This has not stopped RTO supporters from hypocritically proclaiming that transparency makes the manipulation of the electricity market easier to identify and monitor. Elsewhere in the economy, commodity markets are characterized by: published prices; bidders can shop openly among suppliers; non-economic outcomes are not shielded from public/regulatory scrutiny. If the major bidders have substantially more information at hand than many of their competitors, true of nearly all U.S. restructured electric power environments, then the large bidders can use the information to their advantage. This advantage is further strengthened by the ability of the bidders to manipulate the demand curve for balancing energy in many cases.

Freedom of entry: In electricity markets that operate in real time, there is effectively no freedom of entry. When the bid stack is exhausted, additional supplies can only be procured by calling upon the RMR units available to the dispatchers.

Freedom of exit: By definition, an RTO cannot leave the market. Individual suppliers are able to exit the market, but as with entry, the ability to change participation in the short-term is limited.

Perfect competition has been studied extensively since the original work of Alfred Marshall in the nineteenth century. In a perfectly competitive market the supply curve is the sum of the marginal costs of each of the suppliers. If the marginal cost curve meets the demand curve at a point above average total cost, new suppliers will enter the market and if less, suppliers will exit the market. Under perfect competition Price = Marginal Cost = Average Total Cost. The following chart shows the classic graphical representation.[2]



In a perfectly competitive market no one supplier is able to change prices. This leads to a simple demonstration that bids should reflect marginal cost. Given an existing perfectly

[2] It is relatively easy to prove that any reasonable marginal cost curve, with price expressed as a function of quantity, will intersect its corresponding average cost curve at the point of minimum average cost.

competitive price, P, assume that a supplier with a marginal cost equal to C considers bidding at a new price, P'>C. There are three possible outcomes based on whether the existing price P is below C, between C and the supplier's new bid price P'>C, or above P'>C.

> If the competitive price P is below C, the supplier gains nothing from its gambit because no one will pay more for the same product. The supplier's product will not be sold.
>
> If the price is between C and P'>C, the supplier will lose money. It would have made P minus C for the product if it had simply bid at C. Instead, it has not sold anything for the same reason as in the first case and would have been better off simply bidding at C.
>
> If P is larger than P'>C, the supplier gains nothing since the price has been set by the marginal costs of another supplier and the supplier would have been better off simply bidding at the competitive price P.

Overall, the supplier's optimum profit-maximizing strategy under perfect competition is to price the product at marginal cost.

Scarcity Prices

A central defense of the non-economic prices that frequently take place in RTO real time markets is that non-economic prices are required to attract new investment. This argument has spawned many variants to manage its economic fallacies.

At the heart of the argument is a simple economic law that no energy only (often called monomic) market will produce enough revenue in the long term to pay for new investment. The law arises from basic theory of competitive markets. If it is impossible to price capacity – a central feature of monomic markets – the shortfall caused by the missing capacity revenues will create an ongoing shortfall.[3] This is sometimes termed the "missing money" problem. Some advocates of these markets have gone so far as to recommend faking the re-

[3] Looking for the "Voom", Robert McCullough, June 26, 2007.

quirements for balancing energy to raise prices enough to offset the "missing money".[4] Others argue that non-economic pricing represents scarcity rents. The scarcity rents argument is especially pernicious since it assumes that whenever non-economic bids are successful in setting non-economic prices this is proof that shortage has occurred.

The North American Electric Reliability Corporation (NERC) publishes annual and seasonal load resource analyses for most of North America. The question of scarcity is an engineering calculation, rigorously delineated and applied, that has a definitive answer. This has not stopped advocates from rewriting the underlying data to meet their requirements. The Western Market Crisis of 2000-2001 is a case in point. Despite official reports before, during, and after the crisis that showed that no scarcity was present, defenders of non-economic prices have simply created explanations out of whole cloth to explain the crisis. In several instances they parroted Enron's claims. For example:

> The Pacific Northwest Drought of 2000: Precipitation, temperature, and run-off in the Pacific Northwest are the province of the Northwest River Forecast Center, an office of the National Weather Service. The NWRFC makes drought determinations on a monthly basis. These are published on the Web and available to any interested party. No drought took place in 2000. Advocates like William Hogan have repeated this claim again and again knowing full well that it is simply untrue.
>
> Rapid demand growth in California: California peak loads fell from 1998 through 2001. Again, the statement that demand increased rapidly during this period is simply untrue. The authoritative statistics are found in the WECC 10 Year Plans that are publicly available on the Web.
>
> High bids were due to emission costs: This untrue claim is all the more startling given that the actual prices paid for emissions are now public and the facts can be checked empirically. Again, proponents of this explanation for high real time market prices have not reviewed the actual prices.

Since the advent of real time markets in California, the NERC studies have not indicated any scarcity conditions in the U.S. on a planning basis. This is significant since it either indicates

[4] Acting in Time: Regulating Wholesale Electricity Markets, William W. Hogan, May 8, 2007.

that the advocates of non-economic pricing are incorrect or that NERC has been unable to correctly calculate reliability conditions for the past decade.

Market Failure

Evidence from California, Texas, the Central Atlantic, and New York indicates that the high prices are actually market failures (as mentioned, each RTO operates slightly differently; thus the mechanics of market failure will differ):

> California: The history of market failure is so well-known it hardly bears repeating. California's ex-post market faced manipulation on a variety of levels. These included unresponsive dispatch, fallacious bids, imaginary generation, anti-trust, and imaginary loads.
>
> Texas: The Texas PUC is pursuing TXU for exercising market power as a pivotal supplier. TXU, the major player in ERCOT's real time market, is able to set real time prices whenever the remaining bids are insufficient to meet demand.
>
> Central Atlantic states: FERC's 2008 order concerning Edison Mission demonstrates a situation in which a major market participant removed itself from the day ahead market by making uneconomic bids, thus dividing the market among competitors.
>
> New York: In July 2008, NYISO notified FERC that since January 2008 one or several market participants had scheduled transactions on eight "circuitous" paths around Lake Erie. The practice was most prevalent in April and May. The ISO "determined that 80% of the power flow originating in New York and terminating in the PJM service territory traveled over the direct interface connection between New York and New Jersey, regardless of how individuals scheduled its path. This means that the certain market participants were sending their power through the most expensive and congested corridors, but not paying the fees like everyone else. This mechanism did not just allow them to avoid paying higher fees, but also meant the power they were selling was garnering a higher price, because, as market participants increased the volume of power flow on the congested interfaces, the price for power would rise, further adding to the profits of the sellers."[5] The estimated

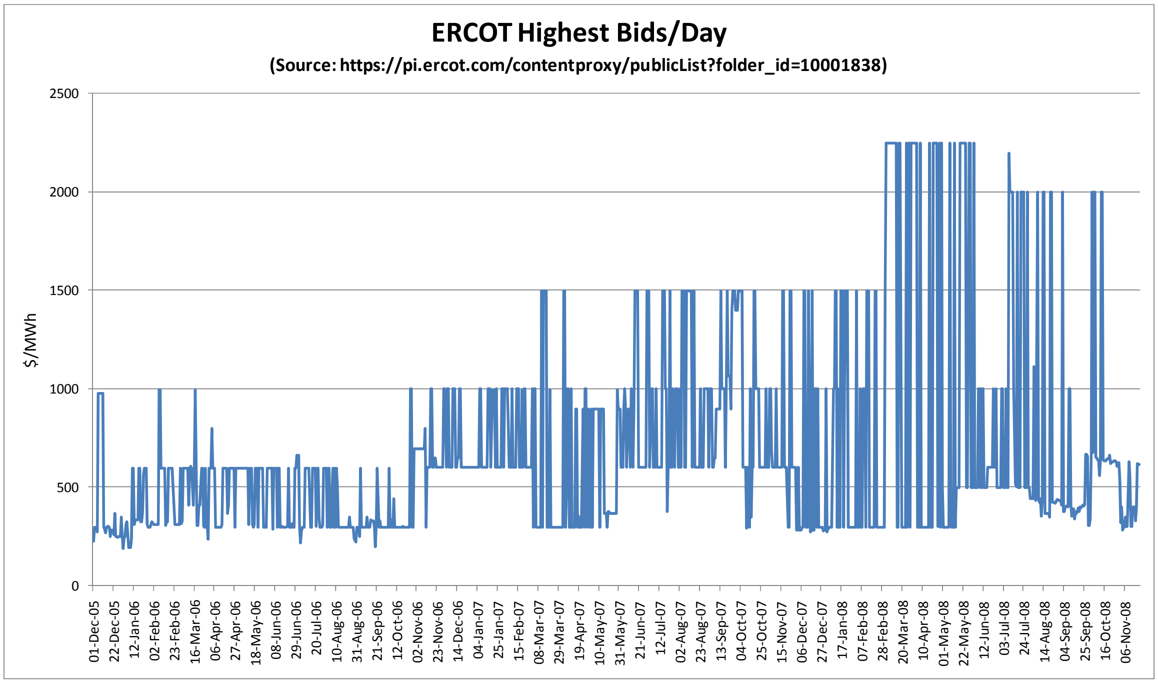
[5] Letter from U.S. Senator Charles Schumer to FERC Chair Joseph Kelliher, August 12, 2008.

> cost to New Yorkers could reach 290 million dollars (higher if the deception was practiced prior to 2008).

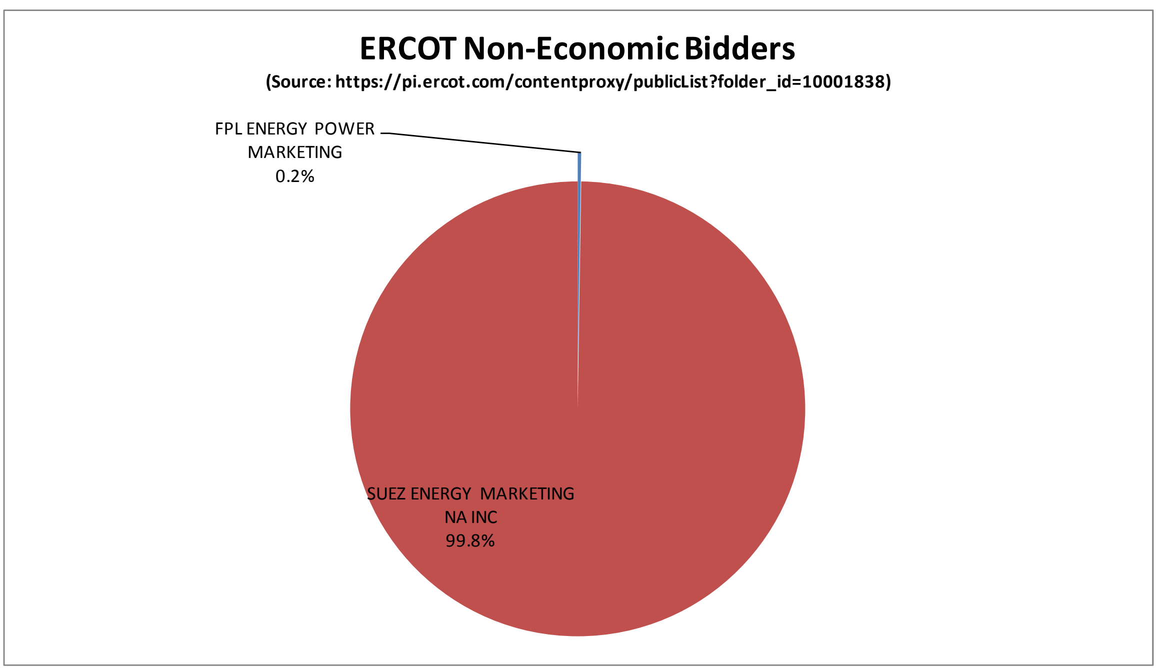
The following charts trace anomalous bidding patterns in five of the six RTOs. In the sixth, the Midwest ISO, there is insufficient data to show anomalous bids since MISO does not report bids that have not been accepted in the market.

1. ERCOT

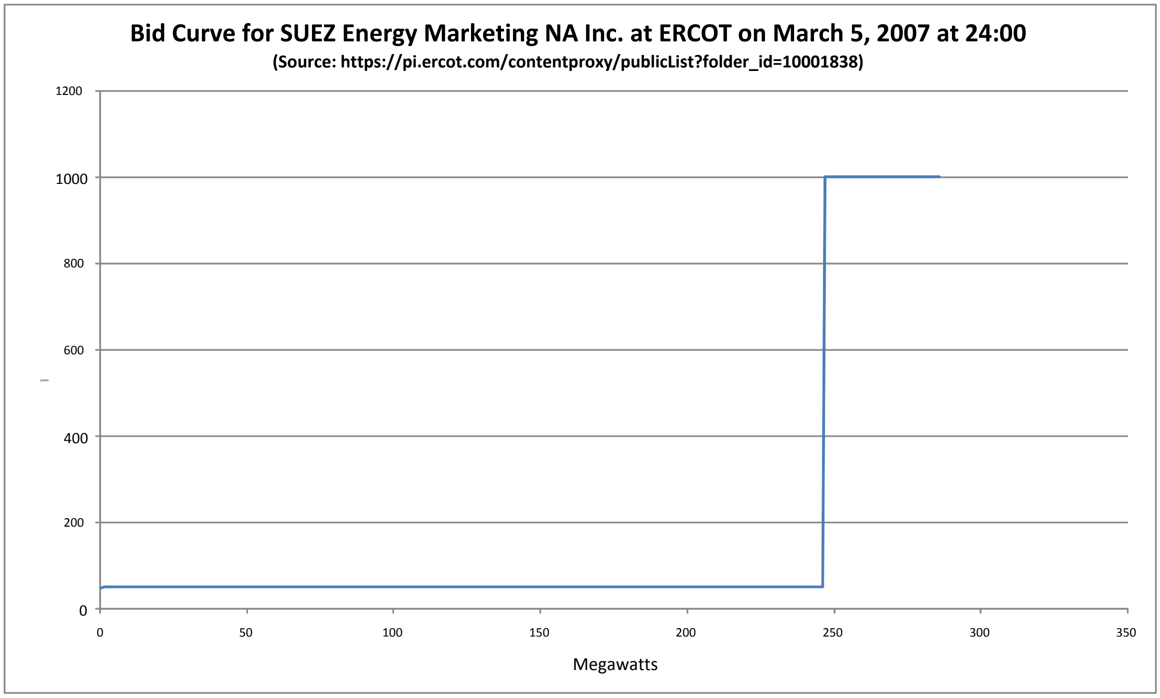
In ERCOT, two firms, Suez Energy Marketing NA Inc., and FPL Energy Power Marketing Inc., provided bids at the price cap during the time period illustrated below. Between December 1, 2005 (starting date of our data) and March 1, 2007 (the date of increase in the price cap from $1,000 to $1,500) Suez Energy Marketing submitted numerous bids at the $1,000 cap and FPL Energy Power Marketing submitted one bid (on October 18, 2006 at 2:00 pm) at the $1,000 cap. Between March 1, 2007 and March 1, 2008 (the date of increase in the price cap from $1,500 to $2,250), while FPL Energy no longer submitted any non-economic bids, Suez Energy submitted an even larger number of them at the new higher $1,500 cap. From March 1, 2008 onward, Suez Energy was the only qualified scheduling entity still submitting bids at the current $2,250 price cap until June 2, 2008.



The great majority of non-economic bids during this period were submitted by just one market participant.



A sample non-economic bid by Suez displays the characteristic "hockey stick" shape with economic bids occurring at lower levels of generation and then a sudden transition to $1,000/MWh.

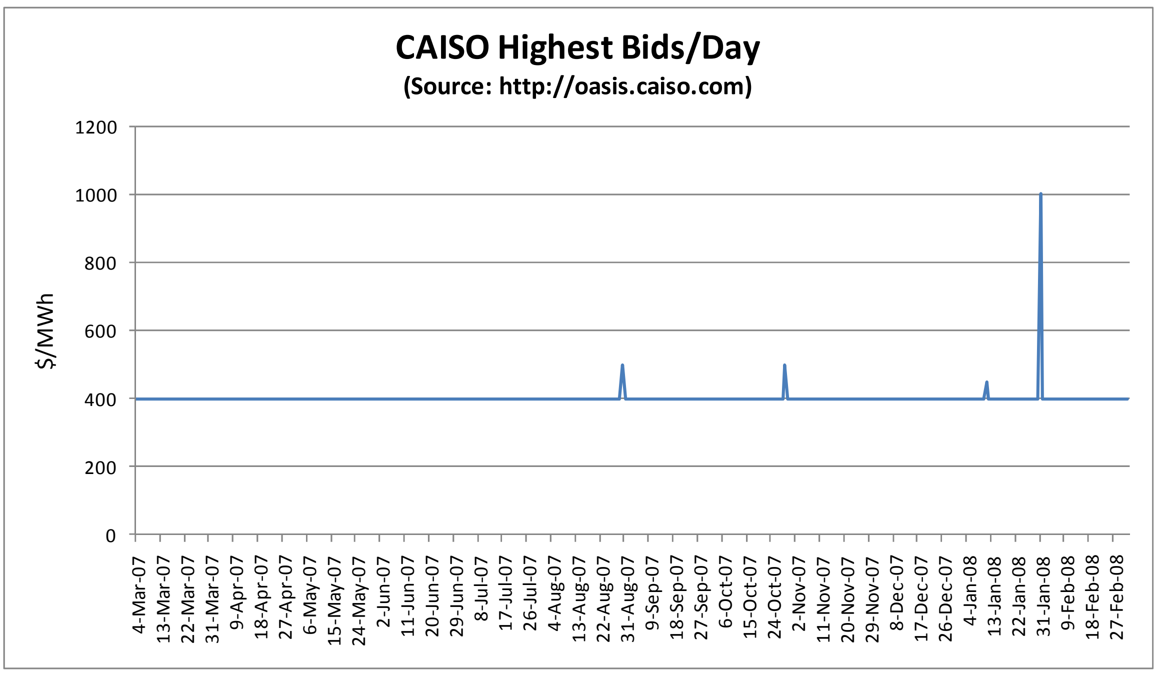


2. CAISO

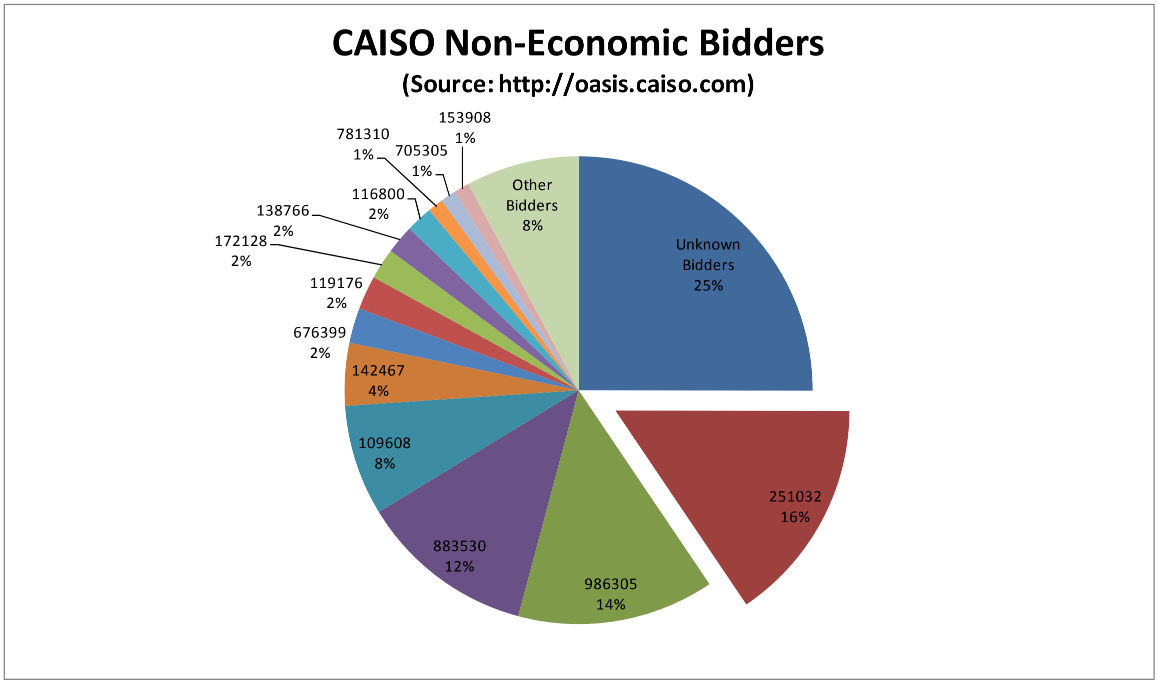
In CAISO, a "soft" cap of $400/MWh was in place during the one-year period between March 4, 2007 and March 3, 2008 we studied. A soft bid cap is one where the market participants may submit bids above the bid cap if they can provide adequate justification based on sellers' costs, but with the understanding that such bids cannot set the market clearing price that will prevail.

Bids were submitted at the bid cap throughout the period. Interestingly, the figure below indicates that there were four days when bids in excess of the bid cap were submitted, presumably with their economic justification. For instance, an astronomical bid of $1,000/MWh that was more than twofold above the effective cap was submitted by the bidder "153908" on January 31, 2008 at 17:00. There is no economic justification for such an extraordinarily high bid.

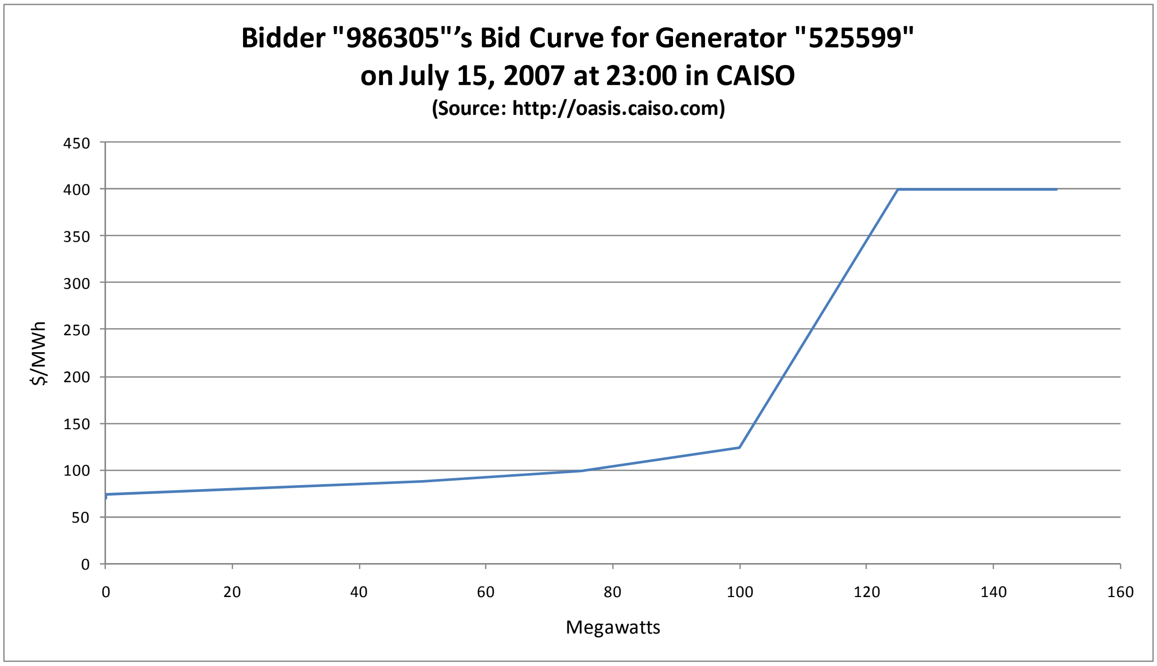
Other anomalous bids exceeding the bid cap included six bids submitted at $500/MWh, five bids at $450/MWh, and five more bids varying in price slightly above the cap. In this one-year period, a total of seventeen bids were submitted above the cap.



CAISO's non-economic bidders constitute a relatively large and diverse collection of scheduling coordinators, as demonstrated in the pie chart below. We also note that almost a quarter of the non-economic bids in CAISO were submitted by scheduling coordinators whose bidder IDs have not been identified in the publicly available bid information.

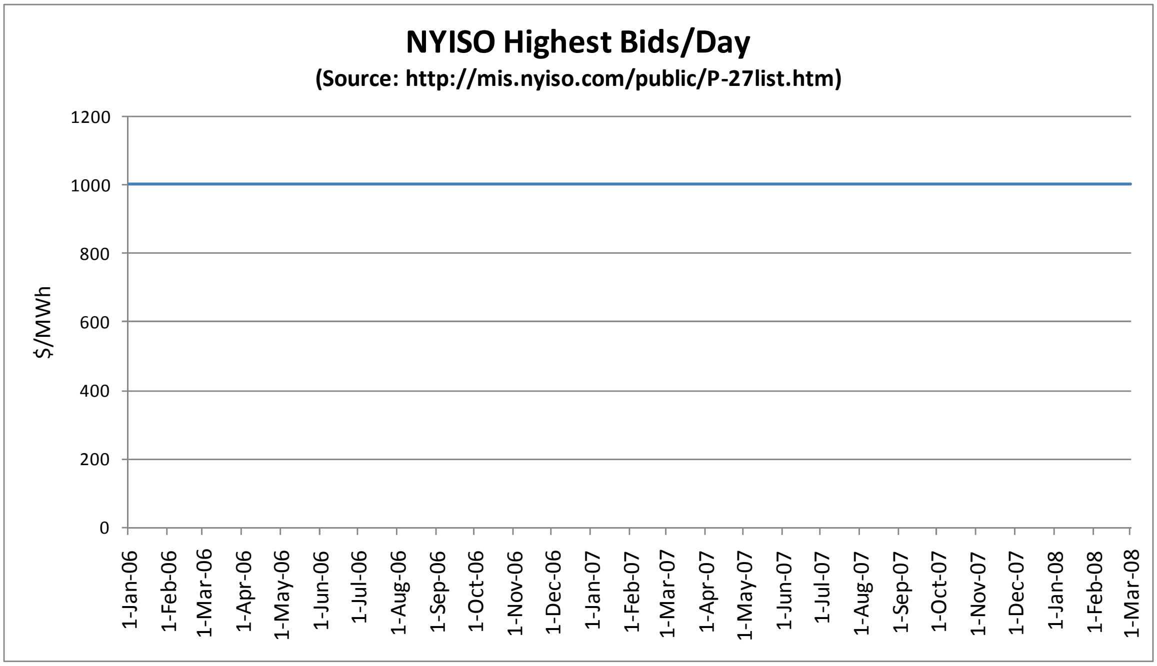


Bids by bidder "986305" for generator "525599" resemble a hockey stick, rising moderately for certain levels of generation and then increasing sharply to meet the bid cap:



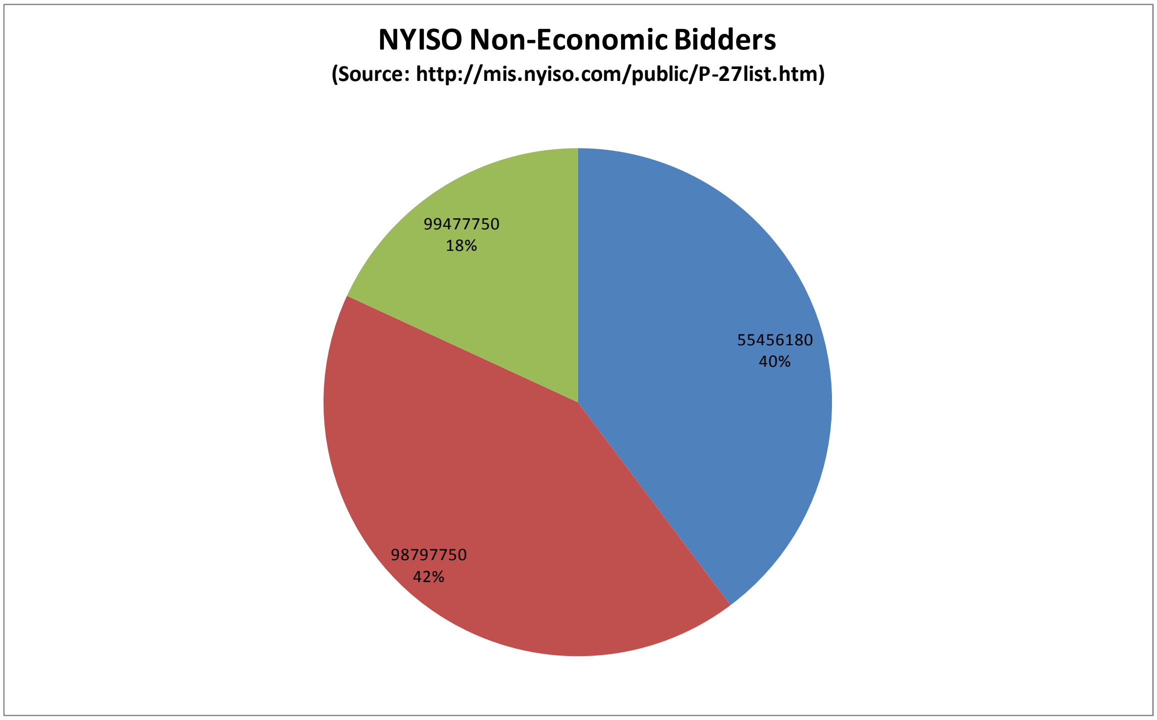
3. New York ISO

NYISO had maximum bids of $1,000/MWh in every hour for the period from January 2006 through March 2008.

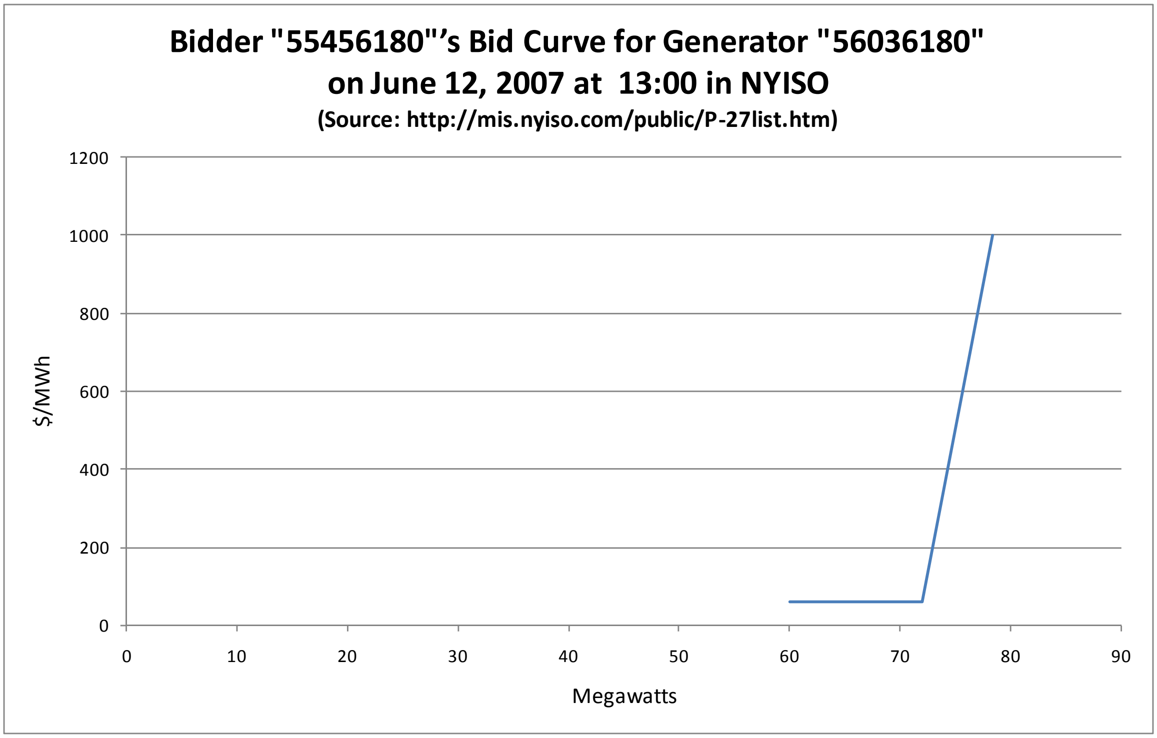


New York is unique in one respect. At least one of their non-economic bidders files bids for every hour of the year – apparently never facing maintenance or forced outages. Obviously, this implies that not only is the bid artificial, but the plant itself may also be artificial.

There were three bidders bidding at the cap in NYISO during our period of study:

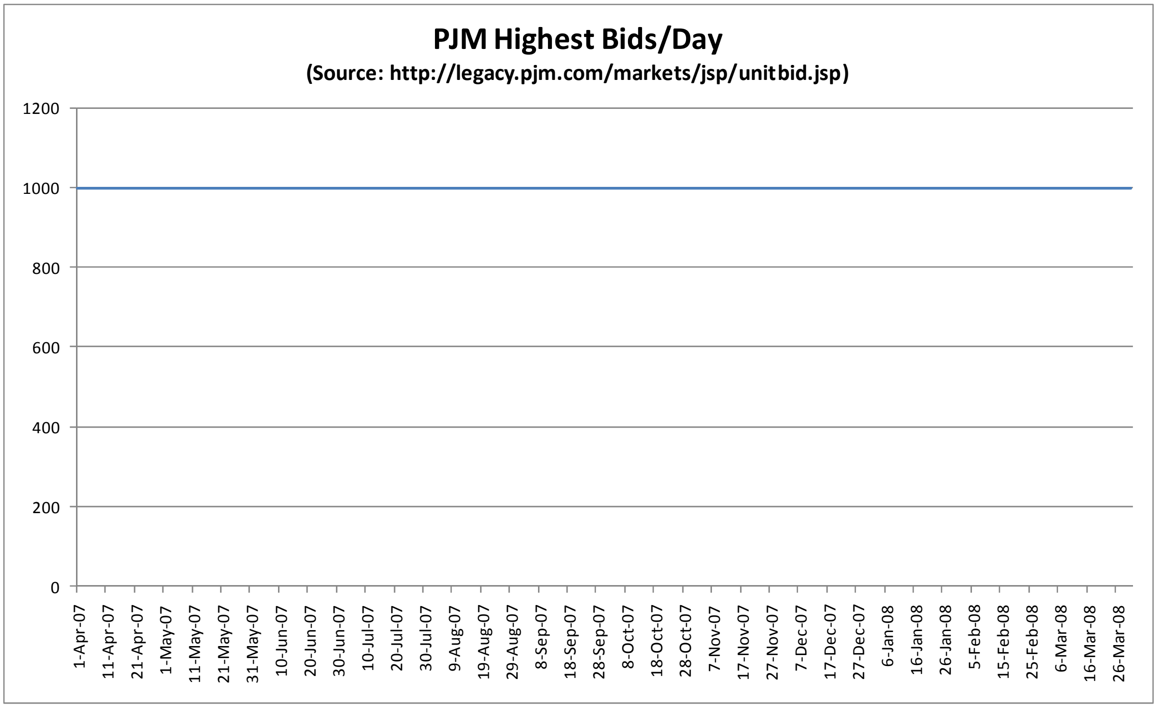


For example, the bid by bidder "55456180" for generator "56036180" on June 12, 2007 is a classic hockey stick bid.



4. PJM

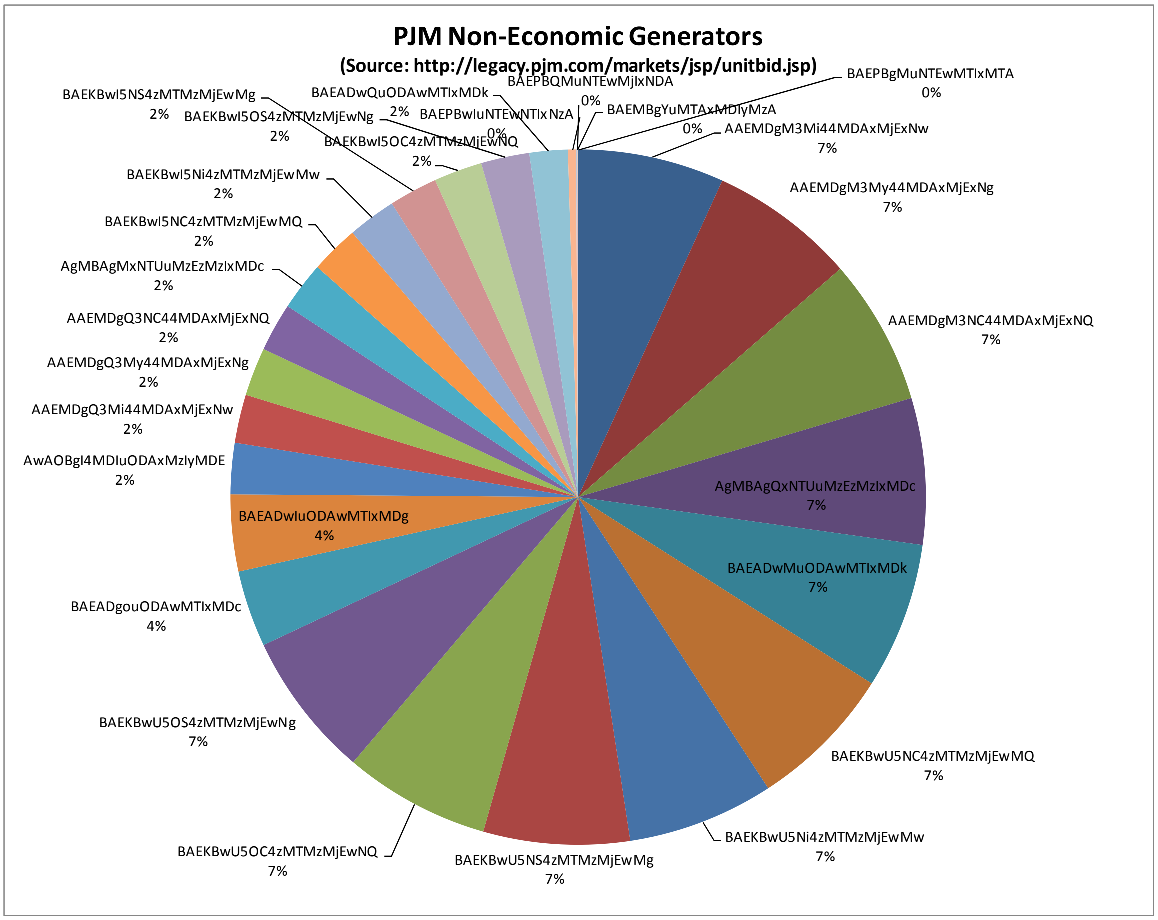
PJM's bid data is badly damaged and difficult to interpret. Unique among the six RTOs, PJM mixes together day ahead and hour ahead bids without labeling them. Discussions with PJM personnel indicate that they have no explanation for this practice. To add to the problems, a number of the bids appeared to be highly unlikely. Some bidders appeared to be using the number "1,000" to terminate their supply curves – or perhaps felt that they had plants with outputs that actually were 1,000 megawatts. Finally, PJM's data is designed to hide the identity of the bidders in an idiosyncratic fashion that is likely to only hide their identities from non-participants.



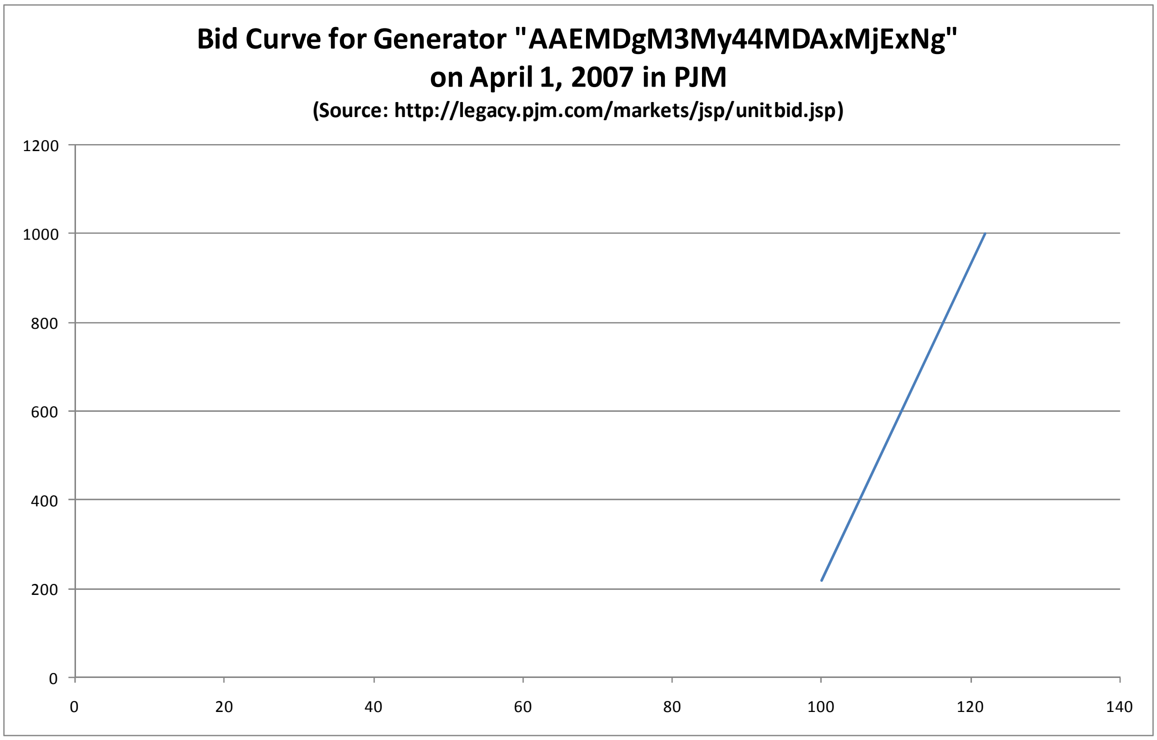
A variety of individual bidders in PJM submitted non-economic bids in a wide variety of patterns.

PJM decided on December 1, 2006 to discontinue posting in public the unique identifier numbers that are assigned to the bidders.

Due to the current unavailability of this information, associating temporal scheduling patterns with their bidders or tracing individual bidders' bidding activities over time are no longer possible. Therefore, the market concentration of non-economic bidding activity in PJM has been represented on a generator level in the pie chart below. Since many generators are owned by just a few market participants, this chart overstates the diversity of the non-economic bids.

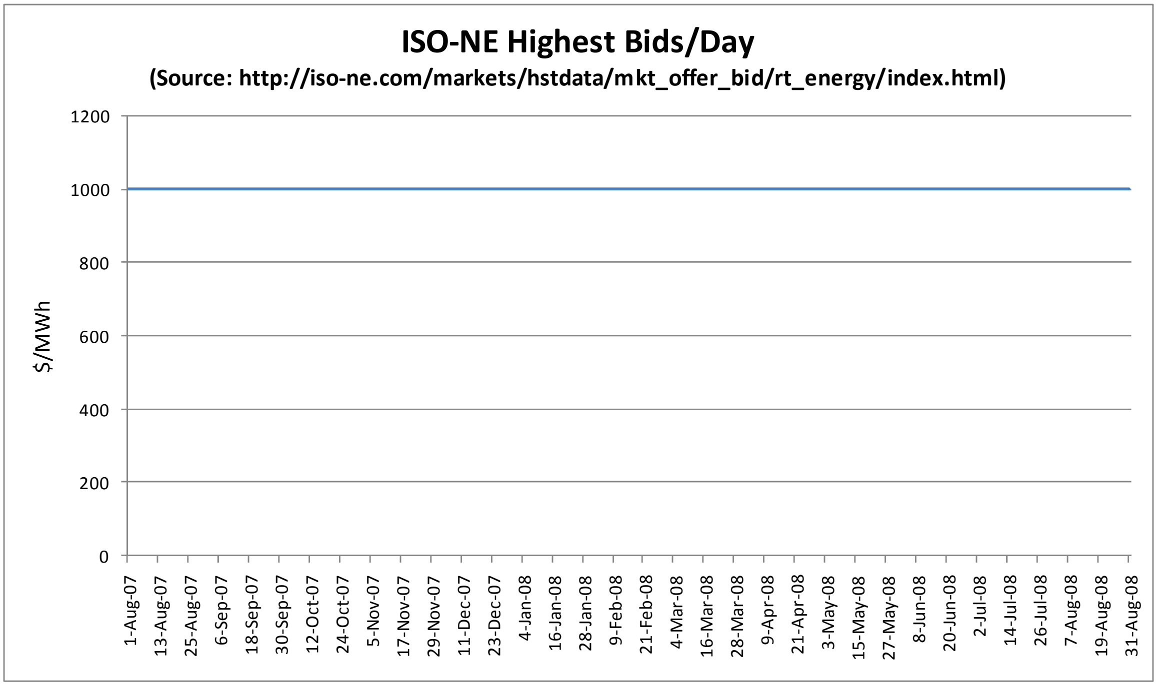


In the following example, one plant increased its bid from less than $100/MWh to $1,000/MWh over 22 megawatts.

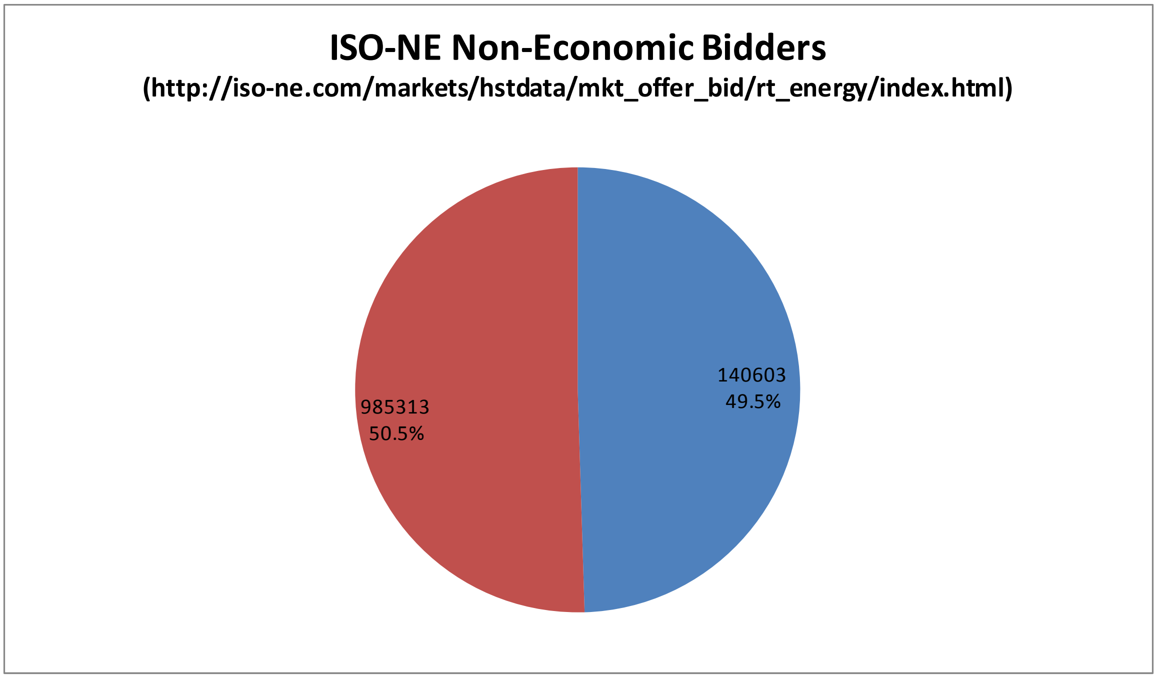


5. ISO New England

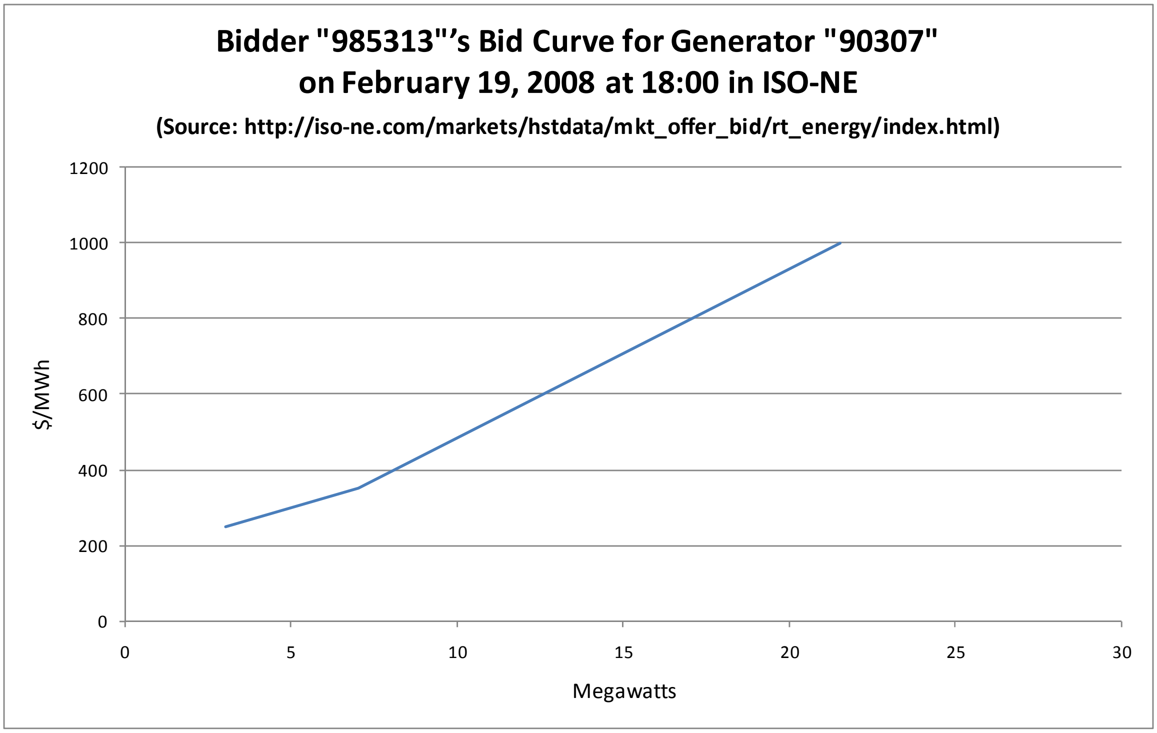
Like New York, ISONE also has bids equal to the bid cap every hour of every day.



Bids at the bid cap are split between two non-economic bidders in a pattern similar to that in ERCOT.



Bidder "985313"'s non-economic supply curve for generator "90307" also represents the classic hockey stick pattern.



While the existence of non-economic bids is often cited as proof that scarcity is a frequent event in balancing energy markets, reliability reports from NERC and each of the regional reliability councils tell another story. For the ten years that balancing energy markets have been active in the six U.S. RTOs, the authoritative reliability council reports have not identified any cases of scarcity on a planning basis. Regardless of that fact, non-economic bids occur continuously – often in every hour.

Such a striking difference between actual behavior and economic theory deserves an explanation. One possible explanation is the behavior of bidders in markets with a vertical supply curve.

Demand and Supply in Real Time

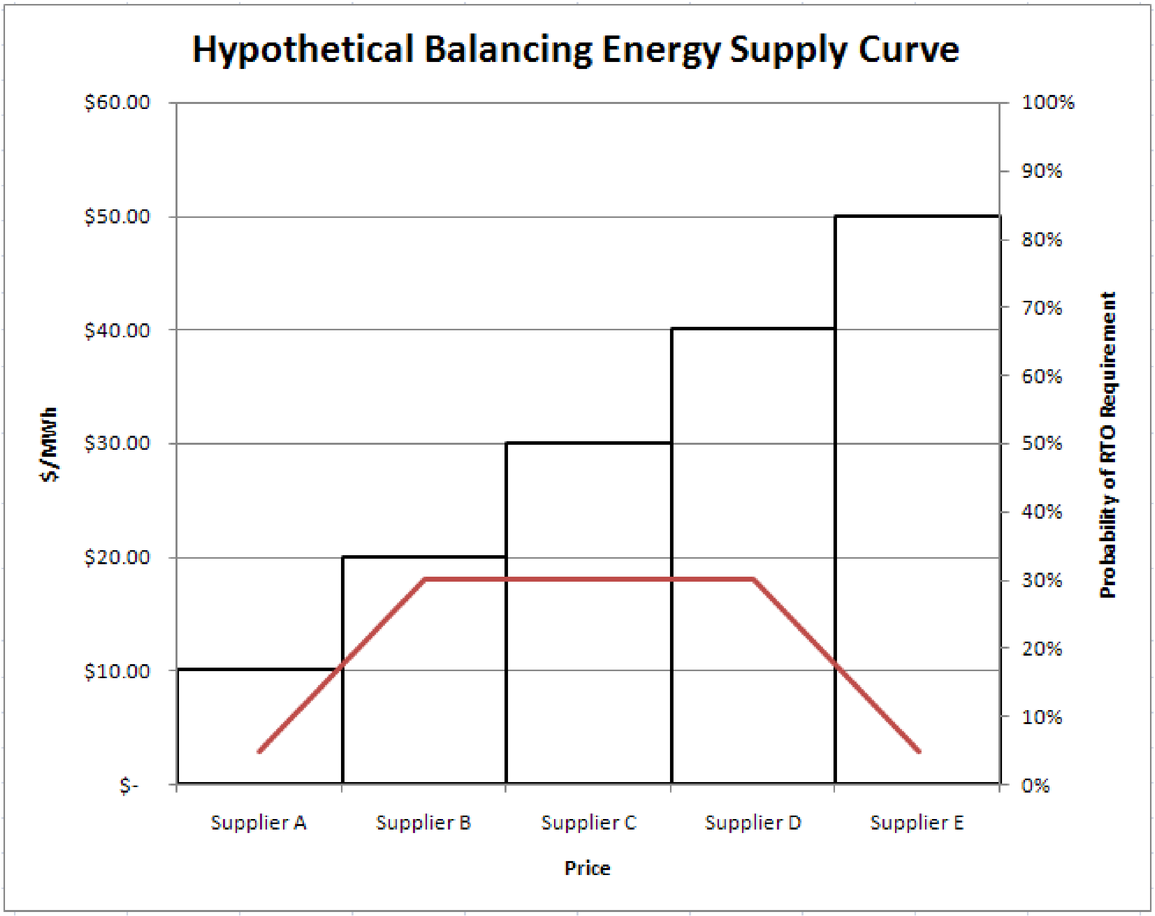
Balancing energy markets differ in several ways from the traditional Marshallian supply and demand chart shown below. The most important difference is that the demand curve is vertical; an RTO's requirements are fixed. When balancing energy is required, it is for a very short period in which the price will not change the quantity needed.

A vertical demand curve is unusual in economics. Few instances can be imagined other than a market for fire protection after the fire has started, or for medical care after an individual is diagnosed with an immediate, life-threatening medical condition. In these cases, there is no upward bound on prices, since an individual is most likely willing to pay anything for help in a fire or in the emergency room.

RTOs address this problem by setting price caps. In policy debates around the U.S., however, it is clearly understood that the price cap is setting pricing behavior. This reflects the dynamics of bidding when the level of demand is fixed.

The chart illustrates a simple example where five suppliers, each with comparable levels of capacity, form a "stair step" supply curve. Each generator has a marginal cost just $10 per megawatt-hour greater than its next more efficient competitor. The bell curve[6] shows the probability of the RTO's demand level hitting its portion of the supply curve.

[6] The example calculations use the same 30% probability for the three middle suppliers and 5% for the highest- and lowest-cost suppliers, with the five probabilities adding up to one.



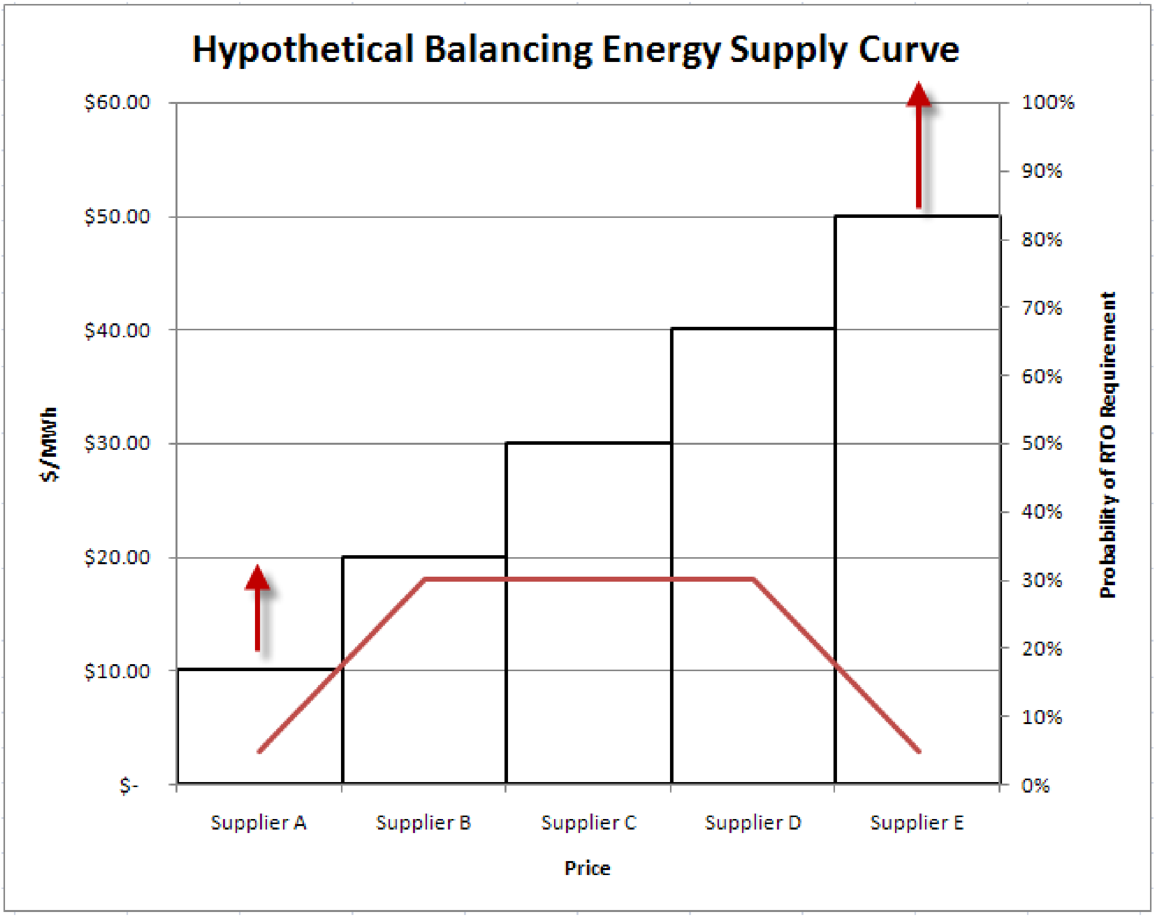
Supplier E has a marginal cost of $50/MWh and can expect to be called upon only 5% of the time. The majority of time, suppliers A through D will be called upon because their marginal costs are considerably lower. By the same token, Supplier A, the most efficient generator, can always expect to be called upon since no other supplier has a lower cost.

In this example, the RTO's expected price for balancing energy is the sum of the marginal costs times the probability of a specific marginal cost setting the market price, or 5% x $10 + 30% x $20 + 30% x $30 + 30% x $40 + 5% x $50 = $30.00 per megawatt-hour.

Each of the suppliers can expect a profit except for Supplier E. Supplier A, of course, operates 100% of the time, and receives $20/MWh for each megawatt-hour it produces. Supplier B only operates 90% of the time. When Supplier B sells into the market the average price is $31.05 and its profit is $11.05 per megawatt-hour. It is immediately apparent that Suppliers A and E can profit considerably more with a better bidding strategy. Supplier E's situation is straightforward. E's plant is only dispatched 5% of the time, but when it is dispatched there is no alternative. This is the definition of a pivotal supplier. Supplier E can raise the bid to $1,000/MWh – the price cap in this hypothetical example – since it will not change the chances of being called upon.

Supplier A has a similar situation. Supplier A is always called upon because it is the least expensive supplier. The most conservative bid would be to raise the price it asks to just under the level of Supplier B's marginal cost. Then during periods when only Supplier A's generation is needed, it will receive twice the revenue.

Both Supplier A's and Supplier E's strategies are riskless because they are able to unilaterally raise the price to the RTO during periods when they are the pivotal suppliers. Their new bidding strategies can never lose money relative to the perfectly competitive supply curve. Their new bidding strategies will have an immediate impact on the expected cost to the RTO since the expected price for balancing energy will increase from $30.00 to $78/MWh.



It is important to note that the change in bidding strategy for Supplier A and Supplier E is good for all of the bidders, because the higher price for Supplier E is enjoyed by each of the other suppliers in the 5% case when its generation sets the price.

The change in profits is highly significant. The profit for Supplier A rises from $20/MWh to $68/MWh. The profit for Supplier B rises from $11.05 to $61.05/MWh. It is also important to note that the changes do not affect the market shares of any of the other suppliers.

Now consider Supplier D whose situation is fraught with difficulty. It would like to follow Supplier E to the price cap, but if Supplier D sets its price to 1 cent under the price cap, Supplier E could underbid Supplier D by 1 cent. In this case, Supplier E would be dispatched 35% of the time (5% when E is a pivotal supplier and 30% when D is underbid).

This is a classic problem in game theory. In its simplest formulation there are four possible outcomes:

Outcome 1: D and E both bid at marginal cost. Supplier D makes 5% x ($50-$40) or $.50. Supplier E makes nothing.

Outcome 2: Supplier D bids at marginal cost and Supplier E bids at the price cap. Supplier D makes $48 and Supplier E makes $47.50.

Outcome 3: Supplier D bids at the price cap and Supplier E bids at marginal cost. Supplier D makes 5% x ($1,000 - $40) or $48.00 per megawatt-hour. Supplier E makes 5% x ($1,000 - $50) or $47.50.

Outcome 4: Both D and E bid as close to $1,000 per megawatt-hour as possible. Supplier D's profit is 35% x ($1,000 - $40) or $336 per megawatt-hour. Supplier E's profit is 35% x ($1,000 - $50) or $332.50.

This is a traditional game theory payoff matrix:

		Supplier E	
		Marginal Cost	Price Cap
Supplier D	Marginal Cost	$ - $ 3.00	$ 47.50 $ 48.00
	Price Cap	$ 47.50 $ 48.00	$ 332.50 $ 336.00

The classical answer is relatively easy. If both players are conservative, the least risk for Supplier D is to choose to bid near the price cap. The least risk for Supplier E is also to bid the price cap. If both firms are conservative, their behavior will set the RTO price to the price cap 35% of the time.

In this case, the price will be set by the lowest cost bidder 5% of the time. The two next-highest-cost bidders will also be bidding at marginal cost. The two highest-cost bidders will be setting the price at $1,000 35% of the time. The weighted average of market prices and their probabilities are $366.00 per megawatt-hour.

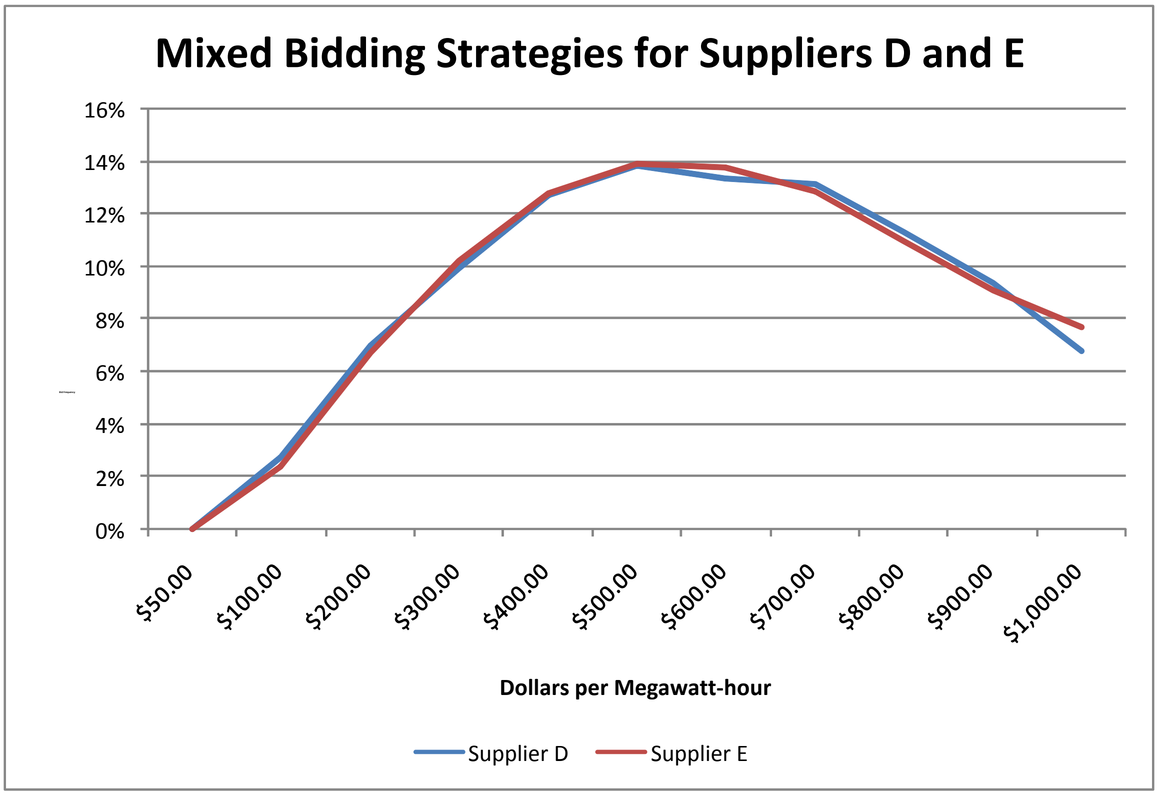
A careful study of the payoff matrix, however, may well lead Supplier E to another strategy. It would prefer an outcome where it bids just under the price cap when Supplier D bids the price cap. Then Supplier E would receive the price cap, or a price just under the cap 35% of the time, but Supplier D would only receive the price cap 5% of the time. This strategy is likely to provoke the following response from Supplier D:

		Supplier E	
		Price Cap - $1	Price Cap
Supplier D	Price Cap - $1	$ 332.15 $ 335.65	$ 47.50 $ 335.65
	Price Cap	$ 332.15 $ 48.00	$ 332.50 $ 336.00

If the two suppliers' options were restricted to only these two prices, the answer would clearly be option four where both suppliers always bid the max. What happens in the real world where the two suppliers have a choice of bids ranging from their marginal cost all the way up to the cap?

This game theory matrix has no easy solution. Supplier E prefers a price just below the cap since it allows a chance that it can take Supplier D's position in the supply curve. Supplier D prefers a price just below the price cap as well. Obviously, this is an unstable situation where each has an incentive to decrease prices ever so slightly to take advantage of the other.

Dynamic games of this type lead to mixed strategies where the expectation that Supplier E will bid at the price cap is sufficient to keep Supplier D pricing at the price cap. The final strategy depends on Suppler E's expectations of the behavior of Supplier D and vice versa. A dynamic simulation of an optimal mixed strategy for Suppliers D and E can be modeled. The following illustrates the result of a 1,000 iteration dynamic game where each has optimized its bidding strategy based on its competitor's past behavior:



In this case, Supplier D's bids average $580 per megawatt-hour and Supplier E's bids are $584.04 per megawatt-hour. The situation is improved for the RTO because prices only average $219.00 per megawatt-hour in this example. It should be noted that no assumption has been made of collusion. The hypothetical generators have acted in their own self-interest with no communications with their competitors. Also, as can be seen by the difference in the games analyzed above, collusion is a powerful force in markets where surveillance is minimal and data on bids and bidders is secret. While the hypothetical example analyzed here is just that, the level of concentration is not unduly high by RTO standards. The Hirschman

Herfindahl Index for this example is 2,000. Concentration indices for RTO market are often significantly higher.

Unfortunately, the technical problems with balancing energy markets are not restricted to the slope of the demand curve. A very serious problem is whether the level of demand is under control of one or more market participants.

Manipulating the Supply Curve

A central feature of the Western Market Crisis was the shortage of prescheduled energy and its necessary replacement in the real time market. Given the complexities of the electric system, attempting to operate it in real time is both expensive and risky. During the 2000-2001 California crisis and the 2003 crisis in Texas, each system was effectively forced to drive "ahead of its headlights".

In California a central feature of several "Fat Boy" schemes used by Enron and others was to purchase energy in the day ahead market and then to submit erroneous schedules for the same energy in the real time market. The term for energy scheduled in error, "inadvertent energy", carries a negative connotation since it implies errors in dispatch and transmission planning that add expense and reduce reliability.

In California, this "Fat Boy" scheme involved thousands of megawatts. The FERC Final Staff Report addressed the problem directly:

> Enron's use of the fat boy trading strategy did not set the market clearing price in the Cal ISO's real-time market. Under California market rules, entities are price takers for the amount of generation in excess of actual load; that is, they are paid the clearing price that is established in the Cal ISO market.[29] Nevertheless, the submission of false schedules, and the Cal ISO's encouragement of such fabrications to circumvent the balanced schedule rule, would be prohibited under Staff's recommendations in the Initial Report. The Initial Report included a recommendation that all tariffs for market-based rates include an express prohibition against submitting

> false information. In addition, all open access transmission tariffs should be amended to include this prohibition. Flawed market rules that are not working as intended should be amended by the Commission, not circumvented by market participants. More significant was the elimination of the market rule that held the three California public utilities in the spot market. As stated in the Initial Report, allowing a greater use of forward contracting resulted in far less reliance on the spot market, thus reducing the economic incentive for this trading strategy.
> While Staff has concluded that the fat boy trading strategy alone did not set the market-clearing price in the Cal ISO's real-time market, and may in fact have been encouraged by at least one Cal ISO employee, this trading strategy nonetheless involves the deliberate submission of false information and falls within the scope of the antigaming provision because it necessarily involves taking "unfair advantage" of the Cal ISO's rules and may otherwise have made the "ISO Markets vulnerable to price manipulation to the detriment of their efficiency."[30]
>
> [29]The day-ahead and real-time imbalance pricing during May 20-23, 2000 illustrates this trading strategy. Unexpected high loads occurred on May 20-21, which caused prices in the Cal ISO real-time market to reach the $750 price cap while the Cal PX day-ahead prices were in the $40 to $50 range. Reacting to these prices, Enron and British Columbia Power Exchange Corporation overscheduled between 1,000 and 2,000 MW of generation as "price takers" in the Cal ISO real-time market on May 22. Because the Cal ISO market continued to exceed the Cal PX day-ahead prices, the fat boy strategy was profitable relative to selling in the Cal PX. On May 23, 2000, these two scheduling coordinators continued to overschedule more than 1,000 MW in the Cal ISO imbalance market. However, the Cal ISO's market dropped to the $200 range, while prices in the Cal PX rose to the $300 to $500 range. Thus, this overscheduling strategy ceased, for a time, to be profitable relative to selling in the Cal PX.
> [30]MMIP 2.1.3.

Similar schemes have been undertaken in both ERCOT and PJM. In all three cases, the objective is to reduce energy procured in the relatively more efficient day ahead markets and force the RTO to purchase requirements in the real time market.

The Texas crisis in 2003 represents an extreme case of demand manipulation. ERCOT's real time markets depend on a complex series of reports filed by the major utilities concerning expected loads and expected prescheduled generation. During the March-April period when non-economic prices occurred, there is substantial evidence from the resource plans and other filings that market participants were deliberately raising the proportion of energy that needed to be procured in the real time market.

The problem lies in the inability of the RTO to police the complex engineering issues of real time markets. At the height of the Western Market Crisis FERC dispatched a team to ask the five merchant operators in the California market whether their abysmal reliability record represented economic withholding. The team reported back that they could find no evidence of withholding, even though availability of the units owned by the five merchant operators only averaged 50% during high load periods. As we now know, a number of plants were later implicated in such withholding activities, including one owned by Enron where the orders to take the unit out of service were issued over a recorded telephone line.

The possibility of non-economic prices in real time markets poses an irresistible temptation to market traders. Metaphorically, it is the equivalent of running a jewelry store on the honor principle. Clearly, diamonds and rubies are likely to tempt a customer to shoplift. For that reason, jewelry store owners employ stringent safeguards to avoid giving shoppers the temptation.

The situation in balancing energy markets is far more difficult. Since many participants in balancing energy markets are also major generators as well as being retailers that serve large loads, the ability to affect the demand for balancing energy is always present. Simultaneously, the absence of a downward sloping demand curve provides little incentive for effective competition. The combination of these factors makes non-economic results common even when the chance of scarcity is exceedingly remote.

Transparency

It is not surprising, therefore, that the calls for secrecy in bids, bidders, and bid resolution have been so strident. Across the U.S. it is common for bidding data, bidders, and the algorithms used to resolve the bids to range from difficult to completely secret. The situation ranges from the highly secretive PJM where the bidders are secret, the bids are reported in an unusable format, and the algorithm is secret, to ERCOT where most data is available after a sixty-day delay and only the algorithm is unknown.

The nature of the RTO process also makes the secret data available on an asymmetric basis. Members of the relevant committees at the RTOs often have access to considerably more information than market participants who are not represented on the committees. Additionally, minutes from system operations committees are not protected by open document and open meeting laws, so that major players have considerably more information about the nature of the markets than minor players.

One interesting facet of the investigation into Enron's activities in ERCOT and the California ISO was the discovery that Enron frequently had preferential access to highly significant market information at both RTOs. Even absent Enron's market manipulation activities in both areas, its access to theoretically secret information gave Enron an enormous strategic advantage.[7]

Lobbyists who energetically support secrecy in balancing energy markets argue that absent secrecy conspirators will be able to share information. This is an odd argument. Information sharing is a very common practice (as shown in the trader tapes that came to light during the Enron investigation). Traders can easily coordinate bids, share privileged information, and affect market outcomes without the permission or knowledge of the RTO.

The Enron/Powerex market sharing agreements in the "Project Stanley" scheme practiced in Alberta explicitly involved sharing market information.[8] The criminal antitrust risk was so

[7] In December 2000, for example, CAISO decided to file imaginary transmission schedules to block exports from California to neighboring states. The policy was only discovered in hearings of the Senate Select Committee to Investigate Price Manipulation of the Wholesale Energy Market in 2003. Enron emails identified briefings concerning this secret practice in the spring of 2001.

[8] Enron's internal analysis of Project Stanley appears in the information released during the Enron investigation. Enron's Project Stanley PowerPoint is entitled "Project Stanley - July 20th Review Meeting Initial market analysis", July 20, 1999.

great that one especially poignant Enron trader tape involved a senior Enron executive worrying about going to jail in two countries.[9]

Recommendations

Two recommendations are indicated by this review of balancing energy markets:

1. Market information can only be kept secret by extreme methods. These methods are not only beyond the control of the RTO, they may be impossible when balancing energy bidders sit on the RTO committees, help determine the level of balancing demand, and employ traders who communicate continuously. The solution is to level the proverbial playing field by allowing consumers, the media, and decision-makers access to the same information.

2. The prevalence of non-economic prices does not reflect a valid signal for scarcity, nor do high real time prices represent actual operations. The optimum solution is for real time bids to be limited by documentable marginal costs including fuel, variable O&M, and verifiable environmental charges.

[9] Taped conversation between Timothy Belden and John Lavorato, Exhibit SNO-221 in FERC EL03-180.