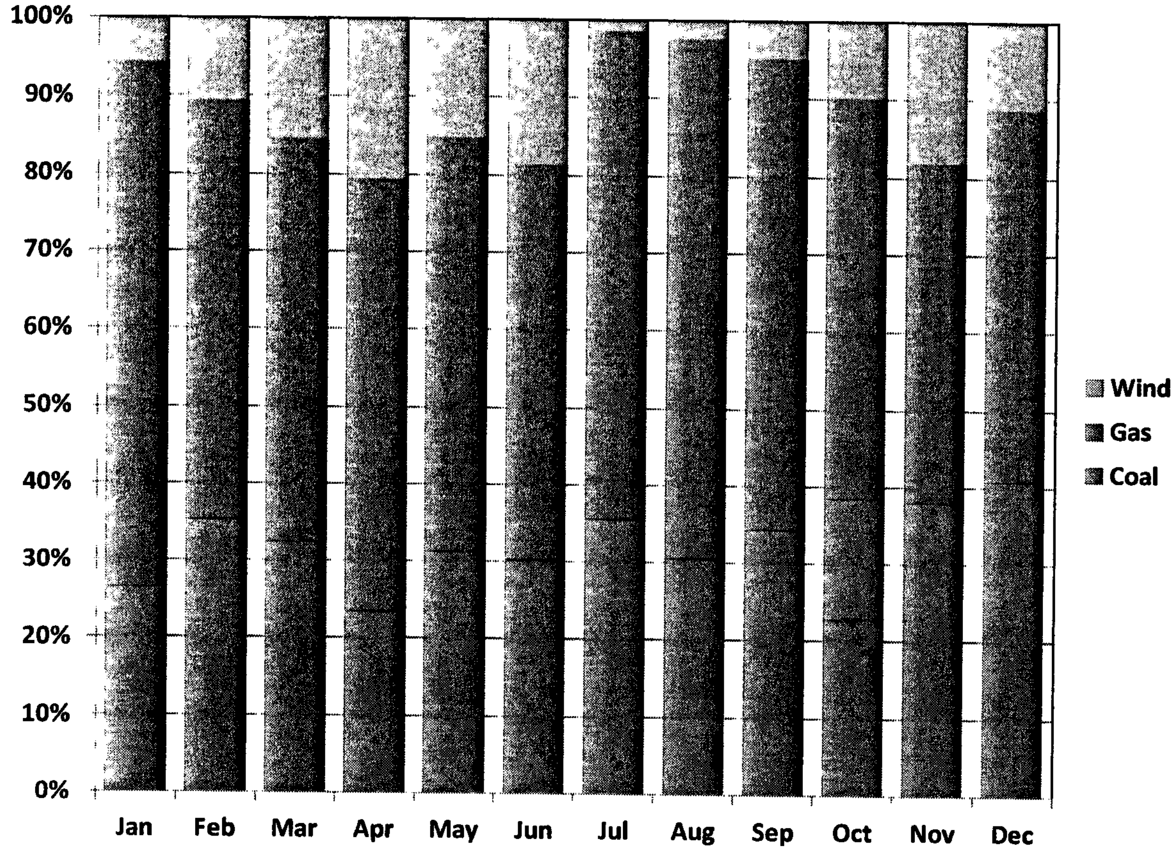
Figure 51 shows, consistent with the previous two years, the frequency with which coal was the marginal fuel averaged just over 30 percent in 2011.

Figure 51: Marginal Unit Frequency by Fuel Type



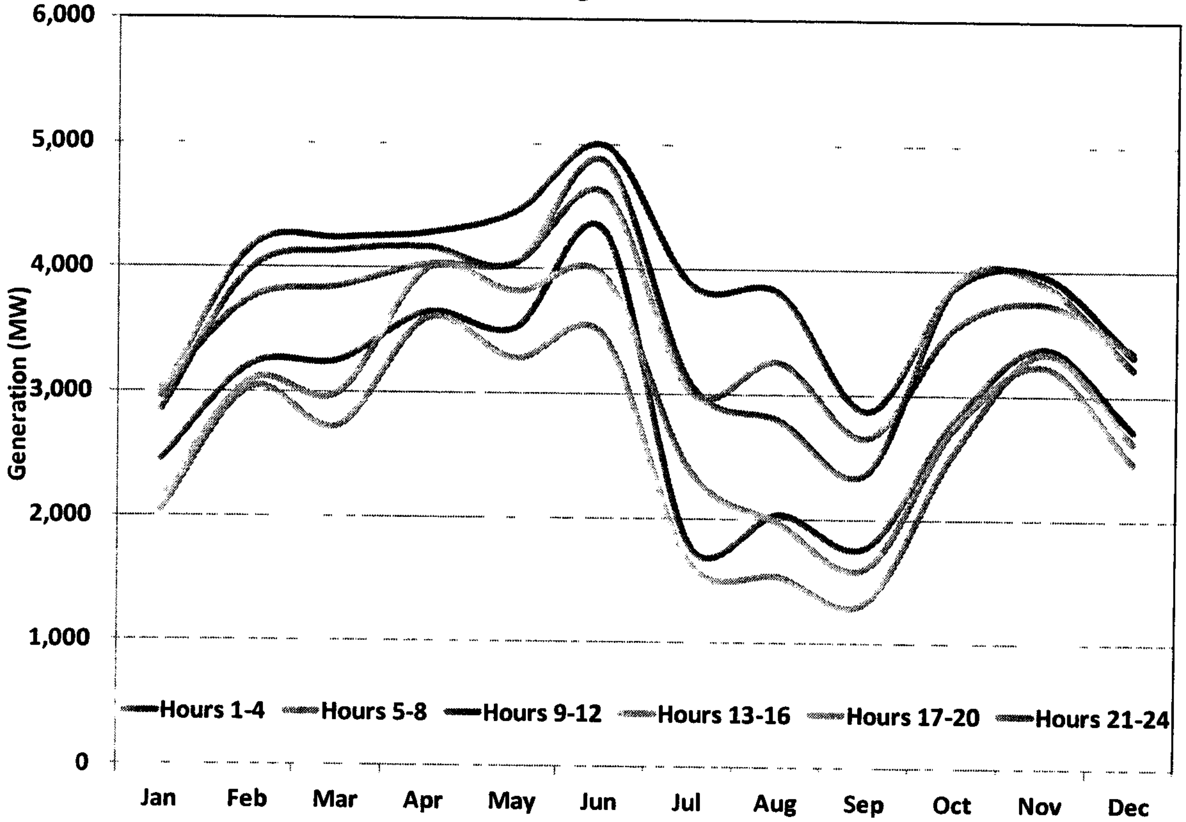
The methodology used in this analysis has been revised to reflect the details of the unit specific dispatch that are available under the nodal market design. For every five-minute interval we determine which units are marginal, that is they are being dispatched and their offer price is contributing to the locational marginal price. With all the marginal units identified, we aggregate by their fuel type to compute monthly percentages.

Natural gas units continue to be marginal the majority of the time. Wind units are marginal up to 20 percent of the time, with the increased frequency coming in months with the highest wind generation output. The contribution of wind generators setting the marginal price, particularly in the West zone, continues a trend observed since 2008.

1. Wind Generation

The amount of wind generation installed in ERCOT exceeded 9 GW by the end of 2011. Although the large majority of wind generation is located in the West zone, there has been more than 1 GW of wind generation recently installed in the South zone. Additionally, a private transmission line went into service in late 2010 allowing nearly another 1 GW to be delivered directly to the South Zone. This section will more fully describe the characteristics of wind generation in ERCOT.

Figure 52: Average Wind Production

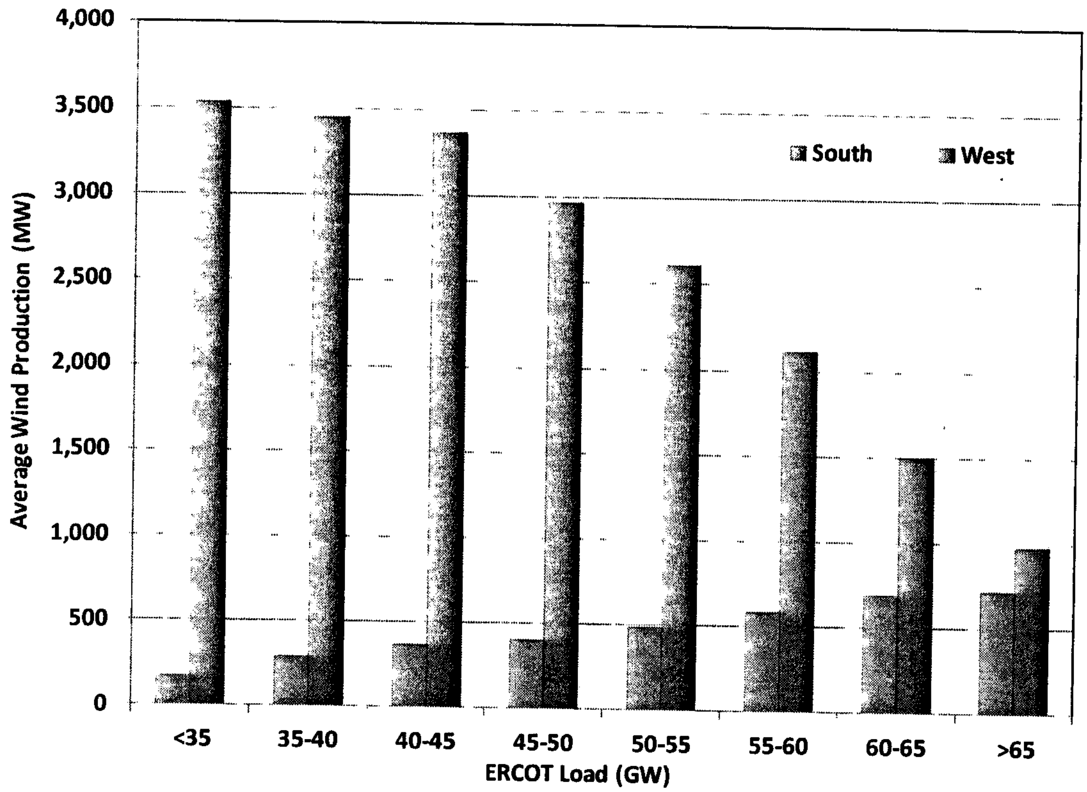


The average profile of wind production is negatively correlated with the load profile, with the highest wind production occurring during non-summer months, and predominately during off-peak hours. Figure 52 shows average wind production for each month in 2011, with the average production in each month shown separately in four hour blocks.[13]

[13] Figure 52 shows actual wind production, which was affected by curtailments at the higher production levels. Thus, the higher levels of actual wind production in Figure 52 are lower than the production levels that would have materialized absent transmission constraints.

Next we compare the differences in output for wind units located in the west and those located in the south.

Figure 53: Summer Wind Production vs. Load



In Figure 53 data is presented for the summer months of June through August, comparing the average output for wind generators located in the West and South zones across various load levels. It shows a strong negative relationship between West zone wind output and increasing load levels. It further shows that the output from wind generators located in the South zone is much more highly correlated with peak electricity demand.

Figure 54: Wind Production and Curtailment

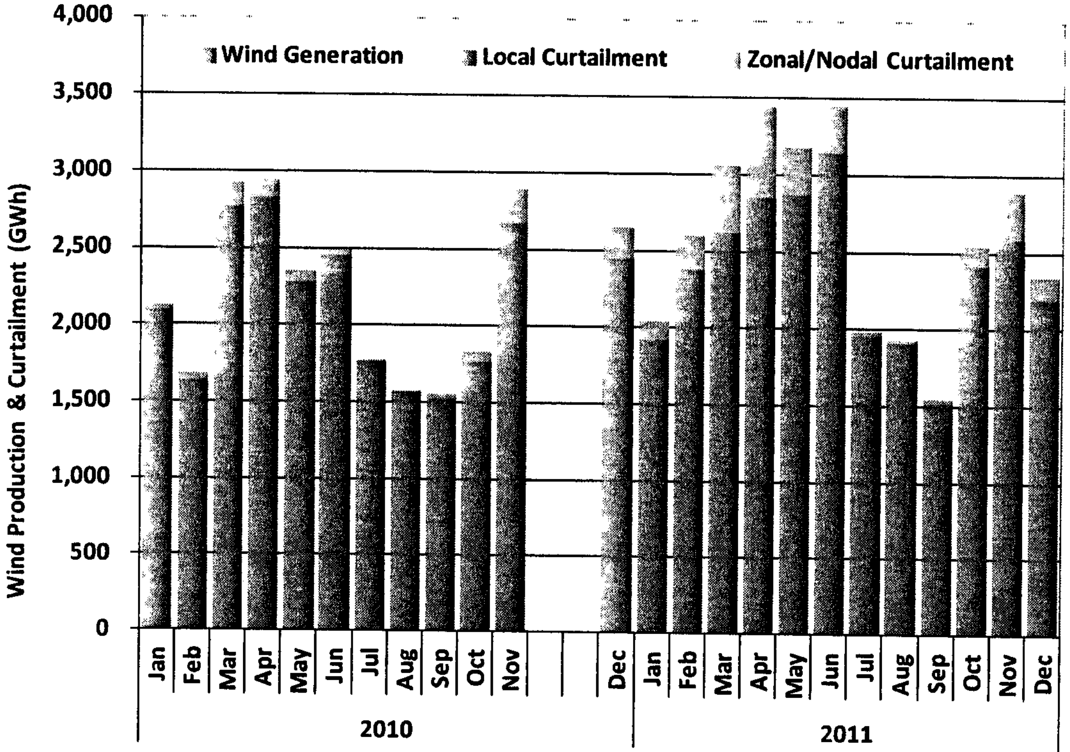
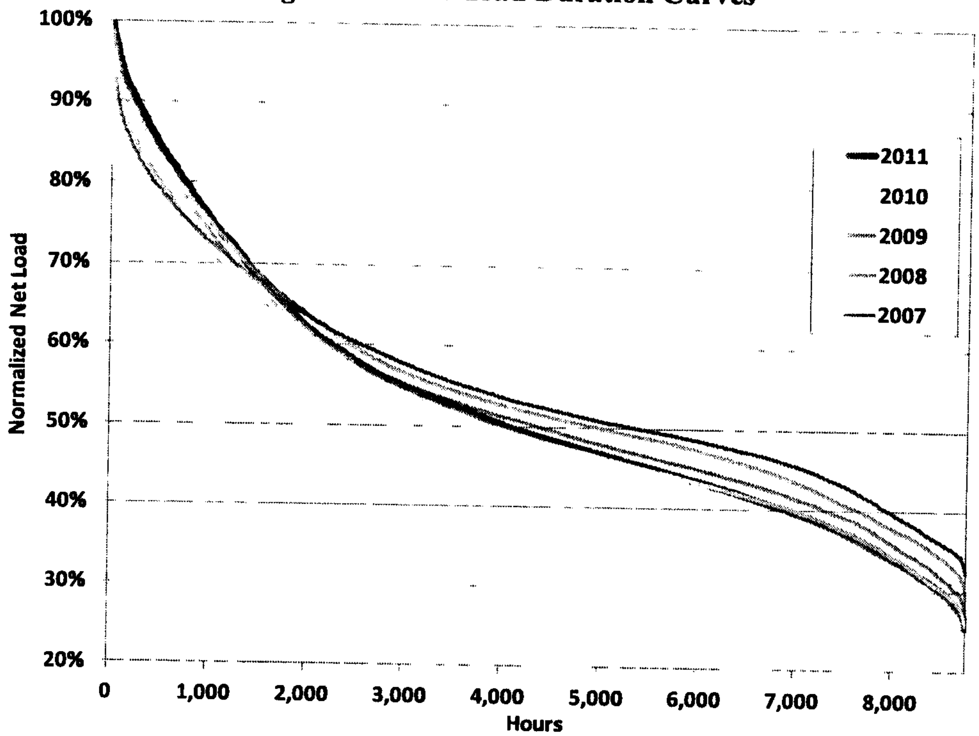


Figure 54 shows the wind production and local and zonal curtailment quantities for each month of 2010 and 2011. This figure reveals that the total quantity of curtailments for wind resources once again increased in 2011 when compared to 2010, even as actual production increased.

Increasing levels of wind resource in ERCOT also has important implications for the net load duration curve faced by the non-wind fleet of resources. Net load is defined as the system load minus wind production. Figure 55 shows the net load duration curves for 2008 through 2011, normalized as a percent of peak load.

Figure 55: Net Load Duration Curves



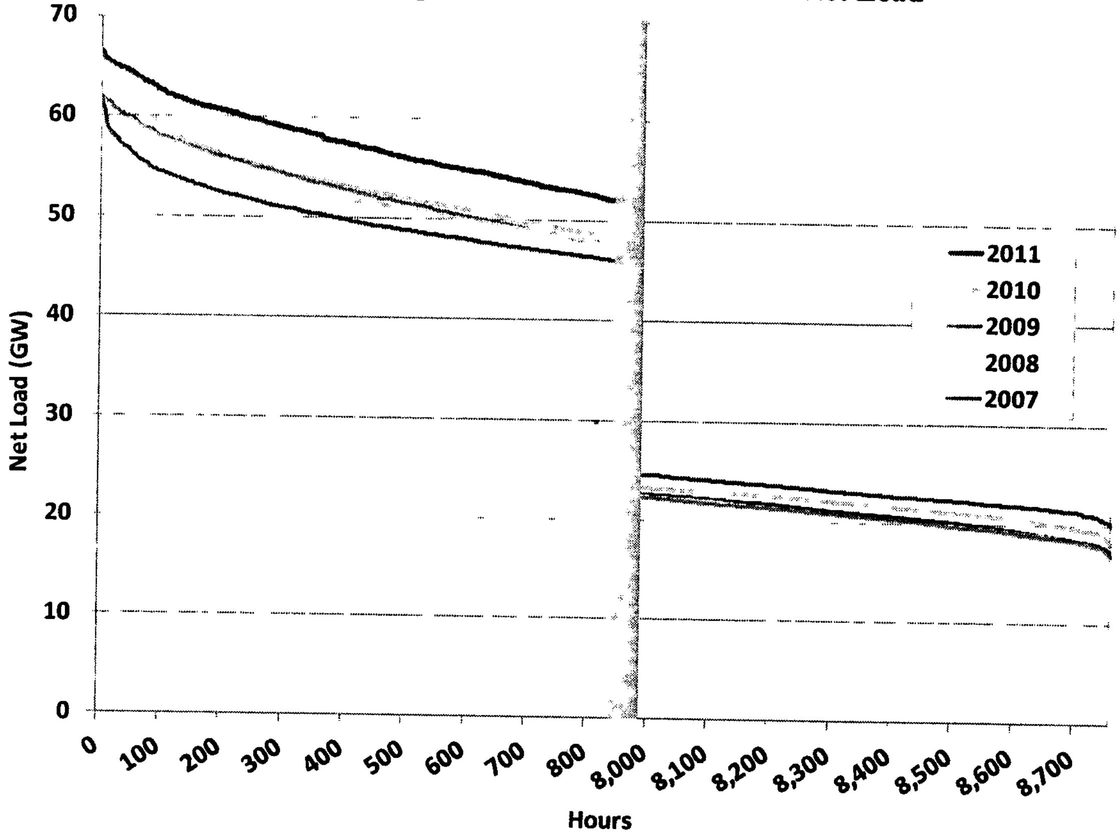
This figure shows the continued erosion of remaining energy available for non-wind units to serve during most hours of the year, with much less impact during the highest loads.

More than 90 percent of the wind resources in the ERCOT region are located in West Texas, and the wind profiles in this area are such that most of the wind production occurs during off-peak hours or other times of relatively low system demand. This profile results in only modest reductions of the net load relative to the actual load during the hours of highest demand, but much more significant reductions in the net load relative to the actual load in the other hours of the year. The trend shown from 2007 to 2011 in Figure 55 may continue with the addition of new wind resources and the reduction in the curtailment of existing wind resources after completion of new transmission facilities.

Focusing on the left side of the net load duration curve shown in Figure 56, the difference between peak net load and the 95th percentile of net load has been between 9.5 and 12.5 GW for the previous five years.

00104

Figure 56: Top and Bottom Ten Percent of Net Load



On the right side of the net load duration curve, the minimum net load has been 17 GW for the past three years, even with sizable growth in total annual load. This continues to put operational pressure on the nearly 25 GW of nuclear and coal fuel generation currently installed in ERCOT.

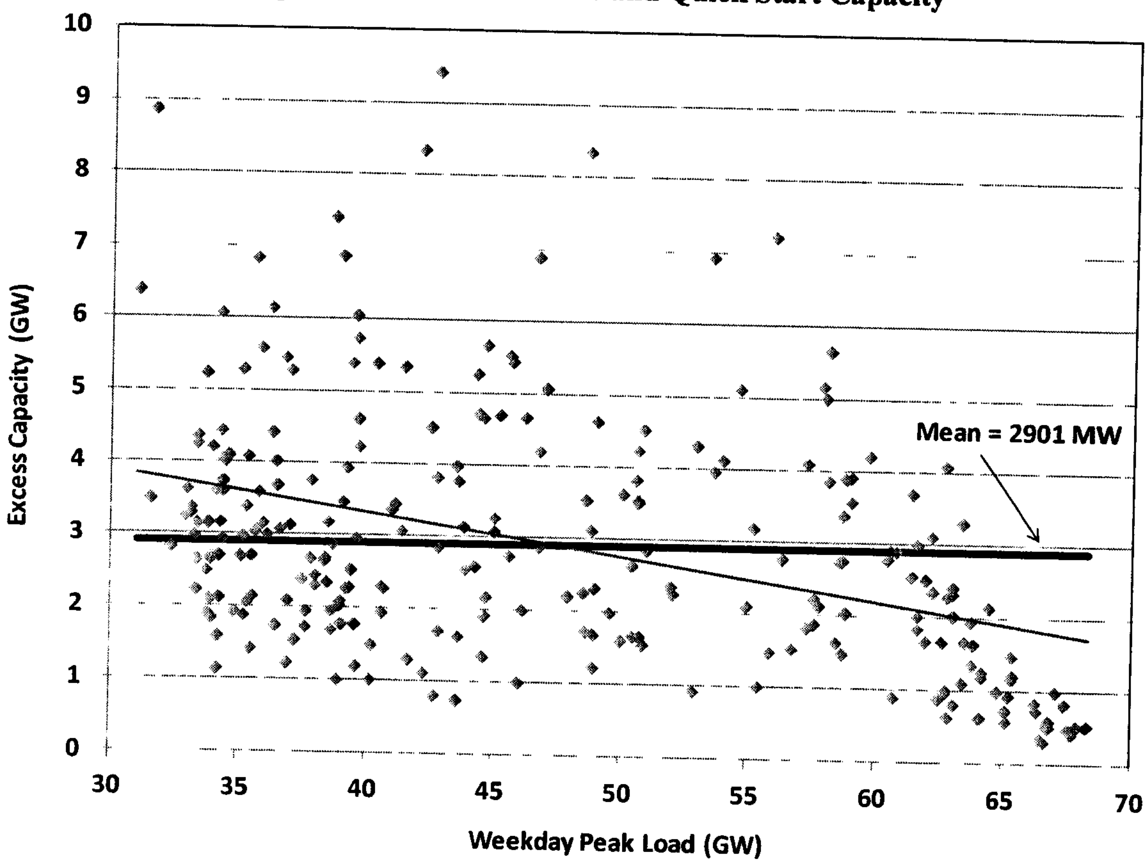
Thus, although the peak net load and reserve margin requirements are projected to continue to increase and create an increasing need for non-wind capacity to meet net load and reliability requirements, the non-wind fleet is expected to operate for fewer hours as wind penetration continues to increase. This outlook further reinforces the importance of efficient energy pricing during peak demand conditions and other times of system stress, particularly within the context of the ERCOT energy-only market design.

2. Daily Generator Commitments

One of the important characteristics of any electricity market is the extent to which it results in the efficient commitment of generating resources. Under-commitment can cause apparent shortages in real-time and inefficiently high energy prices while over-commitment can result in excessive start-up costs, uplift charges, and inefficiently-low energy prices.

This subsection evaluates the commitment patterns in ERCOT by examining the levels of excess capacity. Excess capacity is defined as the total capacity of online plus quick-start generators minus the demand for energy, responsive reserve, up regulation and non-spinning reserve provided from online capacity or quick-start units. To evaluate the commitment of resources in ERCOT, Figure 57 plots the excess capacity compared to peak load during 2011.

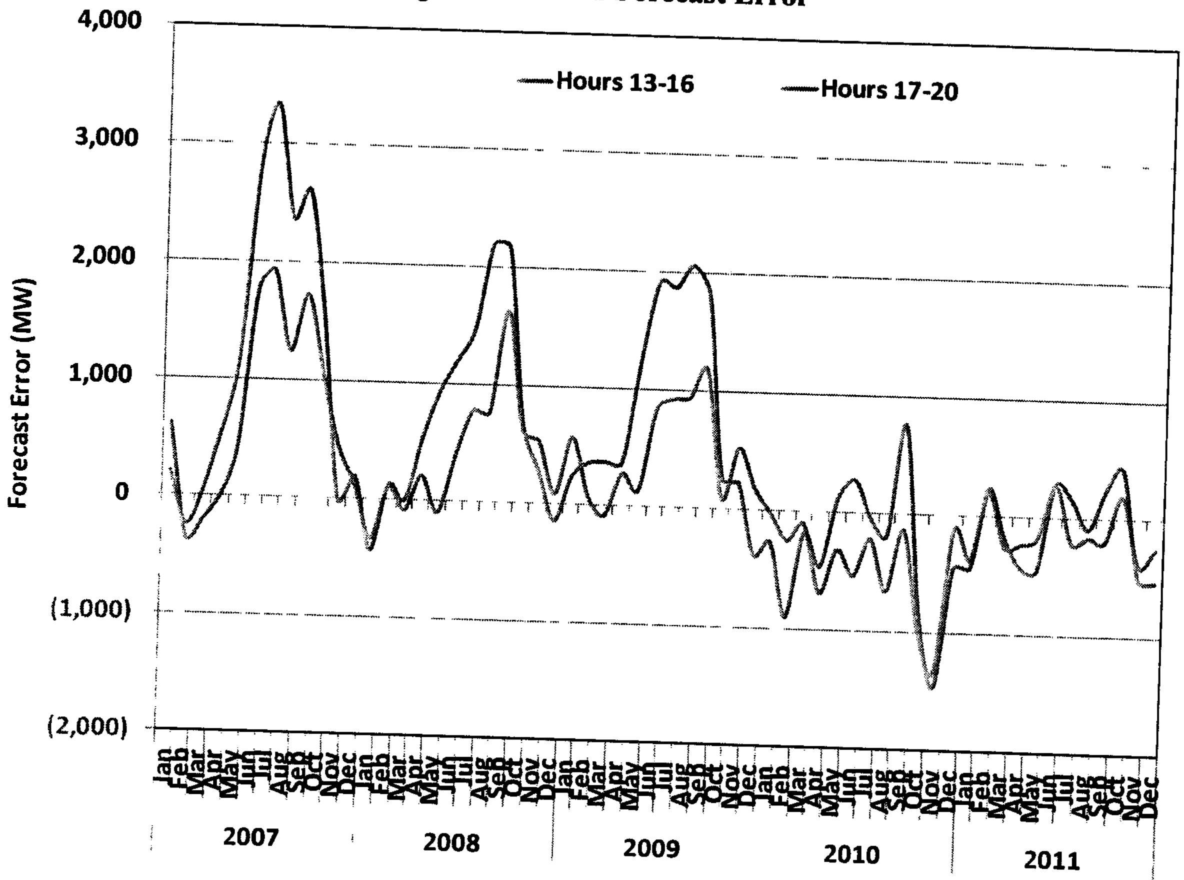
Figure 57: Excess On-Line and Quick Start Capacity



The figure shows the excess capacity in only the peak hour of each weekday because the largest generation commitment usually occurs at the peak hour. Hence, one would expect larger quantities of excess capacity in other hours. Figure 57 shows that the excess on-line capacity during daily peak hours on weekdays averaged 2,901 MW in 2011, which is approximately 7.6 percent of the average load in ERCOT. This is a decrease of 345 MW from the prior year, continuing the trend toward more efficient unit commitment across the ERCOT market. This improvement was expected with the nodal market implementation due to the introduction of a day-ahead energy market offering the opportunity for financially binding, centralized unit commitment decisions.

Even with the expected improvements in unit commitment coming from having a day-ahead market, if ERCOT's day-ahead load forecast continued to show significant bias toward over-forecasting peak load hours[14], we would expect to see over commitment of generation using non-market means.

Figure 58: Load Forecast Error

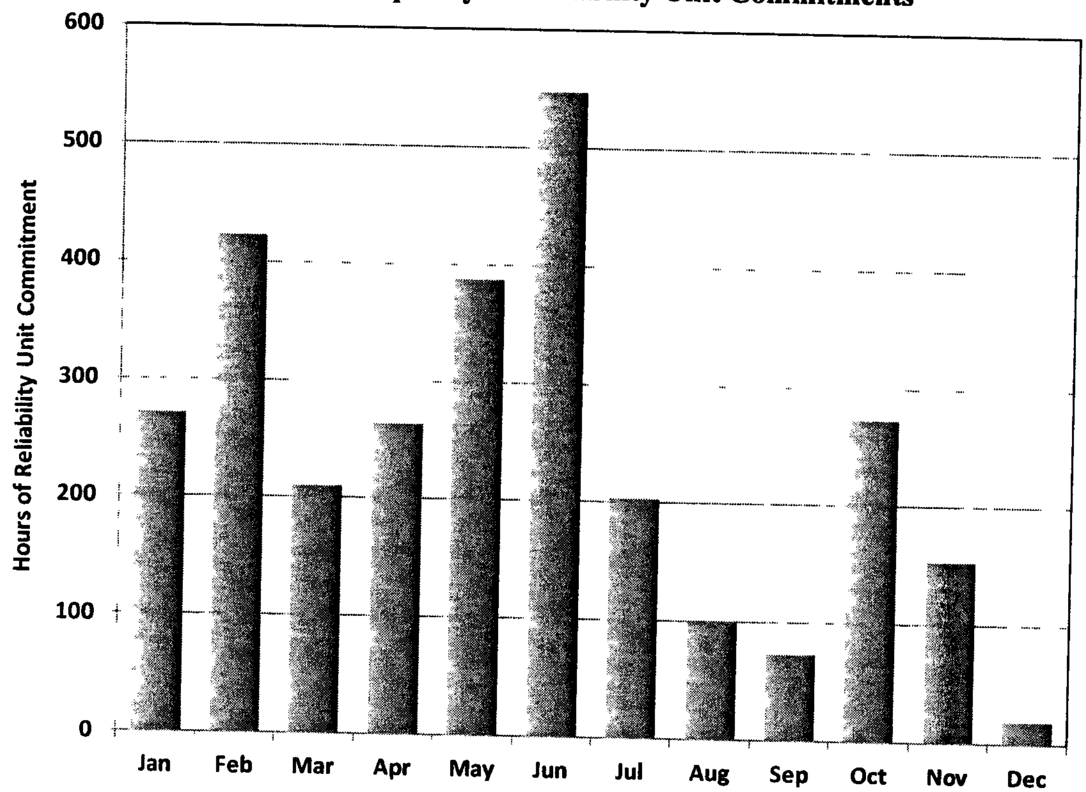


14 See 2010 ERCOT SOM report at 49-51 and 2009 ERCOT SOM report at 68-70.

From Figure 58 we can see the noticeable reduction in ERCOT's load forecast bias. This was due to a procedure change implemented two years ago. ERCOT now specifically identifies and subtracts out the forecast bias and procures additional non-spin capacity in an equal amount.

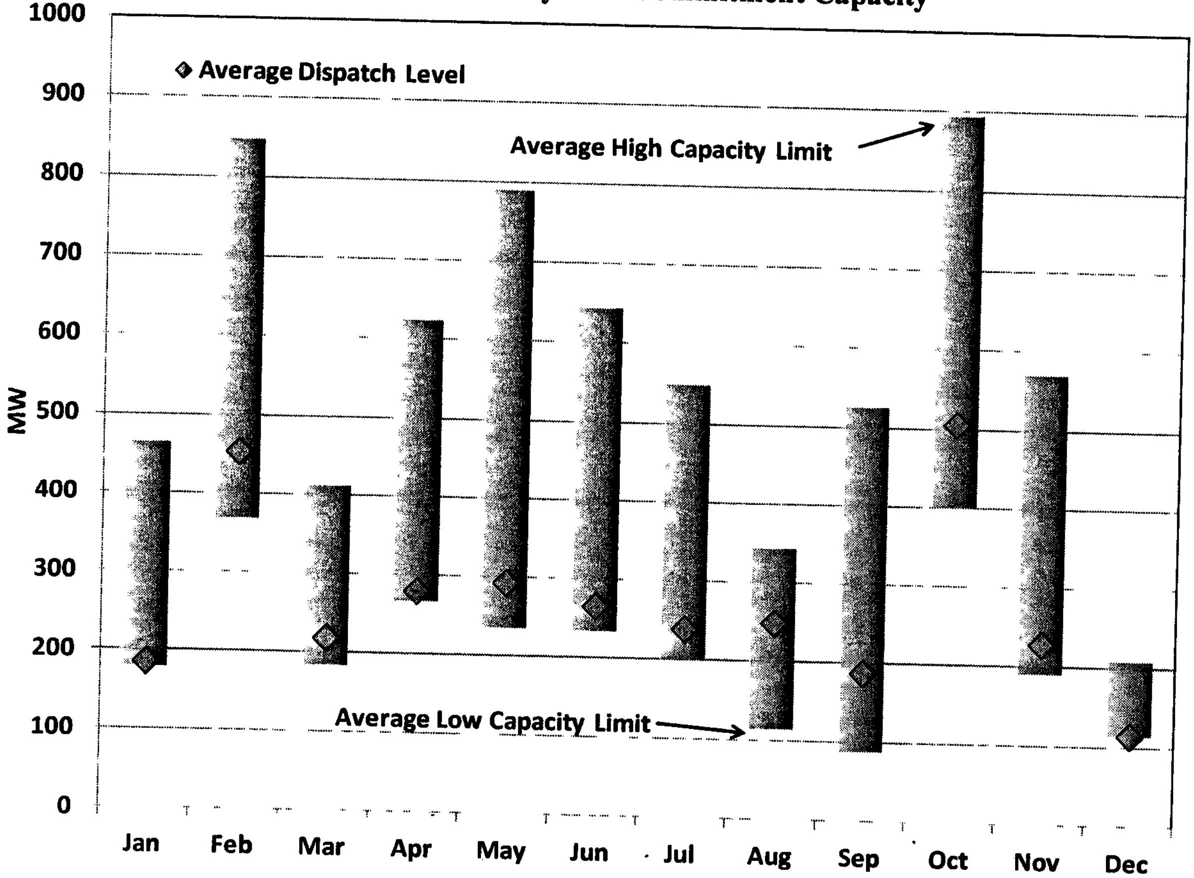
Once ERCOT assesses the unit commitments resulting from the Day-Ahead market, additional capacity commitments are made, if needed, using a reliability unit commitment process that executes both on a day-ahead and hour ahead basis. These additional unit commitments may be made for one of two reasons. Either, additional capacity is required to ensure forecasted total demand will be met, or a specific generator is required to resolve transmission congestion. Figure 59 summarizes, by month, the number of hours with units committed via the reliability unit commitment process.

Figure 59: Frequency of Reliability Unit Commitments



The next analysis compares the average dispatched output of the reliability committed units with their operational limits. In Figure 60 we can see that for most months when units are brought on line via the reliability unit commitment process, they are dispatched between 200 and 300 MWs.

Figure 60: Reliability Unit Commitment Capacity



The larger quantity of committed capacity in February was a result of ERCOT operator action taken to attempt to ensure overall capacity adequacy during both the extreme cold weather event that occurred early in the month, and a subsequent bout of cold weather that occurred one week later. The large amounts of reliability unit committed capacity in October were related to generator outages resulting in reactive power deficiencies in the Dallas-Fort Worth area.

V. RESOURCE ADEQUACY

One of the primary functions of the wholesale electricity market is to provide economic signals that will facilitate the investment needed to maintain a set of resources that are adequate to satisfy the system's demands and reliability needs. We evaluate these economic signals by estimating the "net revenue" new resources would receive from the markets. We begin this section by reviewing factors influencing whether sufficient investment in new resources can be expected to ensure resource adequacy in ERCOT in our analysis of net revenues. Next, our review of the effectiveness of the Public Utility Commission's scarcity pricing mechanism includes two recommendations for market design improvements. We conclude this section with a review of the contributions from demand response toward ensuring resource adequacy in ERCOT.

A. Net Revenue Analysis

Net revenue is the total revenue that can be earned by a new generating unit less its variable production costs. Put another way, it is the revenue in excess of short-run operating costs that is available to recover a unit's fixed and capital costs, including a return on the investment. Net revenues from the energy and ancillary services markets together provide the economic signals that inform suppliers' decisions to invest in new generation or retire existing generation. In a long-run equilibrium, markets should provide sufficient net revenue to allow an investor to receive a return of, and on an investment in a new generating unit that is needed. In the short-run, if the net short-run revenues produced by the market are not sufficient to justify entry, then one or more of these conditions exist:

- New capacity is not needed because there is sufficient generation already available;
- Load levels, and thus energy prices, are temporarily low due to mild weather or economic conditions;
- Market rules or operational practices are causing revenues to be reduced inefficiently; or
- Market rules are not sufficiently linked in short-term operations to ensure long-term resource adequacy requirements are met.

Likewise, the opposite would be true if the markets provide excessive net revenues in the short-run. The persistence of excessive net revenues in the presence of a capacity surplus is an indication of competitive issues or market design flaws. In this section, we analyze the net revenues that would have been received by various types of generators in each zone.

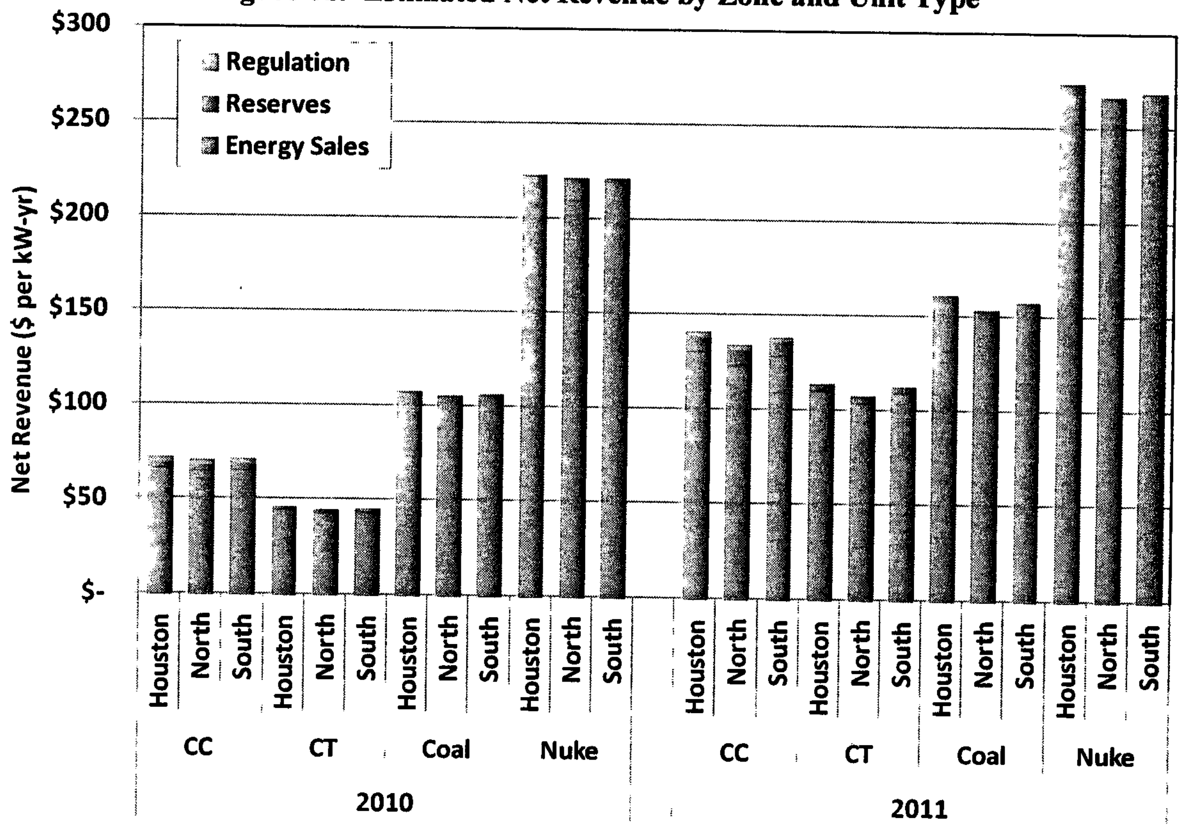
Figure 61 shows the results of the net revenue analysis for four types of hypothetical new units in 2010 and 2011. These are: (a) natural gas fueled combined-cycle, (b) natural gas fueled combustion turbine, (c) coal fueled generator, and (d) a nuclear unit. For the gas-fired technologies, net revenue is calculated by assuming the unit will produce energy in any hour for which it is profitable and by assuming it will be available to sell reserves and regulation in other hours that it is available (*i.e.*, when it is not experiencing a planned or forced outage). For coal and nuclear technologies, net revenue is calculated by assuming that the unit will produce at full output.

Although most suppliers are likely to receive the bulk of their revenues through bilateral contracts, the spot prices produced in the real-time energy market should drive the bilateral energy prices over time and are appropriate to use for this evaluation. For purposes of this analysis, heat rates of 7 MMBtu per MWh for a combined cycle unit, 10.5 MMBtu per MWh for a combustion turbine, and 9.5 MMBtu per MWh for a new coal unit were assumed. Variable operating and maintenance costs of $4 per MWh for the gas units and $5 per MWh for the coal unit and fuel and variable operating and maintenance costs of $8 per MWh for the nuclear unit were assumed. A total outage rate (planned and forced) of 10 percent was assumed for each technology.

The energy net revenues are computed based on the generation weighted settlement point prices from the real-time energy market. Weighting the energy values in this way masks what may be very high locational values for a specific generator location. Some generators may also receive uplift payments because of their specific reliability contributions, either as a reliability must run, or through the reliability unit commitment. This source of revenue is not considered in this analysis. The analysis also includes simplifying assumptions that can lead to over-estimates of the profitability of operating in the wholesale market. The following factors are not explicitly accounted for in the net revenue analysis: (i) start-up costs, which can be significant; and (ii) minimum running times and ramp restriction, which can prevent the natural gas generators from profiting during brief price spikes. Despite these limitations, the net revenue analysis provides a useful summary of signals for investment in the wholesale market.

Figure 61 shows that the net revenue for every generation technology type increased in 2011 compared to each zone in 2010. Prior to 2005, net revenues were well below the levels necessary to justify new investment in coal and nuclear generation. However, high natural gas prices through 2008 allowed energy prices to remain at levels high enough to support new entry for these technologies. The production costs of coal and nuclear units did not change significantly over this period, leading to a dramatic rise in net revenues. Conditions have now changed with the much lower natural gas prices experienced through 2011.

Figure 61: Estimated Net Revenue by Zone and Unit Type



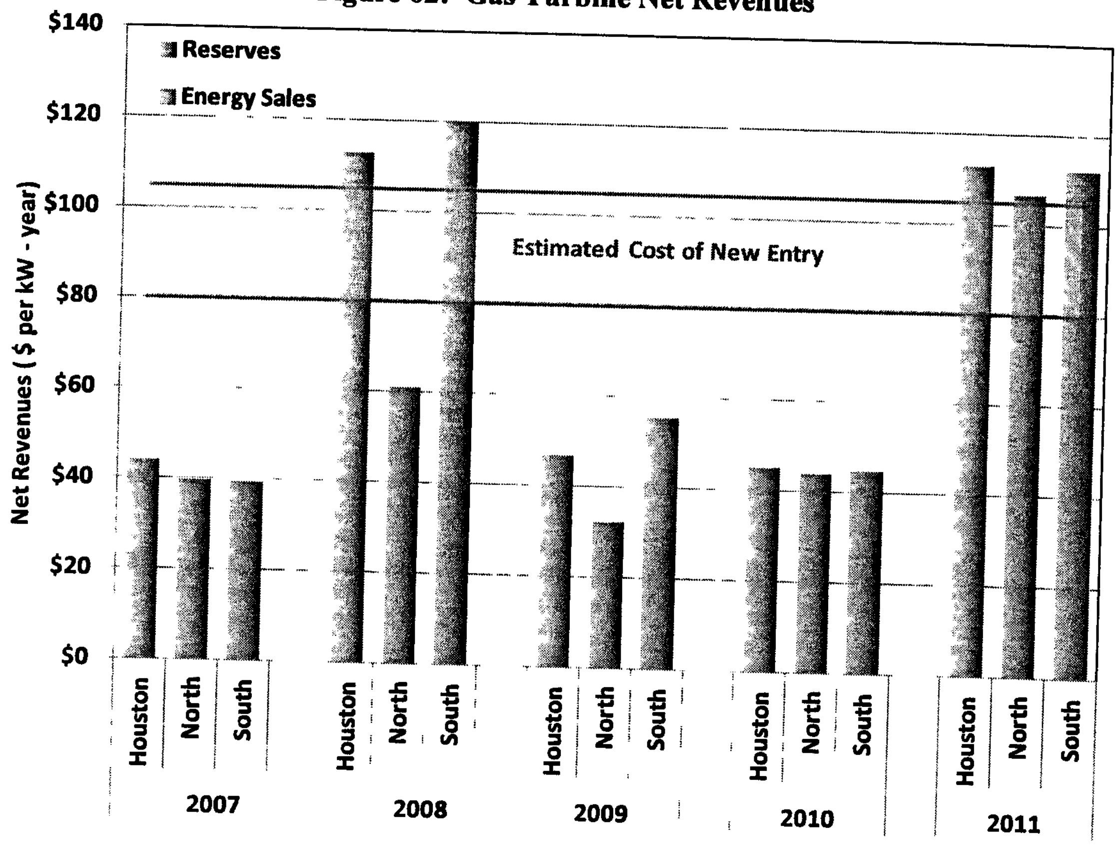
The estimated net revenue for both a new coal or a nuclear unit in ERCOT were well below the levels required to support new entry, despite the relatively frequent shortages in 2011.

- For a new coal unit, the estimated net revenue requirement is approximately $210 to $270 per kW-year. The estimated net revenue in 2011 for a new coal unit was approximately $150 to $160 per kW-year.

- For a new nuclear unit, the estimated net revenue requirement is approximately $280 to $390 per kW-year. The estimated net revenue in 2011 for a new nuclear unit was approximately $270 per kW-year.

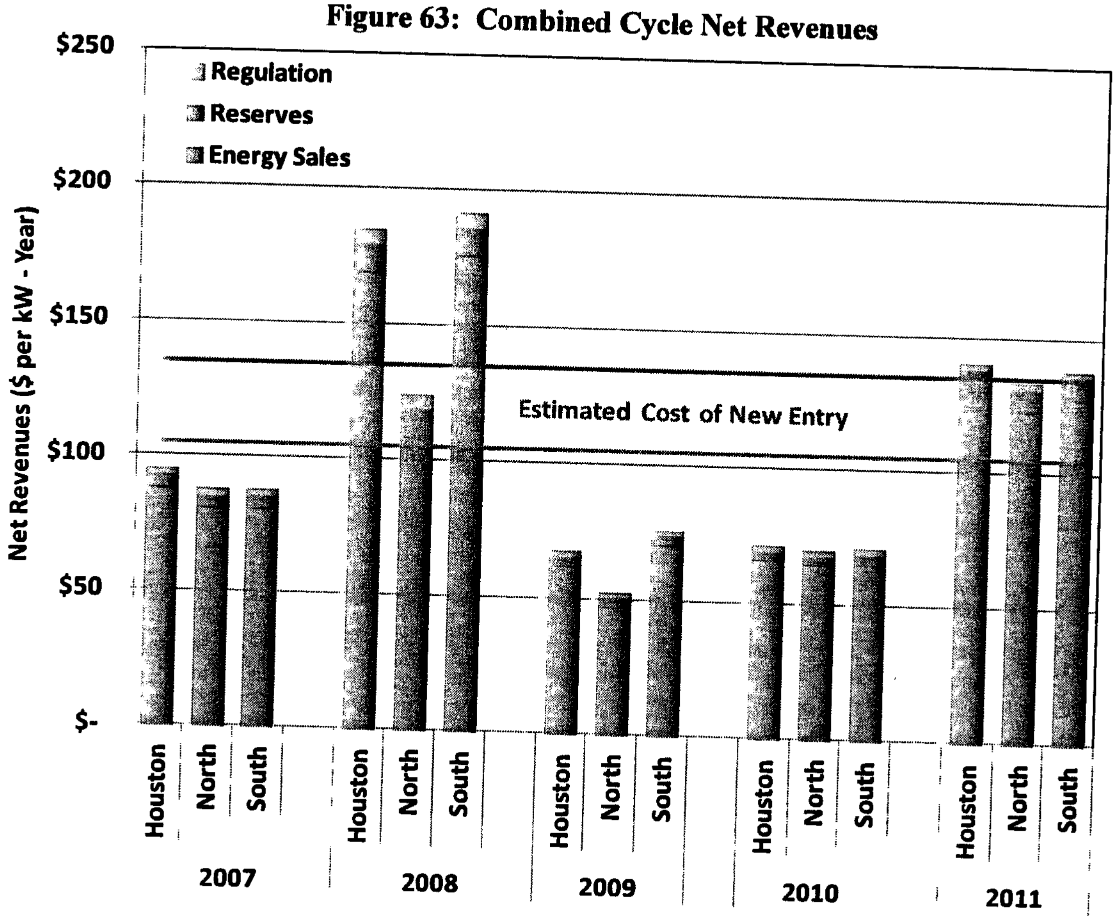
Based on our estimates of investment costs for new units, the net revenue required to satisfy the annual fixed costs (including capital carrying costs) of a new gas turbine unit ranges from $80 to $105 per kW-year. The estimated net revenue in 2011 for a new gas turbine ranged from $107 per kW-year in the North zone to $113 per kW-year in the Houston zone. Figure 62 shows that 2011 is the first time since 2008 that net revenues have been sufficient to support new gas turbine generation.

Figure 62: Gas Turbine Net Revenues



For a new combined cycle unit, the estimated net revenue requirement is approximately $105 to $135 per kW-year. The estimated net revenue in 2011 for a new combined cycle unit ranged from $133 per kW-year in the North to $140 per kW-year in Houston. From Figure 63 we see that 2011 was the first time since 2008 that net revenues have been sufficient to support new combined cycle generation in ERCOT.

Figure 63: Combined Cycle Net Revenues



Even though net revenues for the Houston and South zone in 2008 may have appeared to be sufficient to support new gas fueled generation, it was actually extremely inefficient transmission congestion management and inefficient pricing mechanisms associated with the deployment of non-spinning reserves which led to high prices and resulting higher than warranted net revenues. Discounting the effect that the 2008 results would have had on forward price signals, we find that 2011 is the first time in five years that net revenues have been sufficient to support either new gas turbine or combined cycle generation.

To provide additional context for the net revenue results presented in this section, we also compared the net revenue for two types of natural gas-fired technologies in the ERCOT market with the net revenue that those technologies could expect in other wholesale markets with centrally cleared capacity markets. The technologies are differentiated by their assumed heat rate; 7,000 MMBtu/MWh for combined cycle and 10,500 MMBtu/MWh for simple-cycle combustion turbine.

Figure 64: Comparison of Net Revenue of Gas-Fired Generation between Markets

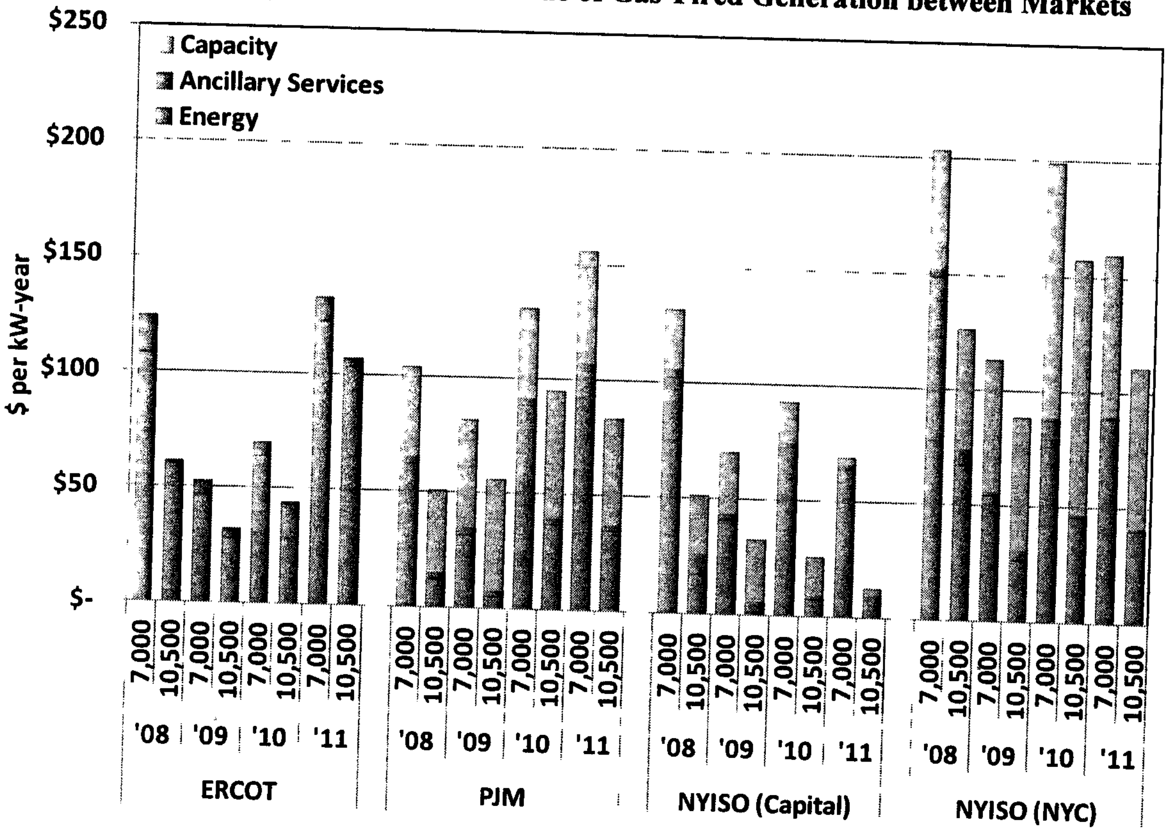


Figure 64 compares estimates of net revenue for the ERCOT North Zone, PJM, and two locations within the New York ISO. The figure includes estimates of net revenue from energy, reserves and regulation, and capacity. ERCOT does not have a capacity market, and thus, does not have any net revenue from capacity sales. Figure 64 shows that net revenues decreased from 2010 to 2011 for both technologies in NY ISO. In PJM, net revenue decreased for combustion turbines and increased slightly for combined cycle technology. In the figure above, net revenues are calculated for central locations. However, there are load pockets within each market where net revenue and the cost of new investment may be higher. Thus, even if new investment is not generally profitable in a market, it may be economic in certain areas. Finally, resource investments are driven primarily by forward price expectations, so historical net revenue analyses do not provide a complete picture of the future pricing expectations that will spur new investment.

B. Effectiveness of the Scarcity Pricing Mechanism

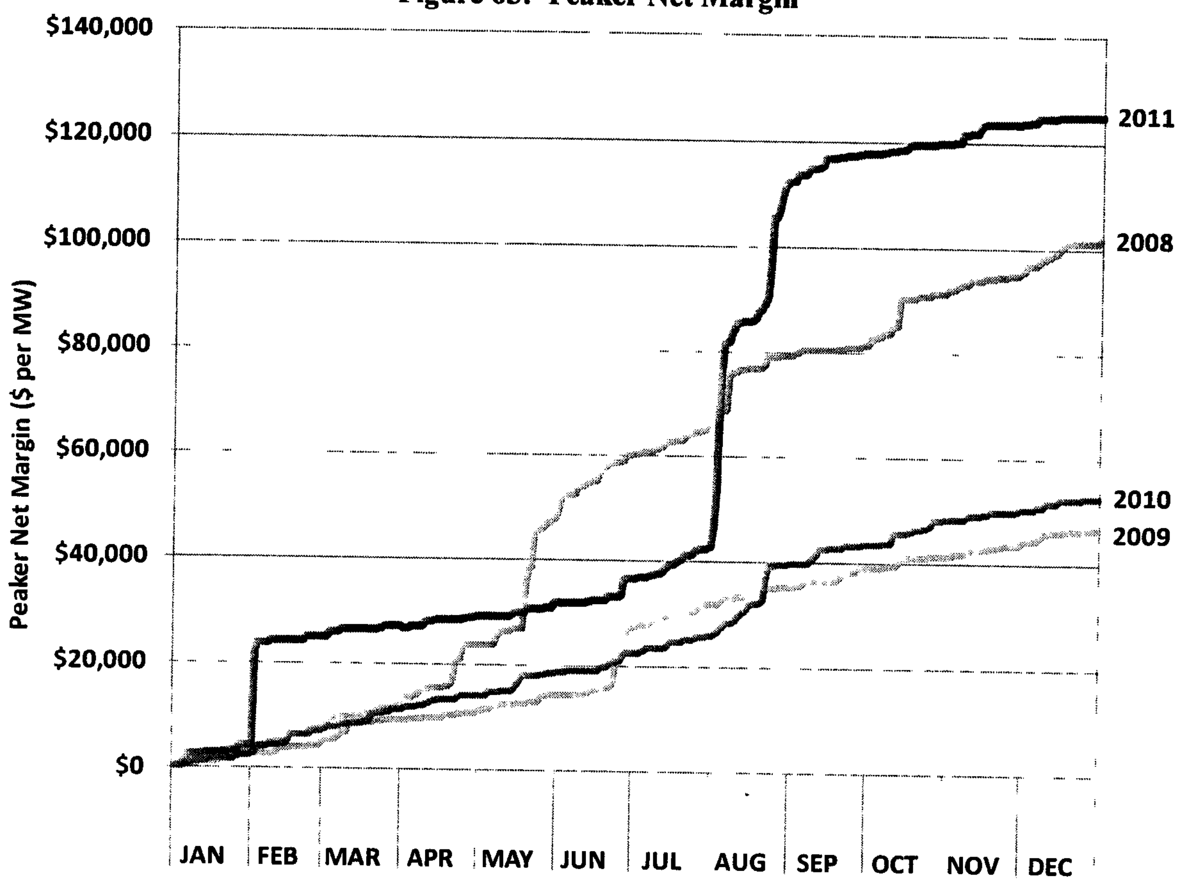
The Public Utility Commission of Texas ("PUCT") adopted rules in 2006 that define the parameters of an energy-only market. These rules include a Scarcity Pricing Mechanism ("SPM") that relaxed the existing system-wide offer cap by increasing it in multiple steps until it reached $3,000 per MWh shortly after the implementation of the nodal market. Additionally, market participants controlling less than five percent of the capacity in ERCOT by definition do not possess ERCOT-wide market power under the PUCT rules. Hence, these participants can submit very high-priced offers that, per the PUCT rule, will not be deemed to be an exercise of market power. However, because of the competition faced by the small market participants, the quantity offered at such high prices, if any, is very small.

PUCT SUBST. R. 25.505 provides that the IMM may conduct an annual review of the effectiveness of the SPM. This subsection provides an assessment of the effectiveness of the SPM in 2011 under ERCOT's energy-only market structure. In markets with a long-term capacity market, fixed capacity payments are made to resources across the entire year independent of the relationship between real-time supply and demand. The objective of the energy-only market design is to allow energy prices to rise significantly higher at times when the available supply is insufficient to simultaneously meet both energy and minimum operating reserve requirements. Allowing energy prices to rise during shortages mirrors the outcome expected if loads were able to actively specify the quantity of electricity they wanted and the price they would be willing to pay. The energy-only market design relies upon these relatively infrequent occurrences of high prices to provide the appropriate price signal for demand response and new investment when required. The expectation of competitive energy market outcomes is no different in energy-only than in markets that include a capacity market. However, capacity markets are designed to ensure a specified planning reserve margin, which may be higher than an energy-only market would achieve. Under this condition the higher planning reserve margin will serve to reduce the frequency of shortages in the energy market.

The SPM includes a provision termed the Peaker Net Margin ("PNM") that is designed to measure the annual net revenue of a hypothetical peaking unit. Under the current rule, if the PNM for a year reaches a cumulative total of $175,000 per MW,[15] the system-wide offer cap is then reduced to the higher of $500 per MWh or 50 times the daily gas price index.

Figure 65 shows the cumulative PNM results for each year from 2008 through 2011 and shows that PNM in 2011 was higher than it has ever been.

Figure 65: Peaker Net Margin



15 The proxy combustion turbine in the Peaker Net Margin calculation assumes a heat rate of 10 MMBtu per MWh and includes no other variable operating costs or startup costs.

As previously described, the net revenue required to satisfy the annual fixed costs (including capital carrying costs) of a new gas turbine unit ranges from $80,000 to $105,000 per MW-year. Thus, as shown in Figure 65 and consistent with the previous findings in this section relating to net revenue, the PNM reached the level sufficient for new entry in 2011. In 2008 peaker net margin and net revenue values were also sufficient to support new peaker entry. However, a significant portion of the net revenue increase in 2008 was associated with extremely inefficient transmission congestion management and inefficient pricing mechanisms associated with the deployment of non-spinning reserves.[16] With these issues addressed in the zonal market, the peaker net margin dropped substantially in 2009 and 2010.

As a general principle, competitive and efficient market prices should be consistent with the marginal cost of the marginal action taken to satisfy the market's demand. In the vast majority of hours, the marginal cost of the marginal action is associated with the dispatch of the last generator required to meet demand. It is appropriate and efficient in these hours for this generator to "set the price." However, this is not true under shortage conditions. When the system is in shortage, the demand for energy and minimum operating reserves cannot be satisfied with the available resources, which will cause the system operator to take one or more of the following actions:

- Sacrifice a portion of the operating reserves by dispatching them for energy;
- Voluntarily curtail load through emergency demand response programs;
- Curtail exports or make emergency imports; or
- Involuntarily curtail load.

A market design that adheres to the pricing principles stated above will set prices that reflect each of these actions. When there is a shortage of supply in the market, the marginal action taken by the system operator is generally to not satisfy operating reserves requirements (*i.e.*, dispatching reserves for energy). Diminished operating reserves results in diminished reliability, which has a real cost to electricity consumers. In this case, the value of the foregone reserves – which is much higher than the marginal cost of the most expensive online generator – should be

[16] See 2008 ERCOT SOM Report at 81-87.

reflected in energy prices to achieve efficient economic signals governing investment in generation, demand response and transmission.

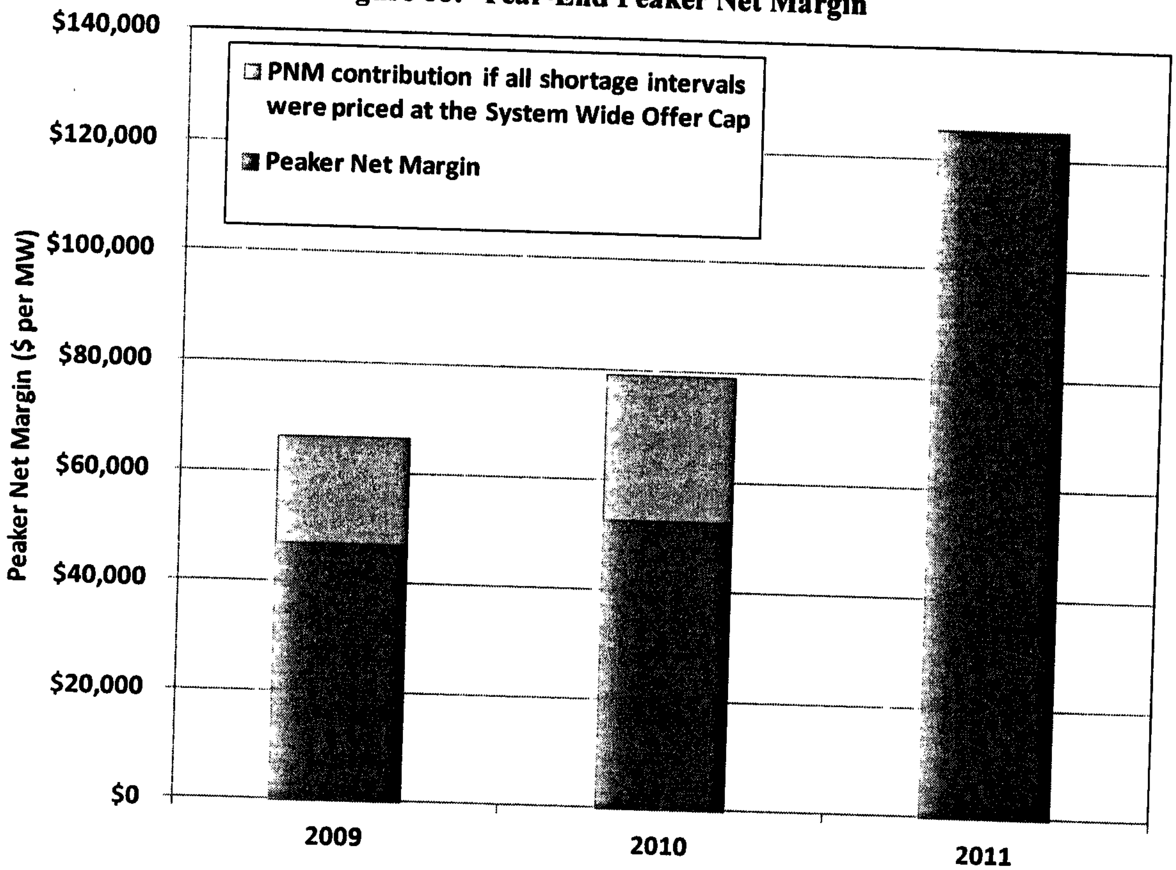
Under the PUCT rules governing the energy-only market, high-priced offers submitted by small market participants are deemed to not be an exercise of ERCOT-wide market power. Under the mechanics of the zonal market design, even during shortage conditions, prices were always set by generator offers. As discussed in previous annual reports[17], relying on owners of small amounts of generation to submit offers that adequately reflect the value of diminished reliability to loads was problematic. This aspect of the zonal market design resulted in energy prices during scarcity conditions that did not reflect the full value of continued reliability of service.

With the implementation of the nodal market, more reliable and efficient shortage pricing has been achieved by establishing pricing rules that recognize when operating reserve shortages exist and allowing energy prices to rise automatically. This approach is more reliable because it is not dependent upon the submission of high-priced offers by small market participants to be effective. It is also more efficient during the vast majority of time in which shortage conditions do not exist because it is not necessary for market participants to effectively withhold lower cost resources by offering relatively small quantities at prices dramatically higher than their marginal cost. At times when there is insufficient capacity available to meet both energy and minimum operating reserve requirements, all available capacity will be dispatched and the clearing price will rise in a predetermined manner to a maximum of the system-wide offer cap. During 2011 prices were at the system-wide offer cap in dispatch intervals which totaled 28.5 hours, or 0.33 percent of the total hours.

Figure 66 presents the results of our analysis of how high PNM would have been under the zonal market design had all shortage intervals been priced at the system-wide offer cap. From this figure we can see the quantity of missed investment signals in 2009 and 2010 making the increase in 2011 seem not quite as dramatic.

[17] See 2010 ERCOT SOM Report at 52-56 and 2009 ERCOT SOM Report at 71-75

Figure 66: Year-End Peaker Net Margin



Although the nodal market implementation brought about more reliable and efficient shortage pricing there remain aspects of the ERCOT real-time energy pricing that can be improved. As discussed later in Section V.C, Demand Response Capability, prices during the deployment of load resources do not reflect the value of reduced reliability which occurs when responsive reserves have been converted to energy.

Similarly, when non-spinning reserves were deployed (converted to energy) prices rarely reflected the marginal cost of the action being taken. Non-spinning reserves are provided primarily by off-line natural gas fired combustion turbines capable of starting in 30 minutes or less. Although the implementation of the nodal market has significantly increased market efficiencies in a number of areas, including the move to a five minute rather than 15 minute energy dispatch, it lacks an efficient economic commitment mechanism for resources such as offline gas turbines and other resources that are not immediately dispatchable in the five minute energy dispatch. This led to prices that were inefficiently low because they did not represent the costs associated with starting and running the gas turbines that were being deployed to meet demand.

After much discussion by market participants, ERCOT staff and PUCT Commissioners, NPRRs 426 and 428 were approved by the ERCOT Board in December for implementation in early 2012. These changes implemented certain requirements for providers of non-spinning reserve to make that capacity available to ERCOT's dispatch software, subject to certain price floors.[18] Providers are now able to specify the price at which they are willing to convert their non-spinning reserve capacity to energy. Further, ERCOT will use this price information to determine which non-spin units to deploy. Real-time energy price formation will be improved, but the current mechanism is sub-optimal from a reliability and efficiency perspective. We continue to recommend that ERCOT develop a mechanism that will rationally commit generation and load resources that can start or curtail within 30 minutes.

This deficiency in ERCOT's nodal market design should be addressed by implementing an additional "look ahead" dispatch functionality for the real-time energy market to produce energy and ancillary services commitment and dispatch results that are co-optimized and recognize anticipated changes in system demands.[19] This additional functionality represents a major change to ERCOT systems; one we recommend together with improved pricing provisions that will allow offers from load resources to set prices if they are required to meet system demand.

As a first step, starting in June 2012 ERCOT will calculate and publish indicative pricing for the next hour, using an hour-long optimization instead of a five-minute optimization. These calculated prices and dispatch levels will initially be non-binding, although it is anticipated that these will become binding as the look ahead dispatch functionality progresses in future phases. This first step is expected to be most valuable to parties that would like to reduce their demand to avoid consuming during high priced intervals.

[18] NPRR 427 provided similar modifications related to the pricing of energy during responsive reserve and up regulation deployments.

[19] See Direct Testimony of David B. Patton, PUCT Docket No. 31540 at pages 35-41.

Although the IMM has raised concerns with the multi-period optimization that ERCOT intends to implement as part of its comprehensive look ahead dispatch approach, we support the first step of publishing indicative prices. Once the software architecture is in place to develop and publish indicative prices, we anticipate evaluating the effect the chosen optimization parameters would have if they were to produce binding dispatch instructions and prices.

An effective look ahead dispatch functionality should also reduce the price dampening effects of energy produced by units operating below their low sustainable operating limit. Although alternatives have been suggested to address this issue in a standalone manner, we believe the better approach will be to develop a comprehensive look ahead dispatch solution.

Additional deficiencies with how the nodal market ensures resource adequacy all relate to ensuring that when all available market-based supply has been exhausted that real-time prices are high enough to reflect the value to demand of continuing to be served.

To bolster the amount of installed capacity available, ERCOT entered into reliability must-run contracts to bring four mothballed generators back to service during the summer's extreme high load conditions. In order for this capacity to have the least amount of impact on market based price formation, the energy provided from these units was offered for dispatch at the system wide offer cap of $3,000 per MWh. These steps were all taken as a result of ERCOT staff action, with oversight from the PUCT and input from the IMM, but without the benefit of specific Protocol language. In late 2011 Protocol revision requests were submitted to clarify the process by which ERCOT can return mothball units to service. Additional protocol revisions have been submitted to require all energy from reliability must-run units and units brought on-line via the reliability unit commitment process to be offered at the system wide offer cap. In combination these changes will serve to ensure that market provided supply offers are exhausted before utilizing any non-market capacity.

Expectations about both the magnitude of the energy price during shortage conditions and the frequency of shortage conditions are what will attract new investment in an energy-only market. In other words, the higher the price during shortage conditions, the fewer shortage conditions that are required to provide the investment signal, and vice versa. As we have continually observed since the SPM was first put in place in late 2006,[20] the magnitude of price expectations is determined by the market rules established by the PUCT, and it is yet to be seen whether the frequency of shortage conditions over time will be sufficient to produce market equilibrium that satisfies the current reliability requirement of maintaining a 13.75 percent planning reserve.

Proceedings are currently underway at the PUCT to review both the magnitudes of prices during operating reserve shortage conditions and the current reliability requirement; specifically whether the assumptions relating to the planning reserve margin calculation are appropriate for the ERCOT energy-only market, and whether the resulting value is to be treated as a target or a minimum requirement. Upon clarification of these issues, policy options will be considered to ensure that the market design elements are properly linked to the chosen resource adequacy objectives.

As extreme as the weather and resulting load was in 2011, the total number of dispatch intervals with system-wide energy prices at the offer cap amounted to 28.5 hours. Although net revenues were sufficient for new gas generation, they were not overly so. Even with the improvements discussed, pricing during shortage intervals may need to be even higher to ensure that investments in new supply and demand resources result in maintaining the desired minimum installed reserve margin.

C. Demand Response Capability

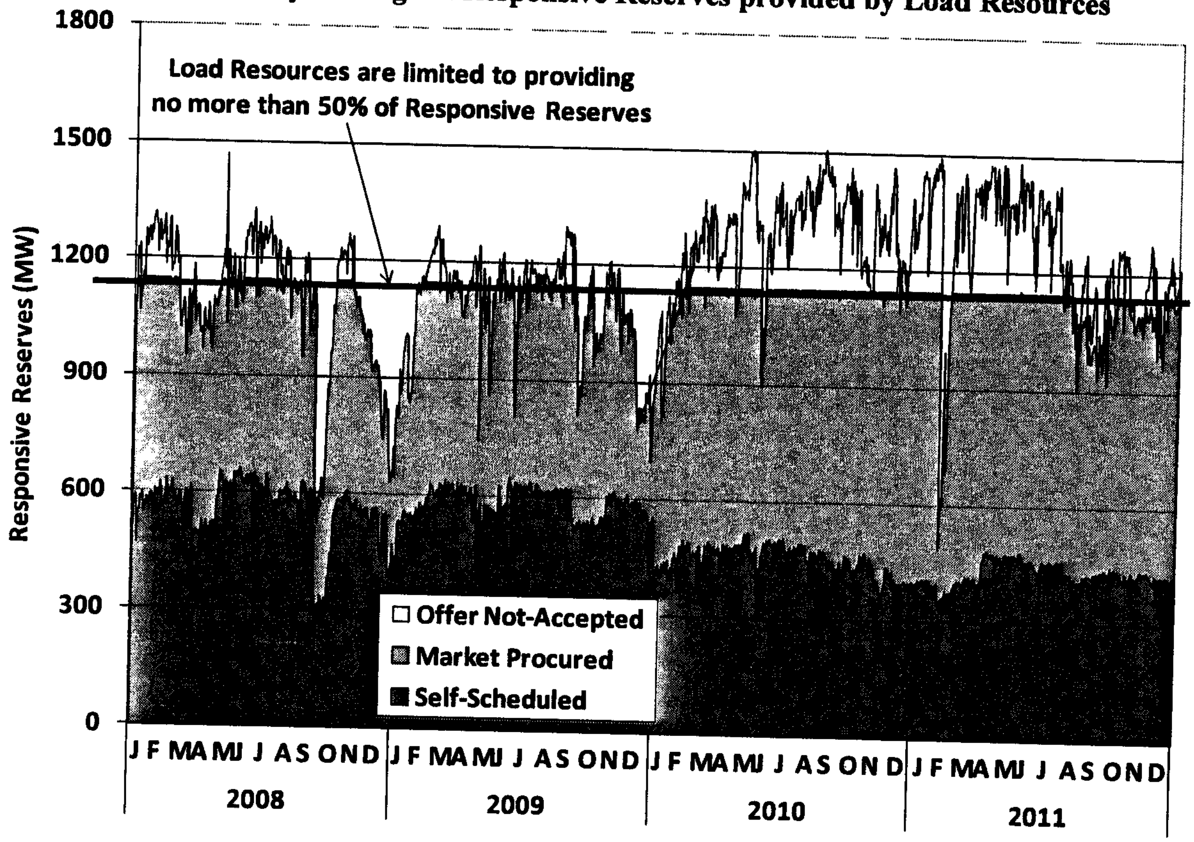
Demand response is a term that broadly refers to actions that can be taken by end users of electricity to reduce load in response to instructions from ERCOT or in response to certain market or system conditions. The ERCOT market allows participants with demand-response capability to provide energy and reserves in a manner similar to a generating resource. The ERCOT Protocols allow for loads to actively participate in the ERCOT administered markets as Load Resources. Additionally, loads may participate passively in the market by simply adjusting consumption in response to observed prices. Unlike active participation in ERCOT administered markets, passive demand response is not directly tracked by ERCOT.

[20] See 2010 ERCOT SOM Report at 48, 2009 ERCOT SOM Report at 66, 2008 ERCOT SOM at 65, and 2007 ERCOT SOM Report at 46.

ERCOT allows qualified load resources to offer responsive reserves and non-spinning reserves into the day-ahead ancillary services markets. Those providing responsive reserves must have high set under-frequency relay equipment, which enables the load to be automatically tripped when the frequency falls below 59.7 Hz, which will typically occur only a few times in each year. Deployments of non-spinning reserves occur much more frequently. To date, load resources have shown a clear preference for providing responsive reserve service.

As of December 2011, approximately 2,400 MW of capability were qualified as Load Resources. Figure 67 shows the amount of responsive reserves provided from load resources on a daily basis in 2011.

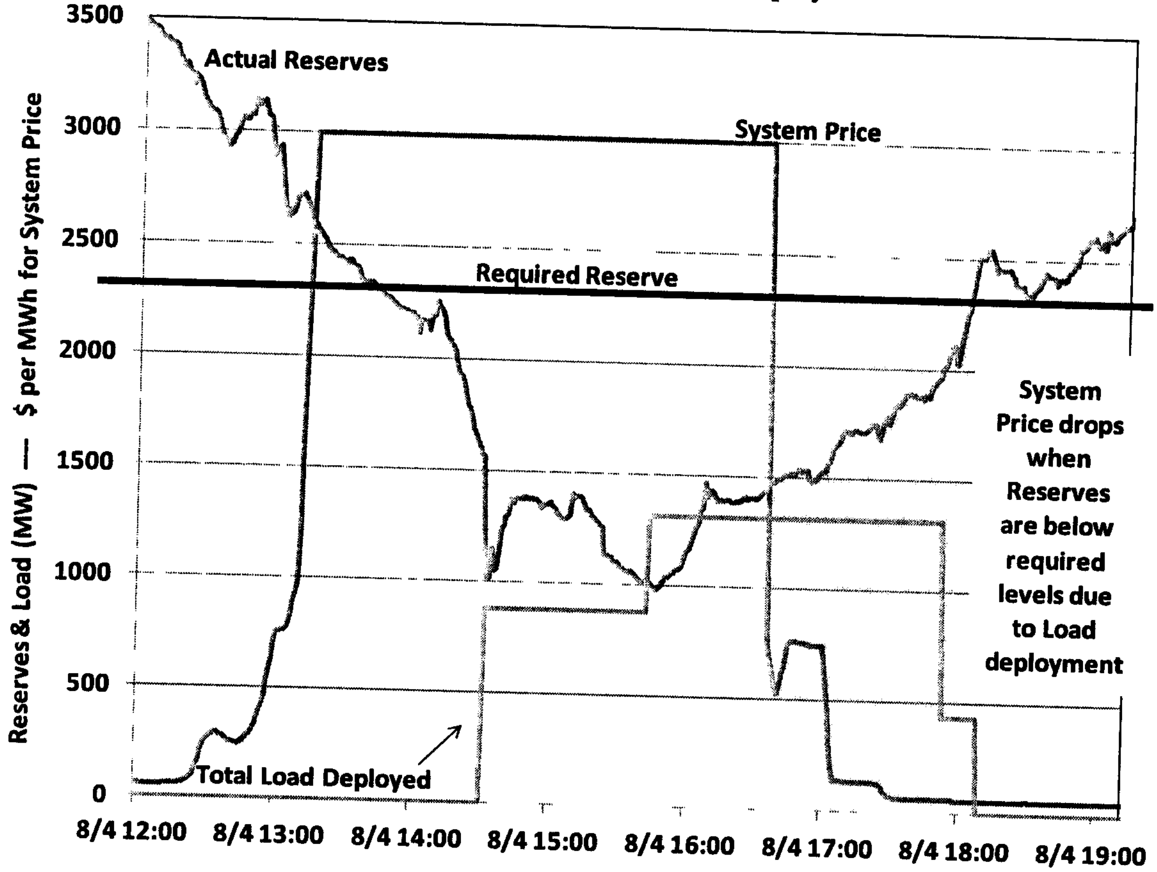
Figure 67: Daily Average of Responsive Reserves provided by Load Resources



The high level of participation by demand response in the ancillary service markets sets ERCOT apart from other operating electricity markets. For reliability reasons the maximum amount of responsive reserves that can be provided by load resources is limited to 1,150 MW. Figure 67 shows that the amount of offers by load resources routinely exceeds this level. Notable exceptions include a decrease in September of 2008 corresponding to the Texas landfall of Hurricane Ike and a more prolonged reduction from November 2008 through January 2009 that was likely a product of the economic downturn and its effect on industrial operations. Another seasonal reduction was observed during late 2009. During 2011 there was a significant reduction in loads offering to provide responsive reserve during early February and again starting in mid-July. Both of these times corresponded with expected high real-time prices. Since load resources provide capacity by reducing their consumption, they have to actually be consuming energy to be eligible to provide the capacity service. During periods of expected high prices the price paid for the energy can exceed the value received from providing responsive reserves.

During both the cold weather event in early February and the record high load period of early August, there were shortages of supply offers available for dispatch and Responsive Reserves were deployed, that is, converted to energy as one of the last steps taken before shedding firm load. During these situations, the value of the foregone reserves – which is much higher than the marginal cost of the most expensive online generator – should be reflected in energy prices to achieve efficient economic signals governing investment in generation, demand response and transmission. Unfortunately, ERCOT's dispatch software does not recognize that load has been curtailed, and computes prices based on supplying only the remaining load. A good example of this situation occurred on August 4th. Figure 68 displays available reserves and the system price for that afternoon and shows that even though reserves were below required levels, system price dropped to $60 per MWh. At this level prices are being set based supply offers and do not reflect the value of the load that is being curtailed to reliably serve the remaining system demand.

Figure 68: Pricing During Load Deployments



We recommend that ERCOT implement system changes that will ensure that all demand response that is actively deployed by ERCOT be incorporated into the dispatch software so that such deployments will be able to set the price at the value of load at times when such deployments are necessary to reliably serve the remaining system demand.

VI. Analysis of Competitive Performance

In this section we evaluate competition in the ERCOT market by analyzing the market structure and the conduct of participants in 2011. We examine market structure by using a pivotal supplier analysis that indicates the frequency with which a supplier was pivotal increased at higher levels of demand, which is consistent with observations in prior years. To evaluate participant conduct we estimate measures of physical and economic withholding. We examine withholding patterns relative to the level of demand and the size of each supplier's portfolio. We conclude this section by reviewing the impacts of the automatic mitigation mechanism included as part of the nodal market.

Based on these analyses, we find the overall performance of the ERCOT wholesale market to be competitive in 2011.

A. Structural Market Power Indicators

We analyze market structure by using the Residual Demand Index ("RDI"), a statistic that measures the percentage of load that could not be satisfied without the resources of the largest supplier. The RDI is used to measure the percentage of load that cannot be served without the resources of the largest supplier, assuming that the market could call upon all committed and quick-start capacity[21] owned by other suppliers. When the RDI is greater than zero, the largest supplier is pivotal (i.e., its resources are needed to satisfy the market demand). When the RDI is less than zero, no single supplier's resources are required to serve the load as long as the resources of its competitors are available.

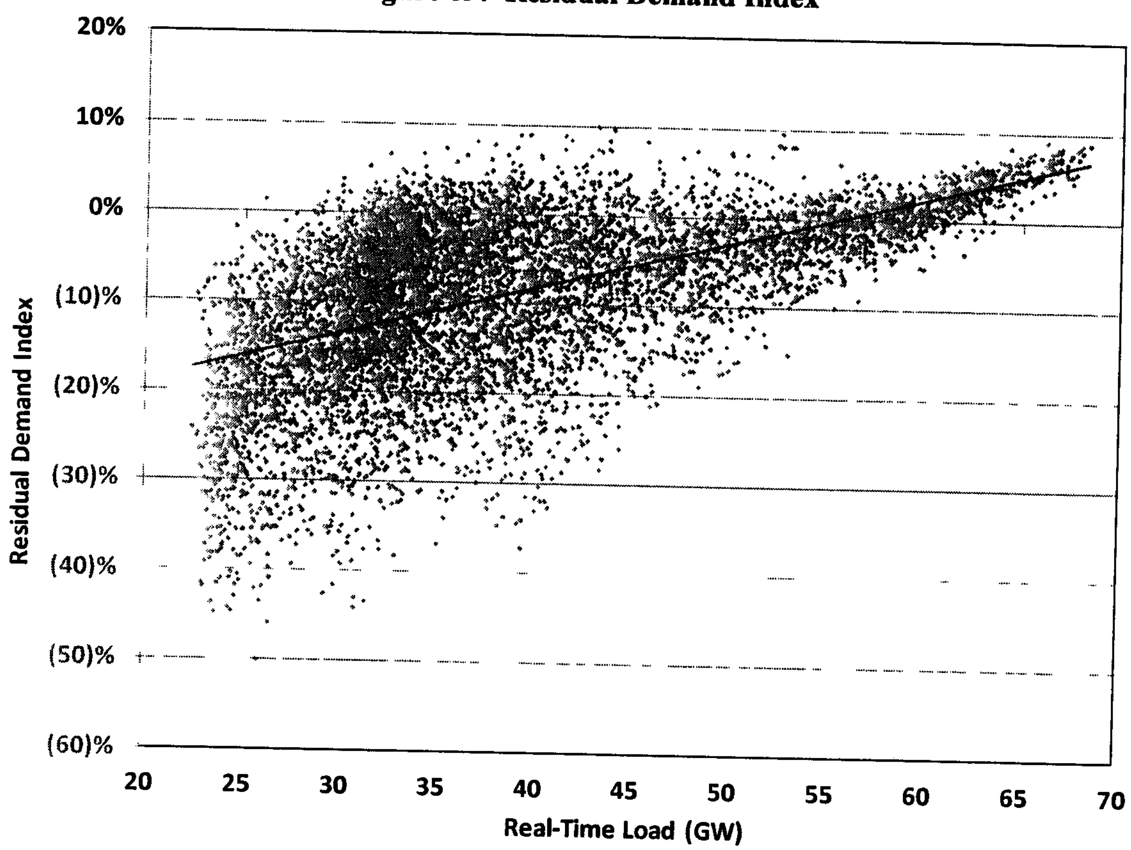
The RDI is a useful structural indicator of potential market power, although it is important to recognize its limitations. As a structural indicator, it does not illuminate actual supplier behavior to indicate whether a supplier may have exercised market power. The RDI also does not indicate whether it would have been profitable for a pivotal supplier to exercise market

21 For the purpose of this analysis, "quick-start" includes off-line simple cycle gas turbines that are flagged as on-line in the current operating plan with a planned generation level of 0 MW that ERCOT has identified as capable of starting-up and reaching full output after receiving a dispatch instruction from the real-time energy market.

power. However, it does identify conditions under which a supplier would have the ability to raise prices significantly by withholding resources.

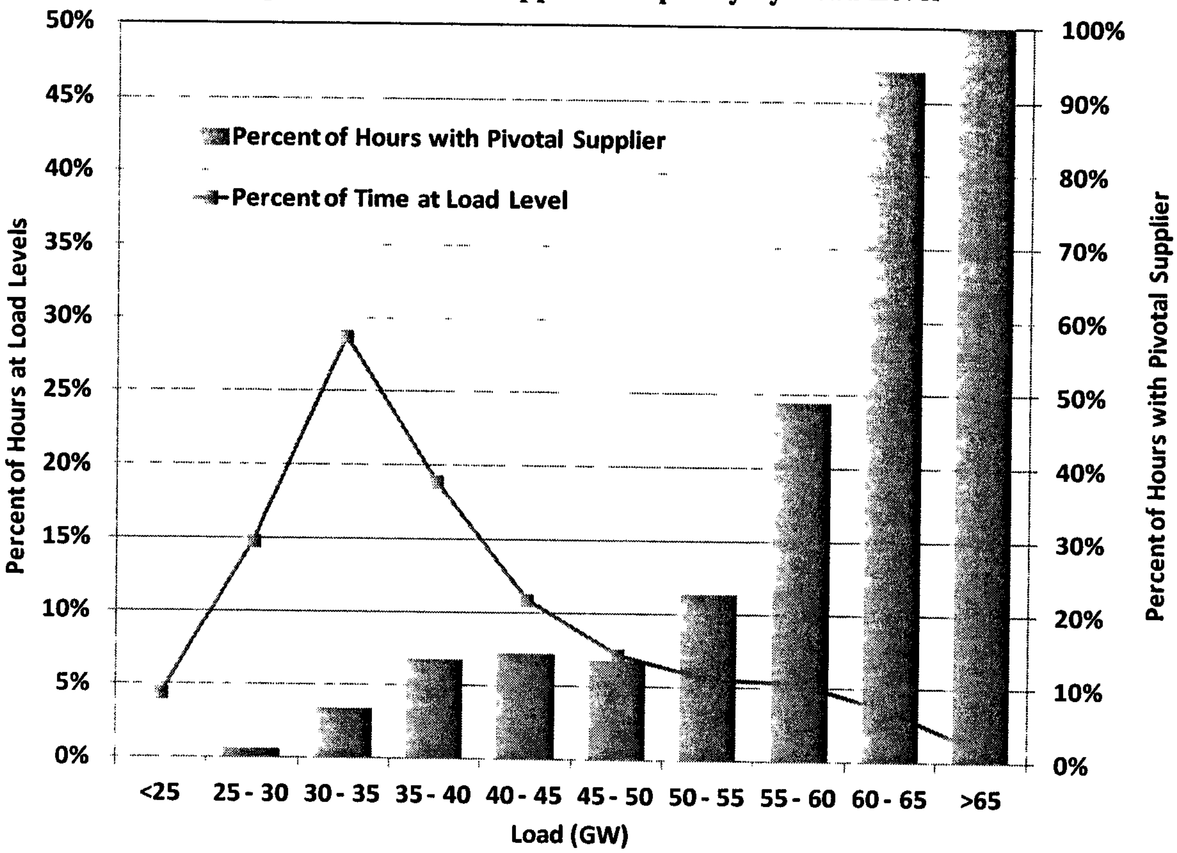
Figure 69 shows the RDI relative to load for all hours in 2011. The trend line indicates a strong positive relationship between load and the RDI. This analysis shown below is done at the QSE level because the largest suppliers that determine the RDI values own a large majority of the resources they are offering. It is possible that they also control other capacity through bilateral arrangements, although we do not know whether this is the case. To the extent that the resources scheduled by the largest QSEs are not controlled or providing revenue to the QSE, the RDIs will tend to be slightly overstated.

Figure 69: Residual Demand Index



Below, Figure 70 summarizes the results of our RDI analysis by displaying the percent of time at each load level there as a pivotal supplier. At loads greater than 65 GW there is a pivotal supplier 100 percent of the time. The figure also displays the percent of time each load level occurs. Combining these values we find that there was a pivotal supplier in approximately 15 percent of all hours of 2011. As a comparison, the same system-wide measure for the Midwest ISO resulted in zero hours with a pivotal supplier.

Figure 70: Pivotal Supplier Frequency by Load Level



It is important to recognize that inferences regarding market power cannot be made solely from this data. Bilateral contract obligations can affect a supplier's potential market power. For example, a smaller supplier selling energy in the real-time energy market and through short-term bilateral contracts may have a much greater incentive to exercise market power than a larger supplier with substantial long-term sales contracts. The RDI measure shown in the previous figure does not consider the contractual position of the supplier, which can increase a supplier's incentive to exercise market power compared to the load-adjusted capacity assumption made in this analysis.

00131

In the next analysis of RDI, we impose ramp rate limitations on the capacity available to meet load. As shown in Figure 71, the ramp constrained RDI shows the same pattern of becoming increasingly positive at higher load levels generally presents the same pattern, but is much more likely to be positive as the total capacity available to the market is smaller than in the previous analysis. We observe that the ramp rate constrained RDI was usually positive, indicating the presence of a pivotal supplier, except when load was below 25 GW.

Figure 71: Ramp-Constrained Residual Demand Index

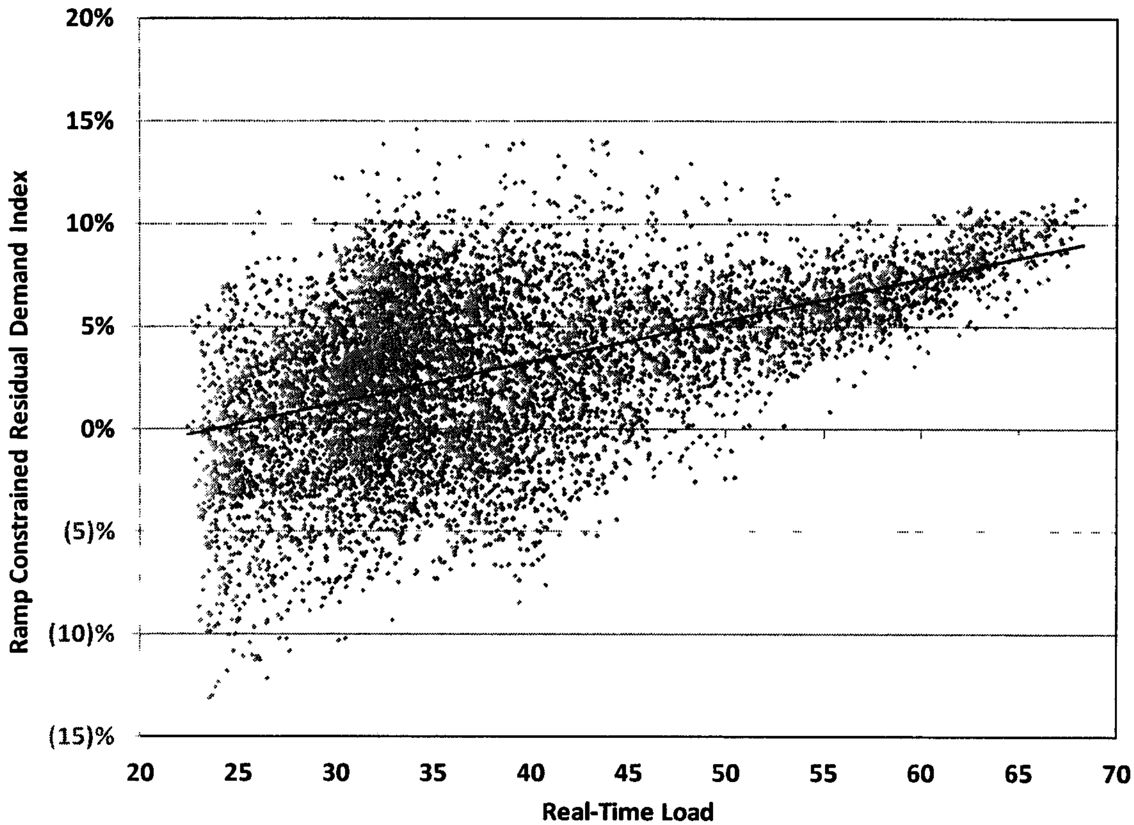
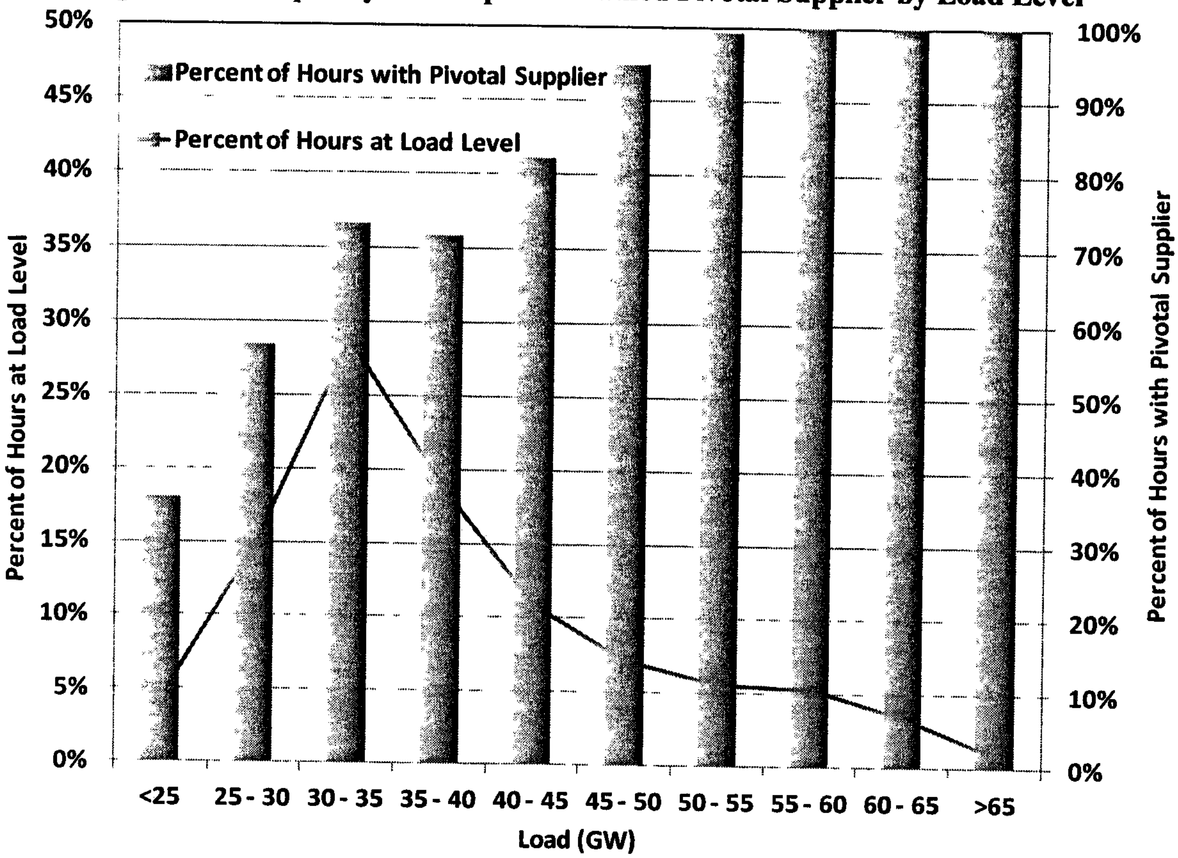


Figure 72 displays the percent of time at each load level there as a pivotal supplier when ramp rate constraints are considered. At loads greater than approximately 50 GW there is a pivotal supplier 100 percent of the time. Ramp rate constrained RDI indicates that there was a pivotal supplier in approximately 75 percent of all hours in 2011. It is important to note that this ramp rate constraint is being imposed for every dispatch interval, or approximately every 5 minutes.

Figure 72: Frequency of Ramp Constrained Pivotal Supplier by Load Level



B. Evaluation of Supplier Conduct

The previous subsection presented a structural analysis that supports inferences about potential market power. In this section we evaluate actual participant conduct to assess whether market participants have attempted to exercise market power through physical or economic withholding. First, we examine unit deratings and forced outages to detect physical withholding and then we evaluate the "output gap" to detect economic withholding.

In a single-price auction like the real-time energy market, suppliers may attempt to exercise market power by withholding resources. The purpose of withholding is to cause more expensive resources to set higher market clearing prices, allowing the supplier to profit on its other sales in the real-time energy market. Because forward prices will generally be highly correlated with spot prices, price increases in the real-time energy market can also increase a supplier's profits in the bilateral energy market. The strategy is profitable only if the withholding firm's incremental profit due to higher price is greater than the lost profit from the foregone sales of its withheld capacity.

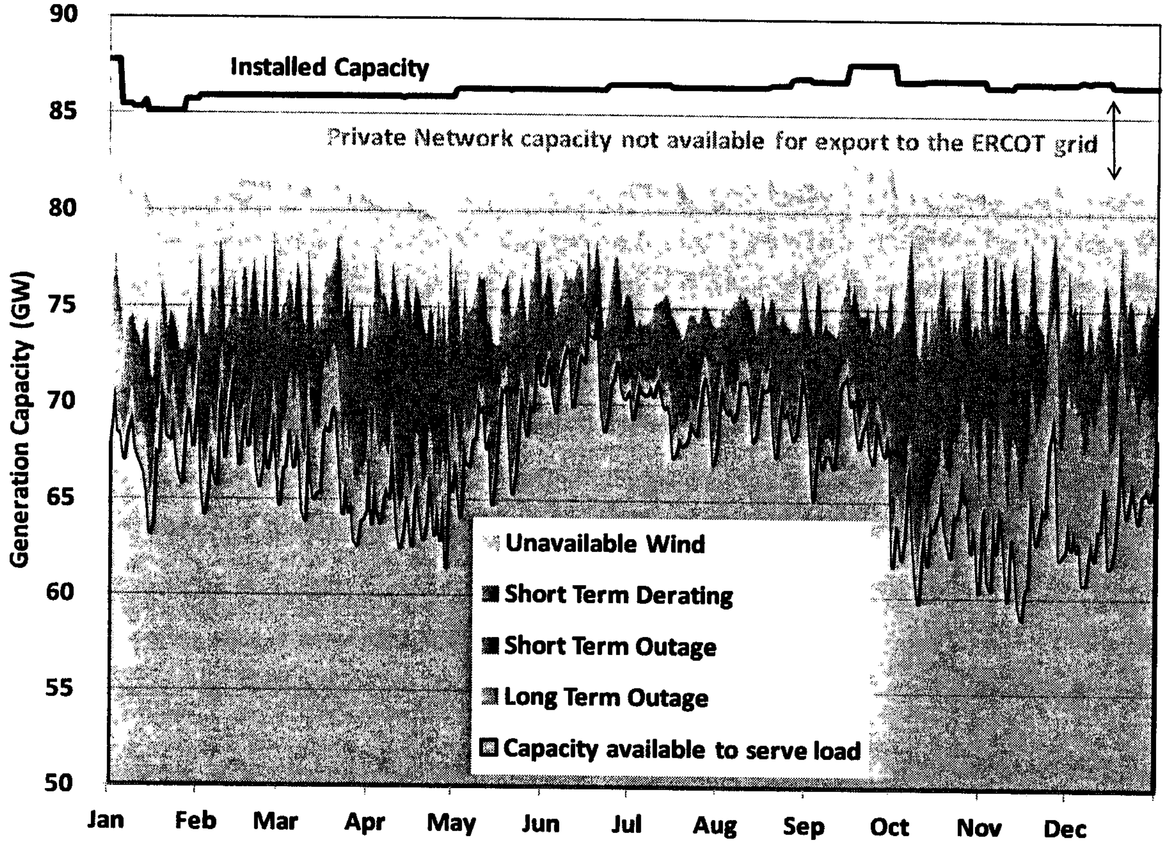
1. Generation Outages and Deratings

A substantial portion of the installed capability is frequently unavailable due to generator deratings. A derating is the difference between the maximum installed capability of a generating resource and its actual capability in a given hour. Generators may be fully derated (rating equals 0) due to a forced or planned outage. It is also very common for generating capacity to be partially derated (e.g., by 5 to 10 percent) because the resource cannot achieve its installed capability level due to technical or environmental factors (e.g., component equipment failures or ambient temperature conditions). A large portion of derated capacity is related to wind generation. It is rare for wind generators to produce at their installed capacity rating due to variations in available wind input. In this subsection, we evaluate long-term and short-term deratings to inform our evaluation of ERCOT capacity levels.

Figure 73 shows a breakdown of total installed capability for ERCOT on a daily basis during 2011. This analysis includes all in-service and switchable capacity. From the total installed capacity we subtract away (a) capacity from private networks not available for export to the ERCOT grid, (b) wind capacity not available due to the lack of wind input, (c) short-term deratings, (d) short-term outages – planned and forced, and (e) long-term – greater than 30 day outages. What remains is the capacity available to serve load.

Outages and deratings fluctuated between 3 and 15 GW, as shown in Figure 73, while wind unavailability varied between 2 and 9 GW. Short term outages were largest in April and October and small during the summer, which are consistent with expectations. Short term deratings increased in August and early September as the extreme heat and drought conditions limited some generators from being able to produce at full capacity.

Figure 73: Reductions in Installed Capability



The next analysis focuses specifically on short-term outages and deratings. Figure 74 shows the average magnitude of the outages and deratings lasting less than 30 days for the year and for each month during 2011.

Figure 74: Short-Term Outages and Deratings

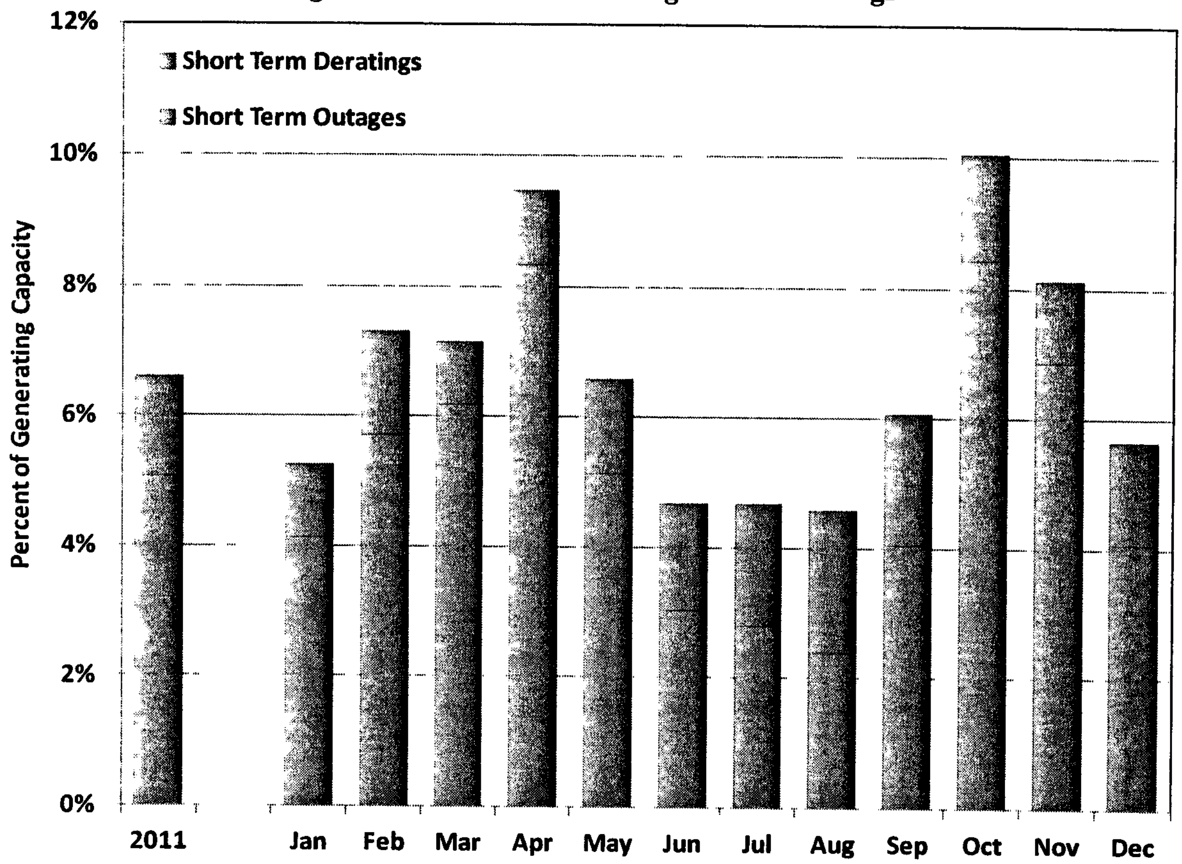


Figure 74 shows that total short-term deratings and outages were as large as 10 percent of installed capacity in the spring and fall, dropping to as low as 5 percent for the summer. Most of this fluctuation was likely due to anticipated planned outages.

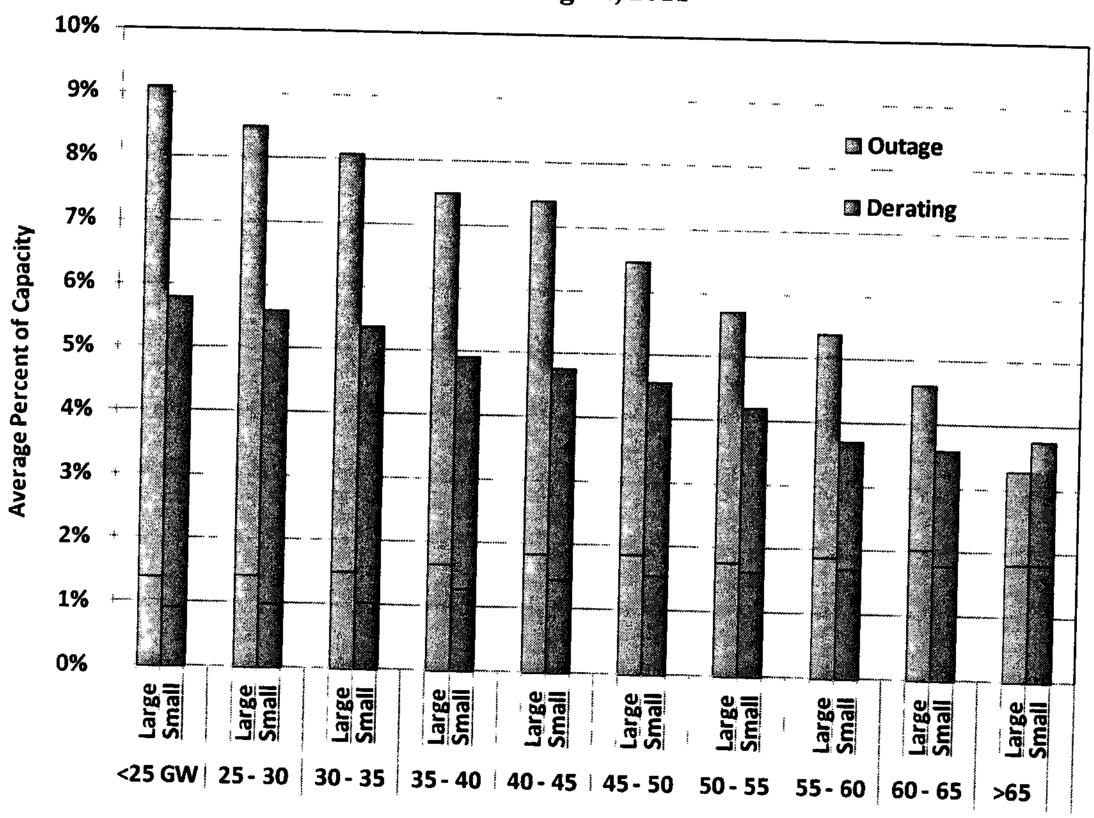
2. Evaluation of Potential Physical Withholding

Physical withholding occurs when a participant makes resources unavailable for dispatch that are otherwise physically capable of providing energy and that are economic at prevailing market prices. This can be done either by derating a unit or declaring it as forced out of service. Because generator deratings and forced outages are unavoidable, the goal of the analysis in this section is to differentiate justifiable deratings and outages from physical withholding. We test for physical withholding by examining deratings and outage data to ascertain whether the data are correlated with conditions under which physical withholding would likely be most profitable.

The RDI results shown in Figure 69 through Figure 72 indicate that the potential for market power abuse rises at higher load levels as the frequency of positive RDI values increases. Hence, if physical withholding is a problem in ERCOT, we would expect to see increased deratings and outages at the highest load levels. Conversely, because competitive prices increase as load increases, deratings and outages in a market performing competitively will tend to decrease as load approaches peak levels. Suppliers that lack market power will take actions to maximize the availability of their resources since their output is generally most profitable in these peak periods.

Figure 75 shows the average relationship of short-term deratings and forced outages as a percentage of total installed capacity to real-time load level for large and small suppliers. Portfolio size is important in determining whether individual suppliers have incentives to withhold available resources. Hence, the patterns of outages and deratings of large suppliers can be usefully evaluated by comparing them to the small suppliers' patterns.

Figure 75: Outages and Deratings by Load Level and Participant Size June to August, 2011



Long-term deratings are not included in this analysis because they are unlikely to constitute physical withholding given the cost of such withholding. Wind and private network resources are also excluded from this analysis because of the high variation in the availability of these classes of resources. The large supplier category includes the four largest suppliers in ERCOT. The small supplier category includes the remaining suppliers.

Figure 75 suggests that as demand for electricity increases, all market participants tend to make more capacity available to the market. For small suppliers, the combined short-term derating and forced outage rates decreased from approximately 6 percent at low demand levels to less than 4 percent at load levels above 55 GW. Large suppliers have derating and outage rates that are higher than those of small suppliers across the range of load levels, up to the very highest. For large suppliers, the combined short-term derating and forced outage rates decreased from 9 to just over 3 percent, across all load levels.

Except for at the very highest load levels, the combined outage and derating percentage for small providers is lower than for the large providers. That pattern is different than in previous years, but given the overall magnitude not immediately troubling. Some of the difference may be due to data available from the nodal market systems being different than what was available from zonal market systems.

3. Evaluation of Potential Economic Withholding

To complement the prior analysis of physical withholding, this subsection evaluates potential economic withholding by calculating an "output gap." The output gap is defined as the quantity of energy that is not being produced by in-service capacity even though the in-service capacity is economic by a substantial margin given the real-time energy price. A participant can economically withhold resources, as measured by the output gap, by raising its energy offers so as not to be dispatched.

Resources are considered for inclusion in the output gap when they are committed and producing at less than full output. Energy not produced from committed resources is included in the output gap if the real-time energy price exceeds by at least $50 per MWh that unit's mitigated offer cap which serves as an estimate of the marginal production cost of energy from that resource.

Before presenting the results of the Output Gap analysis, a description of the two-step aspect of ERCOT's dispatch software is required. In the first step, the dispatch software calculates output levels (Base Points) and associated locational marginal prices (LMPs) using the participants' offer curves and only considering transmission constraints that have been deemed competitive. These "reference prices" at each generator location are compared with that generator's mitigated offer cap, and the higher of the two is used to formulate the offer curve to be used for that generator in the second step in the dispatch process. The resulting mitigated offer curve is used by the dispatch software to determine the final output levels for each generator taking all transmission constraints into consideration.

If a market participant has sufficient market power, it might raise its offer in such a way to increase the reference price in the first step. Although in the second step, the offer appears to be mitigated, the market participant has still influenced the market price. This output gap is measured by the difference between the capacity level on a generator's original offer curve at the first step reference price and the capacity level on the generator's cost curve at the first step reference price. However, this output gap is only indicative because no output instructions are sent based on the first step. It is only used to screen out whether a market participant is withholding in a manner that may influence the reference price.

From the results of this analysis, shown in Figure 76, we observe only very small amounts of capacity at only the very highest loads that would be considered part of this output gap. These small quantities raise no competitive concerns.

Figure 76: Incremental Output Gap by Load Level and Participant Size – Step 1

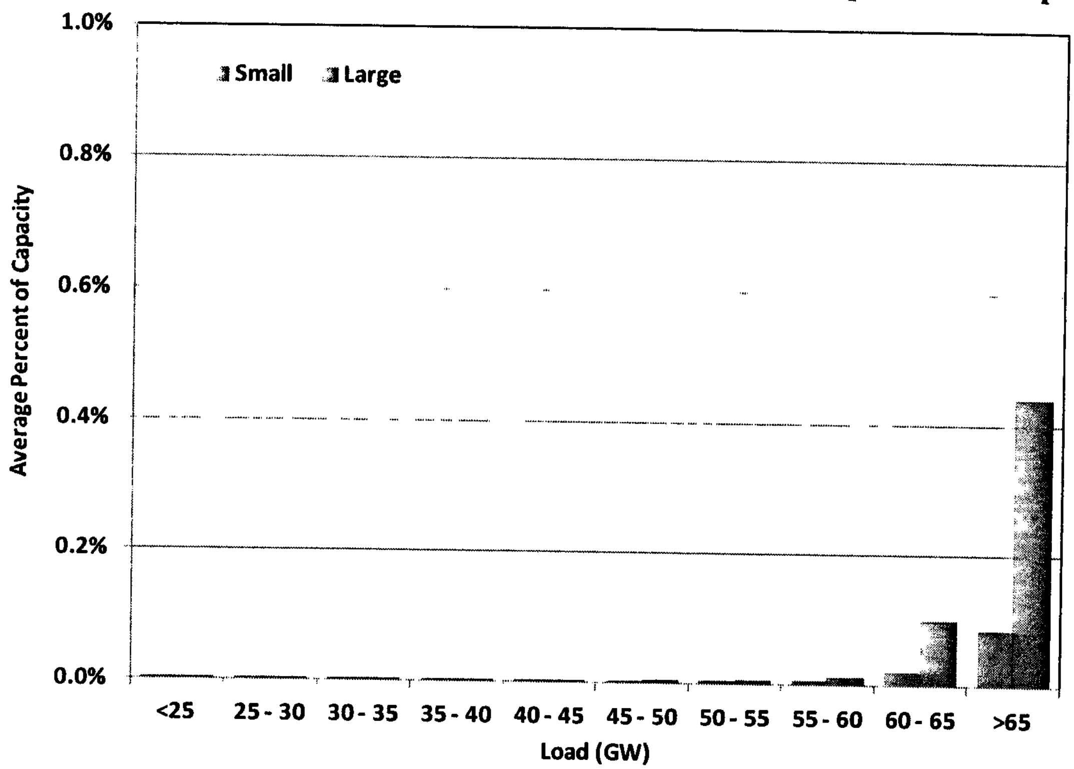
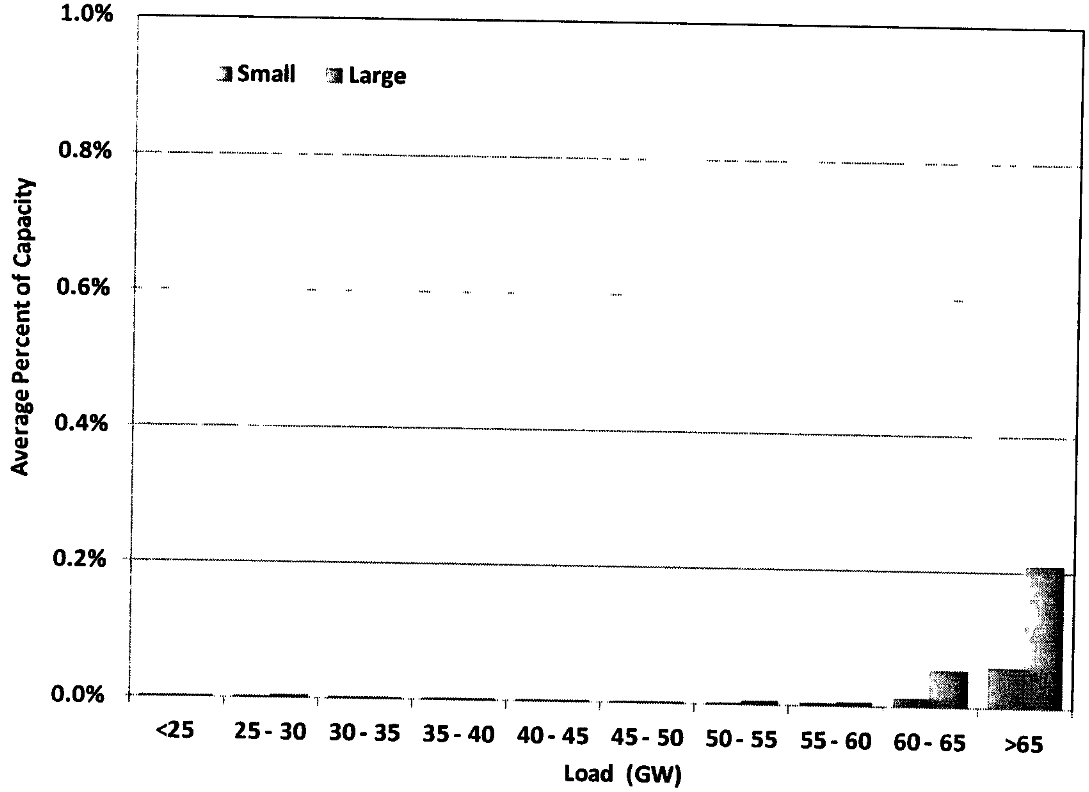


Figure 77 shows the ultimate output gap, measured by the difference between a unit's operating level and the output level had the unit been competitively offered to the market. In the second step of the dispatch, the after-mitigation offer curve is used to determine dispatch instructions and locational prices. As previously illustrated, even though the offer curve is mitigated there is still the potential for the mitigated offer curve to be increased as a result of a high first step reference price due to a market participant exerting market power.

Similar to the previous analysis, Figure 77 shows the magnitude of the output gap to be very small, even at the highest load levels. These small quantities raise no competitive concerns.

Figure 77: Incremental Output Gap by Load Level and Participant Size – Step 2

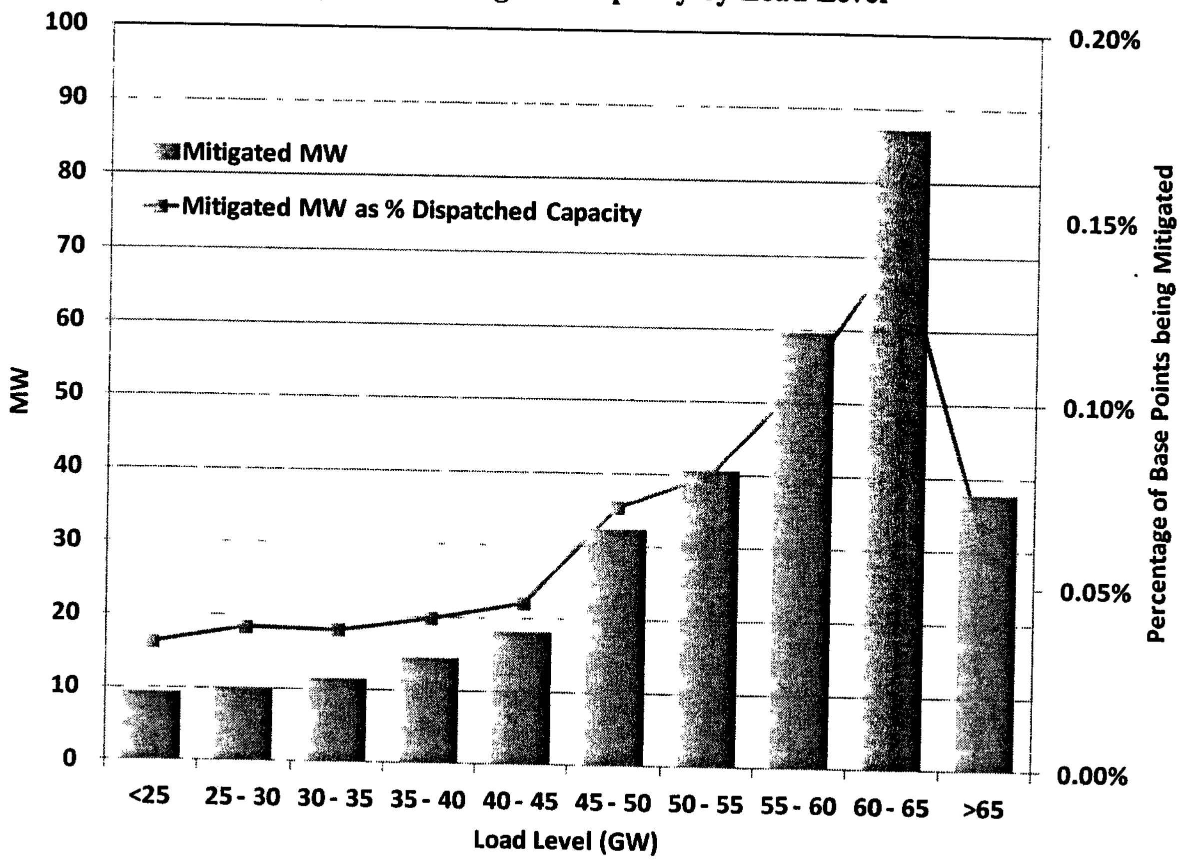


C. Mitigation

As described in the previous subsection, the dispatch software includes an automatic, two step price mitigation process. This approach is intended to limit the ability of a specific generator to raise prices in the event of a transmission constraint that requires their output to resolve. In this section we analyze the quantity of capacity affected by this mitigation process.

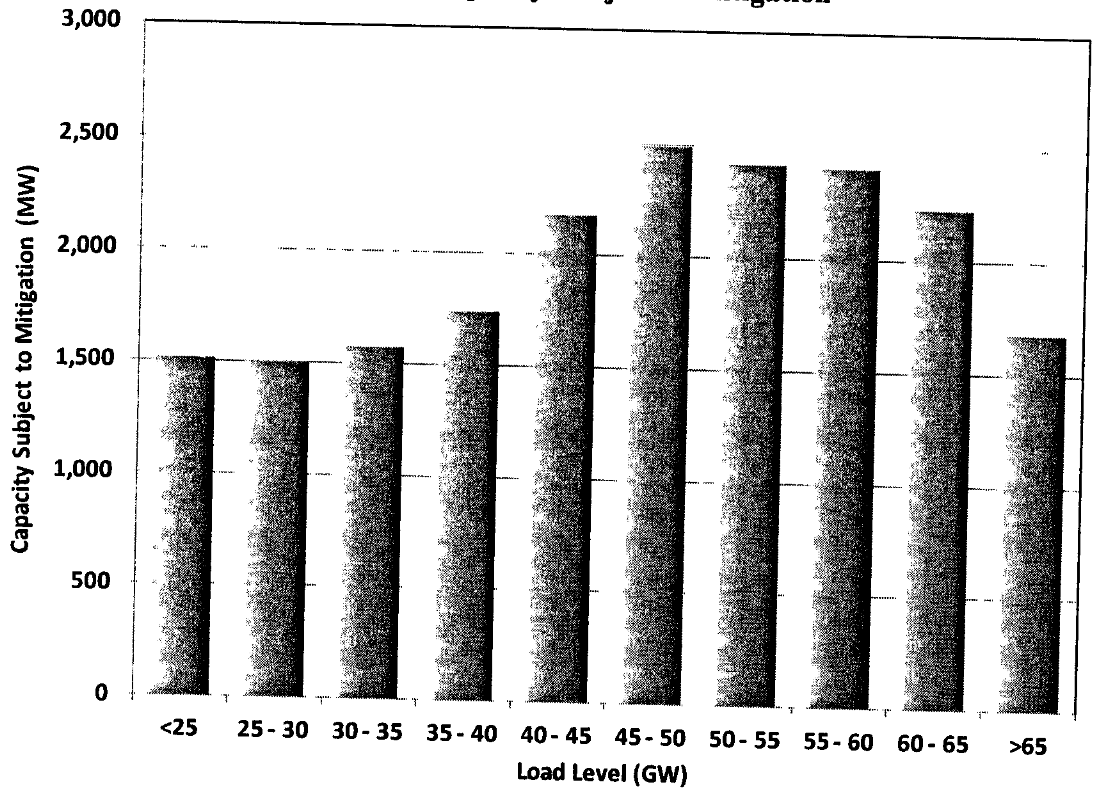
Our first analysis computes how much capacity, on average, is actually mitigated during each dispatch interval. The results, shown in Figure 78, are provided by load level. The quantities of capacity actually mitigated are relatively small, averaging 10 MW at low loads and increasing to almost 90 MW at loads between 60 and 65 GW. At the very highest load levels, above 65 GW, average amounts of mitigated capacity drop to less than 40 MW. This decrease is likely due to the reluctance by ERCOT operators to activate certain transmission constraints during very high system load conditions and mitigation only has an effect when a non-competitive transmission constraint is active.

Figure 78: Mitigated Capacity by Load Level



In the previous figure only the amount of capacity that can be dispatched within one interval is counted as mitigated. In our next analysis we compute the total capacity subject to mitigation. These values are determined by comparing a generator's mitigated and unmitigated (as submitted) offer curves and determining the point at which they diverge. We then take the difference between the total unit capacity and the capacity at the point the curves diverge. This calculation is performed for all units and aggregated by load level, as shown in Figure 79. From this figure we observe that at most 6 percent of capacity necessary to serve load is subject to mitigation.

Figure 79: Capacity Subject to Mitigation



Although executing all the time, the automatic price mitigation aspect of the two step dispatch process only has an effect when a non-competitive transmission constraint is active. One concern with this process is that the mere existence of an active non-competitive transmission constraint can result in mitigating certain units inappropriately. The mitigation process is intended to limit the ability of a generator to affect price when their output is required to manage congestion. The process does not currently address the situation where there are a competitively sufficient number of generators on the other side of the constraint and mitigates all their offers. This results in unnecessary mitigation which is a situation that should be addressed. One way to improve the mitigation process would be to introduce an impact test to determine whether units are relieving or contributing to a transmission constraint, and only subject the relieving units to mitigation.