POTOMAC ECONOMICS

2011 STATE OF THE MARKET REPORT FOR THE ERCOT WHOLESALE ELECTRICITY MARKETS

POTOMAC ECONOMICS, LTD.

Independent Market Monitor for the
ERCOT Wholesale Market

July 2012

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

A. Introduction

This report reviews and evaluates the outcomes of the ERCOT wholesale electricity markets in 2011, and is submitted to the Public Utility Commission of Texas ("PUCT") and the Electric Reliability Council of Texas ("ERCOT") pursuant to the requirement in Section 39.1515(h) of the Public Utility Regulatory Act. It includes assessments of the incentives provided by the current market rules and procedures, and analyses of the conduct of market participants. This report also assesses the effectiveness of the scarcity pricing mechanism pursuant to the provisions of PUCT Substantive Rule 25.505(g).

ERCOT transitioned from the zonal market design that had been in place since 2001 and implemented the nodal market design on December 1, 2010. Thus, this is the first annual report that contains an entire year of nodal market operations. Key findings and statistics from 2011 include the following:

- The ERCOT wholesale market performed competitively in 2011.
- The ERCOT-wide load-weighted average real-time energy price was $53.23 per MWh in 2011, a 35 percent increase from $39.40 per MWh in 2010. The increase was primarily driven by extreme weather in February and August which led to operating reserve deficiencies that resulted in real time energy prices reaching $3,000 per MWh for sustained periods of time.
- The average price for natural gas was 9.2% lower in 2011 than in 2010, decreasing from $4.34 per MMBtu in 2010 to $3.94 per MMBtu in 2011.
- Total ERCOT load in 2011 was 5.0 percent higher than 2010. Peak load increased by 4.0 percent, setting a new all time system hourly peak of 68,379 MW on August 3rd.
- The West to North interface constraint was the most frequently occurring transmission constraint in 2011. It was active at some point during every month and was binding more than 20 percent of the time.

- ★ More reliable and efficient shortage pricing mechanisms than existed in the zonal market allowed energy prices to rise automatically up to the system-wide offer cap during periods of operating reserve shortages. Prices at the system-wide offer cap were experienced in dispatch intervals which totaled 28.5 hours in 2011, or 0.33 percent of the total hours.

- ★ Net revenues provided by the market in 2011 were sufficient to support investment in either new simple-cycle natural gas-fired turbines or natural gas-fired combined-cycle generation. This was largely the result of the increase in shortage pricing in 2011.

B. Review of Real-Time Market Outcomes

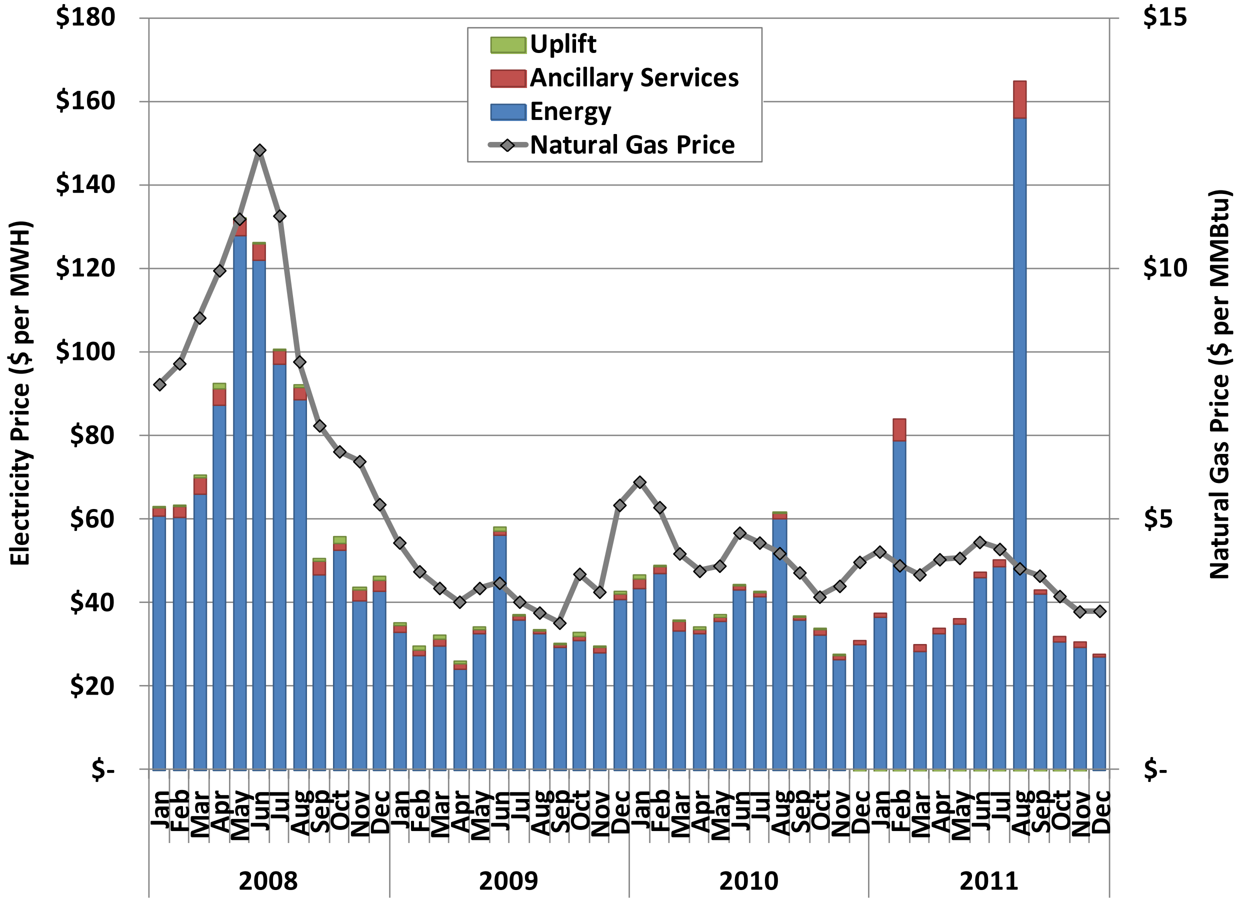
As is typical in other wholesale markets, only a small share of the power produced in ERCOT is transacted in the spot market. However, the pricing outcomes in the real time energy market are very important because they set the expectations for prices in the forward markets where most transactions take place. Unless there are barriers preventing arbitrage of the prices between the spot and forward markets, the prices in the forward market should be directly related to the prices in the spot market.

The average real-time energy prices by zone in 2008 through 2011 are shown below:

	Average Real-Time Electricity Price			
	2008	**2009**	**2010**	**2011**
ERCOT	**$77.19**	**$34.03**	**$39.40**	**$53.23**
Houston	$82.95	$34.76	$39.98	$52.40
North	$71.19	$32.28	$40.72	$54.24
South	$85.31	$37.13	$40.56	$54.32
West	$57.76	$27.18	$33.76	$46.87
Natural Gas	$8.50	$3.74	$4.34	$3.94

The largest component of the all-in cost of wholesale electricity is the energy cost, which is reflected by the locational marginal prices determined in the real-time energy market. ERCOT average real-time market prices were 35 percent higher in 2011 than in 2010. The ERCOT-wide load-weighted average price was $53.23 per MWh in 2011 compared to $39.40 per MWh in 2010. February and August experienced the largest increases to real-time energy prices in 2011, averaging 67 and 160 percent higher than the prices in the same months in 2010. Price increases in both months were driven by extreme weather conditions which led to operating reserve deficiencies resulting in real-time energy prices reaching $3,000 per MWh for sustained periods of time.

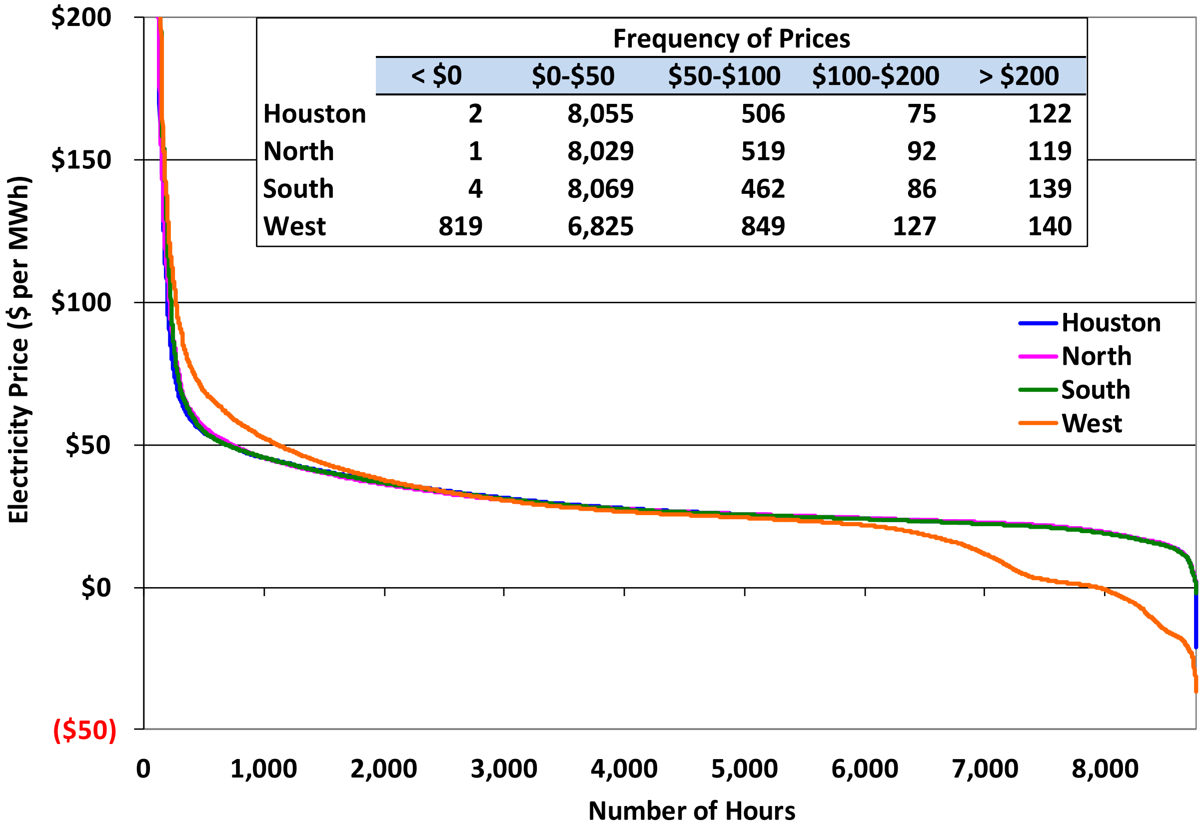
Average All-in Price for Electricity in ERCOT



The increase in real-time energy prices was partially offset by lower fuel prices in 2011. Natural gas price decreased 9 percent in 2011, averaging $3.94 per MMBtu in 2011 compared to $4.34 per MMBtu in 2010. Although lower natural gas prices contributed to lower real-time energy prices in many hours, these reductions were smaller than the price effects of the shortages in February and August.

To depict how real-time energy prices vary by hour in each zone, the next figure shows the hourly average price duration curve in 2011 for four ERCOT load zones. The Houston, North and South load zones had similar prices over the majority of hours. The price duration curve for the West Zone is noticeably different than the other zones, with more hours with prices greater than $50 per MWh and over 800 hours (9 percent of the time) when the average hourly price was less than zero. The occurrences of relatively low prices in the West zone are generally caused by high wind output in the West that frequently results in severe congestion on transmission interfaces from the West zone to the other zones in ERCOT. The occurrences of relatively higher prices in the West zone are caused by local transmission constraints that typically occur under low wind and high load conditions. Specifics about these transmission constraints are provided in Section III, Transmission and Congestion.

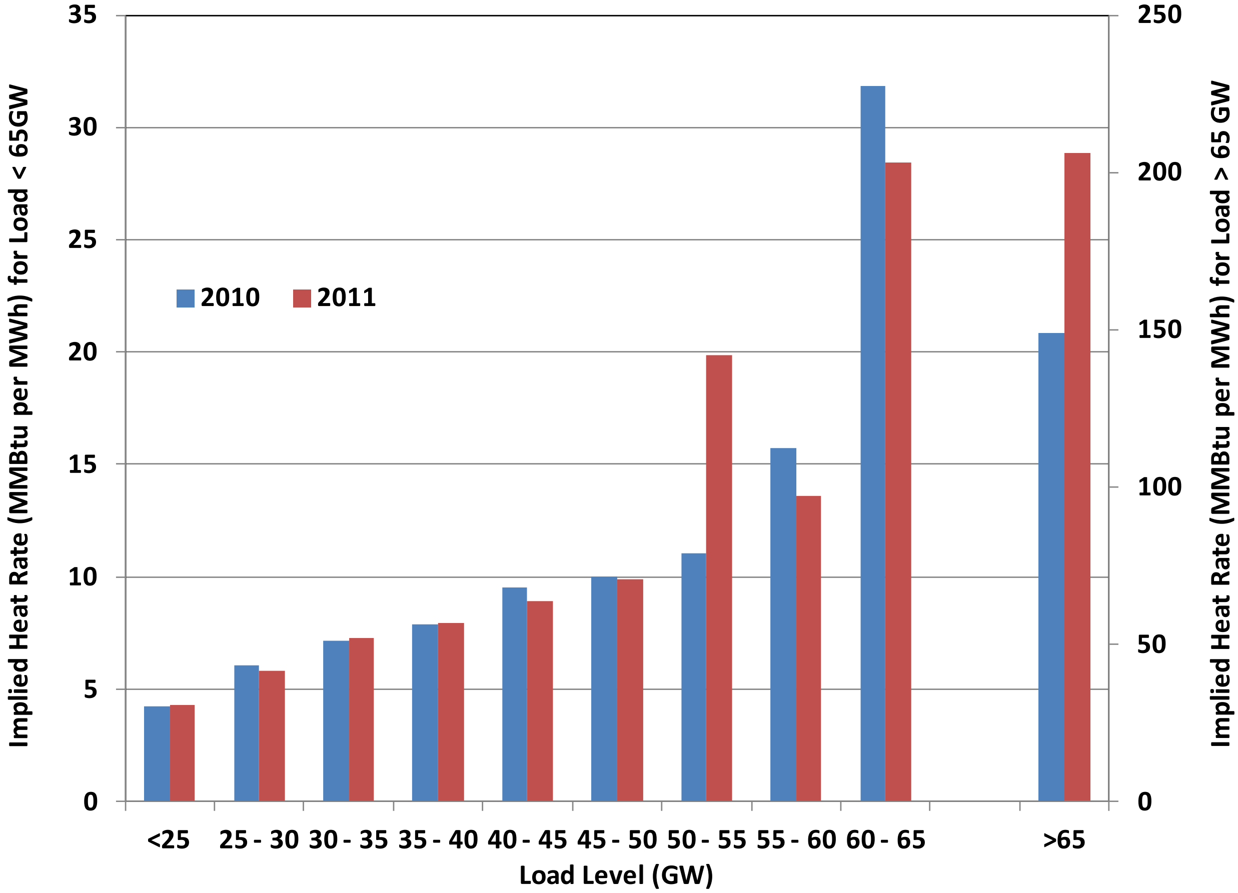
Zonal Price Duration Curves



The examination of the real-time energy market continues with an evaluation of implied heat rates at various load levels. The implied heat rate is a metric that shows changes in energy prices that are not due to changes in fuel prices. It is calculated by dividing the real-time energy price by the natural gas price. The figure below provides the average heat rate at various system load levels for 2011 and 2010. In a well performing market, a clear positive relationship between these two variables is expected since resources with higher marginal costs must be dispatched to serve higher loads. Although there is generally a positive relationship, a noticeable disparity for loads between 50 and 55 GW can be observed. During the extreme cold weather event in early

February, loads were at this level while prices reached $3,000 per MWh for a sustained period of time. Small reductions in heat rates for most load levels during 2011 compared to 2010 were observed and may be attributed to the enhanced efficiency of the nodal market.

Heat Rate and Load Relationship

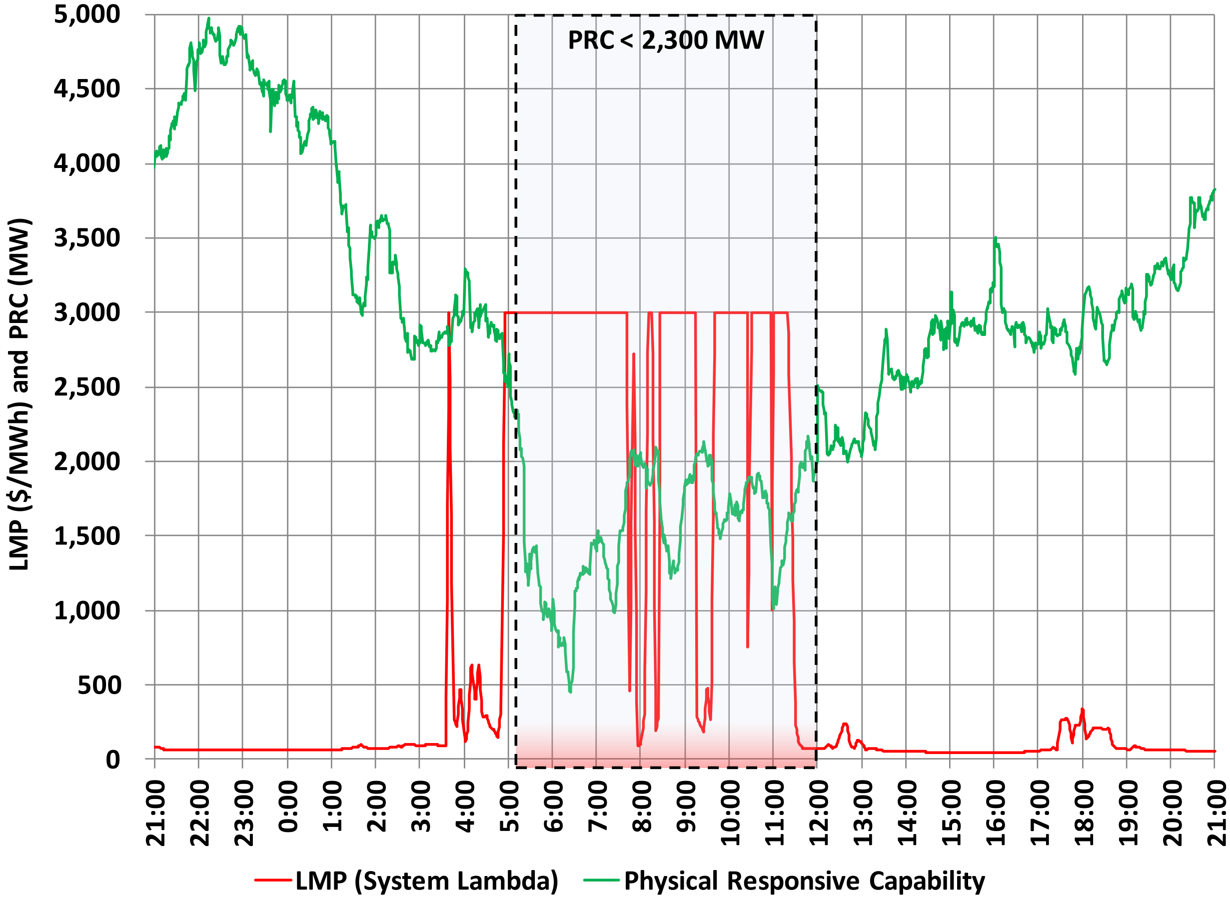


February Cold Weather Event

A significant operational challenge greeted the nascent nodal market in the early morning of February 2, 2011, when the ERCOT region experienced extreme cold weather conditions, record electricity demand levels, and the loss of numerous electric generating facilities across the ERCOT region. These events combined to result in the deployment of load resources contracted to provide responsive reserve service and Emergency Interruptible Load Service ("EILS") and culminated with 4,000 MW of firm load being shed for several hours. The resulting market outcomes had a sizable effect on the overall annual results.

During the morning of February 2nd, ERCOT operating reserve levels were reduced to perilously low levels for a sustained period of time. ERCOT's primary measure of overall operating reserves is Physical Responsive Reserve ("PRC"), and ERCOT will remain in various levels of EEA once PRC drops below 2,300 MW. The figure below shows the wholesale market prices and PRC from 21:00 on February 1 through 21:00 on February 2, 2011.

Prices and PRC (2/1/11 21:00 – 2/2/11 21:00)

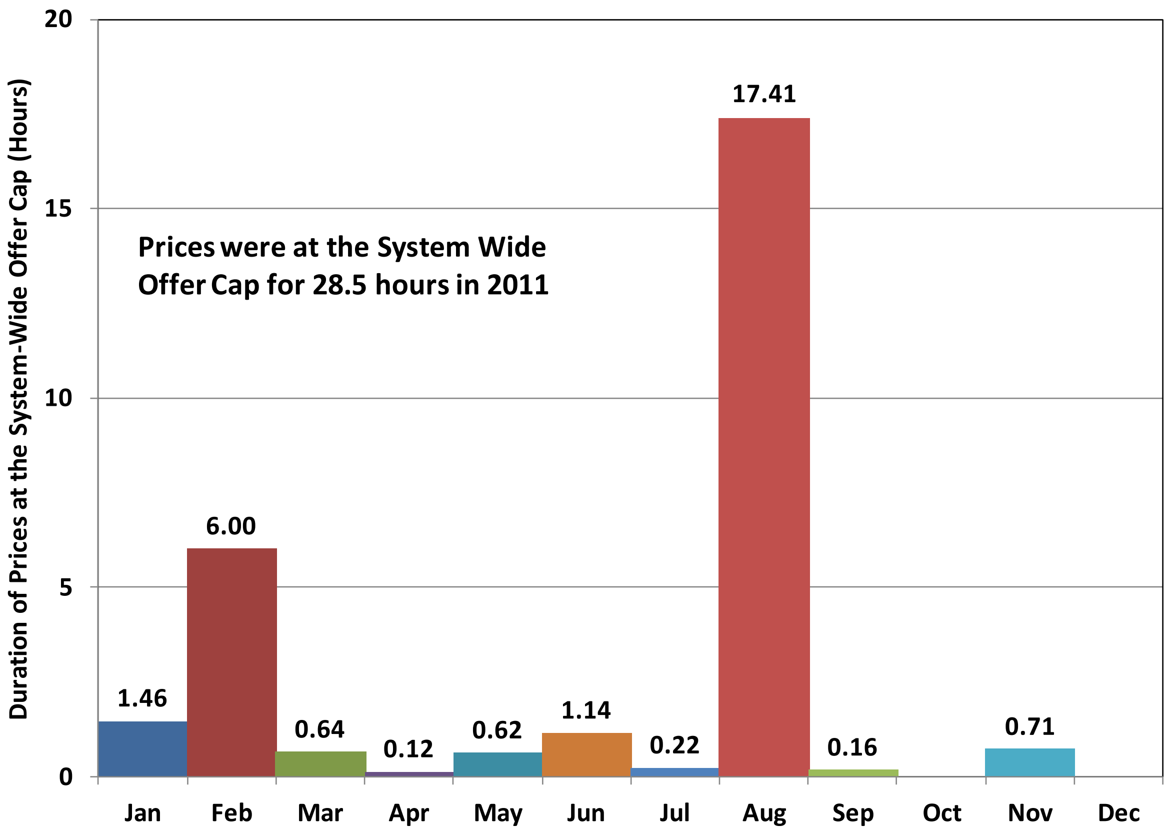


These wholesale market pricing outcomes were consistent with the ERCOT energy-only market design. The wholesale market prices began communicating the degradation in system reliability as early as 3:30 a.m. By 4:55 a.m. – 15 minutes prior to the reduction of PRC below the minimum acceptable level of 2,300 MW and 50 minutes prior to the first stage of firm load shedding – prices were consistently communicating the rapidly deteriorating system reliability conditions. Finally, as load levels naturally reduced and reserve levels were restored, prices dropped back to levels typical of non-shortage conditions.

August Weather Conditions and Shortages

The summer of 2011 will be remembered as the hottest and driest on record in ERCOT. These extreme weather conditions led to record high demand for electricity during August. There were 50 hours in 2011 with electricity demands that exceeded the highest hourly demand that occurred in 2010.

Duration of Prices at the System Wide Offer Cap



During these high demand conditions there is an increased likelihood that the available generation capacity is not sufficient to meet customer demands for electricity and maintain the required reliability reserves. The nodal market causes energy prices to rise toward the system-wide offer cap as available operating reserves approach minimum required levels to reflect the degradation in system reliability. Presented in the figure above is the aggregated amount of time represented by all dispatch intervals where the real-time energy price was at the system-wide offer cap, displayed by month. Of the 28.5 hours of the annual total time at the system-wide offer cap, more than 17 hours (60 percent) occurred during August.

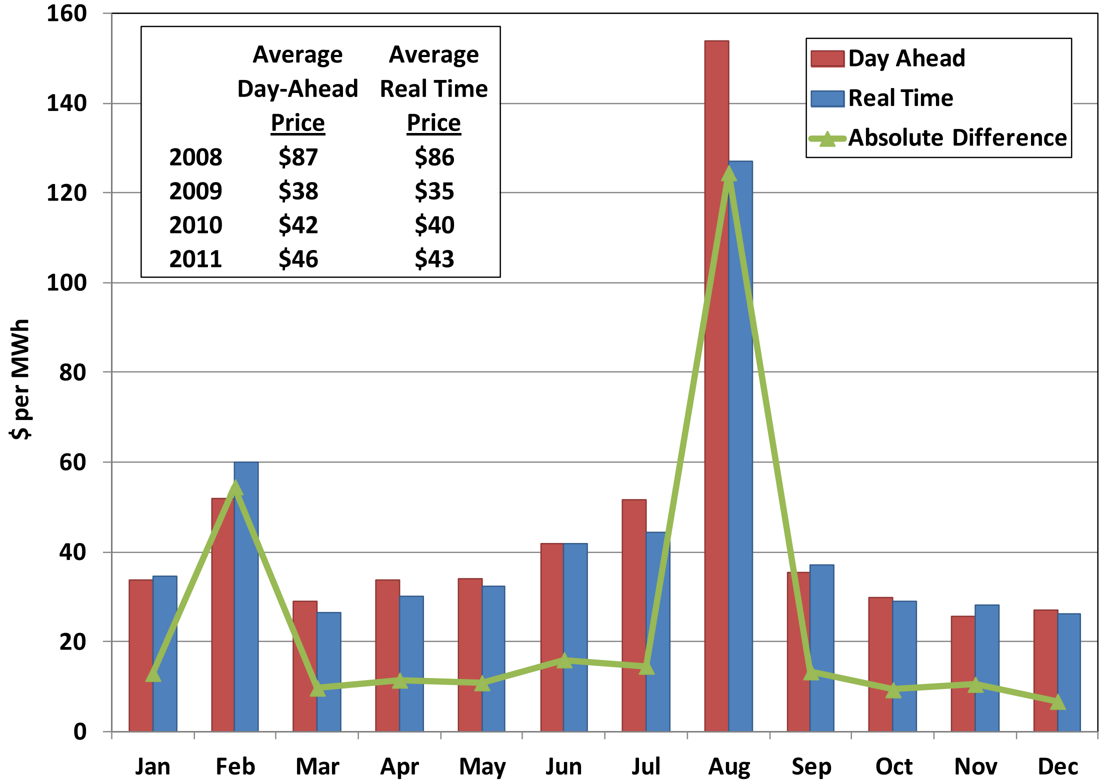
C. Review of Day-Ahead Market Outcomes

The performance of the day-ahead market is important because it coordinates the commitments of the ERCOT generation and most wholesale energy bought or sold through the ERCOT markets is settled in the day-ahead market. Day-ahead market performance is primarily evaluated by the degree to which its outcomes converge with those of the real-time market because the real-time market reflects actual physical supply and demand for electricity. In a well-functioning market, participants should eliminate sustained price differences by making day-ahead purchases or sales to arbitrage them over the long-term.

To measure the short-term deviations between real-time and day-ahead prices, the average of the absolute value of the difference between the day-ahead and real-time price on a daily basis is also calculated.

The figure below shows the price convergence between the day-ahead and real-time market, summarized by month. The simple average of day-ahead prices in 2011 was $46 per MWh, compared to the simple average of $43 per MWh for real-time prices. This slight premium is consistent with expectations due to the much higher volatility of real-time prices. Risk is lower for loads purchasing in the day-ahead and higher for generators selling day-ahead. The higher risk for generators is associated with the potential of having a forced outage and buying back energy at real-time prices. This may explain why the highest premiums occurred during the highest priced months. Overall, the day-ahead premiums were very similar to the differences observed in 2009 and 2010.

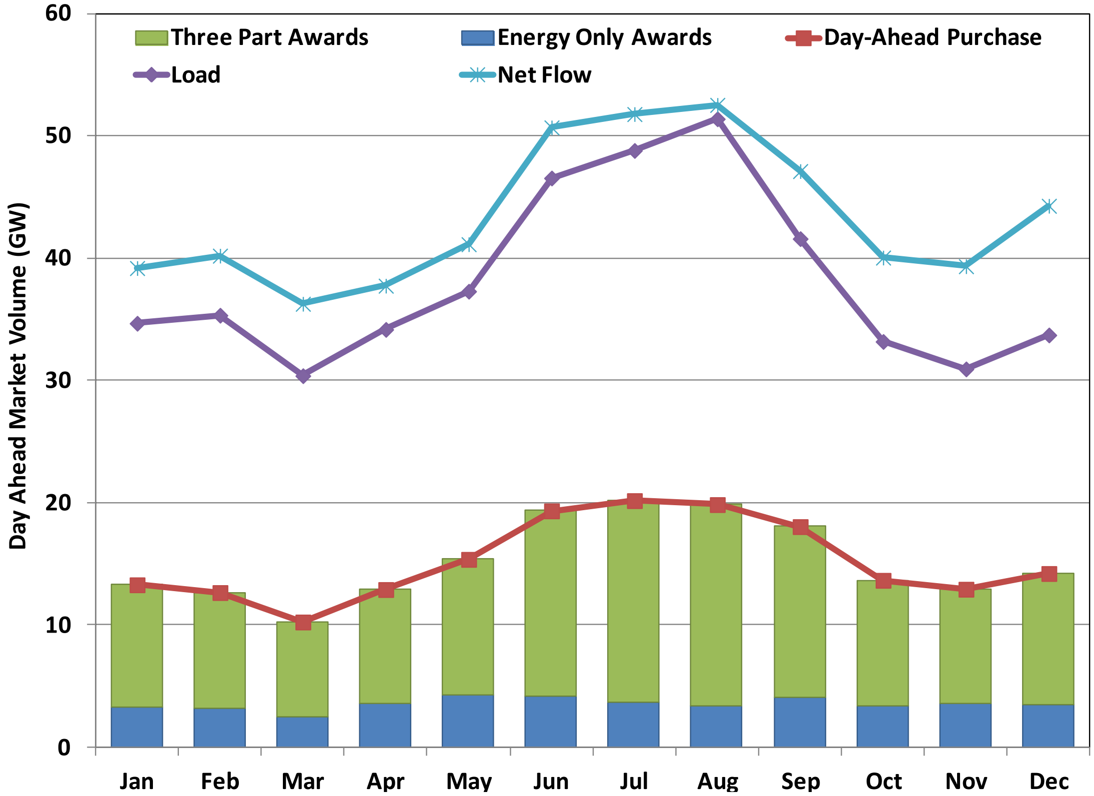
Convergence between Forward and Real-Time Energy Prices



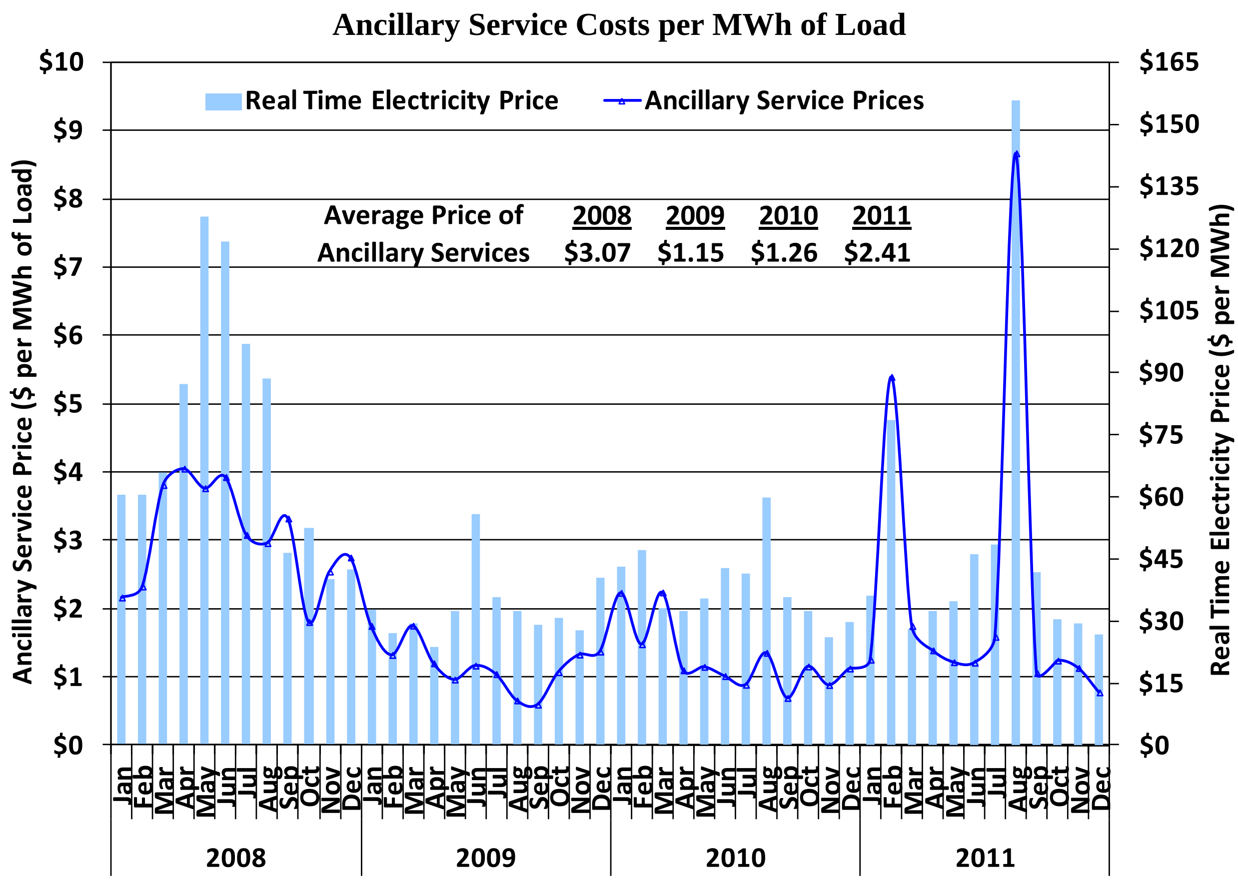
The average absolute difference between day-ahead and real-time prices was $24.50 per MWh in 2011; much higher than in the previous two years where the average absolute difference was $12.25 and $12.37 in 2010 and 2009, respectively. This large increase was the result of the significant periods of very high real-time prices during February and August. Removing the contribution from these two months reduces the average absolute difference to $11.49 per MWh in 2011.

Summarized in the figure below is the volume of day-ahead market activity by month. It shows that day-ahead purchases are approximately 40 percent of real-time load. These energy purchases are met through a combination of generator specific and virtual offers. Once the effects of net energy flows associated with purchases of PTP Obligations are included, total volumes transacted in the day-ahead market are, on average, greater than real-time load.

Volume of Day-Ahead Market Activity by Month



Ancillary Service capacity is procured as part of the day-ahead market clearing. The figure below shows the monthly total ancillary service costs per MWh of ERCOT load and the average real time energy price for 2008 through 2011. Total ancillary service costs are generally correlated with real-time energy price movements, which are highly correlated with natural gas price movements. The average ancillary service cost per MWh of load increased to $2.41 per MWh in 2011 compared to $1.26 per MWh in 2010, an increase of 91 percent. Total ancillary service costs increased from 3.2 percent of the load-weighted average energy price in 2010 to 4.5 percent in 2011.

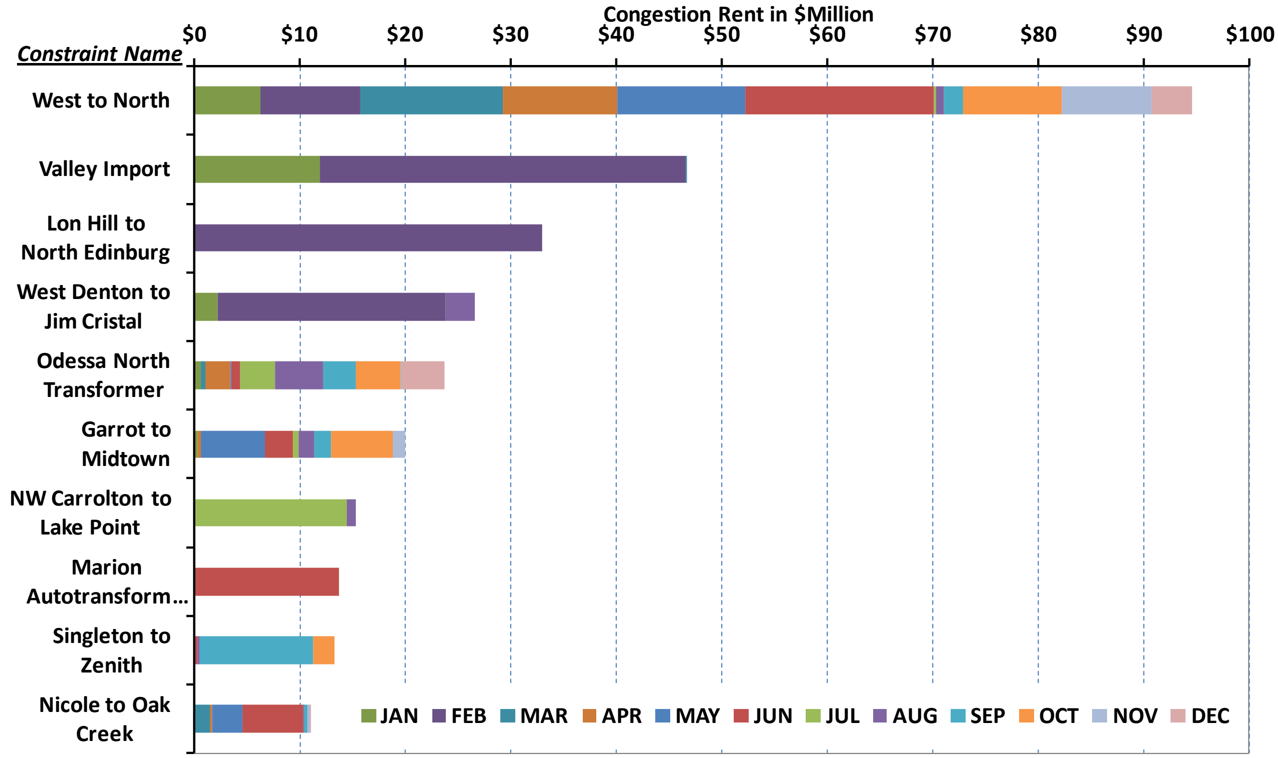


D. Transmission and Congestion

There were more than 300 different transmission constraints active at some point during real-time operations in 2011. The median financial impact of all these constraints, as measured by congestion rent, was approximately $300,000.

The figure below displays the ten most costly real-time constraints and indicates that the West to North interface constraint had the highest financial impact during 2011. The West to North interface constraint is very similar to the competitively significant constraint that existed since the inception of ERCOT's zonal market. Through the years it has been a major impediment to delivering all the wind generation located and produced in the western reaches of ERCOT to the load centers. The West to North interface constraint was the most frequently occurring constraint in 2011. It was active at some point during every month of 2011 and was binding more than 20 percent of the time in 2011.

Top Ten Real-Time Constraints

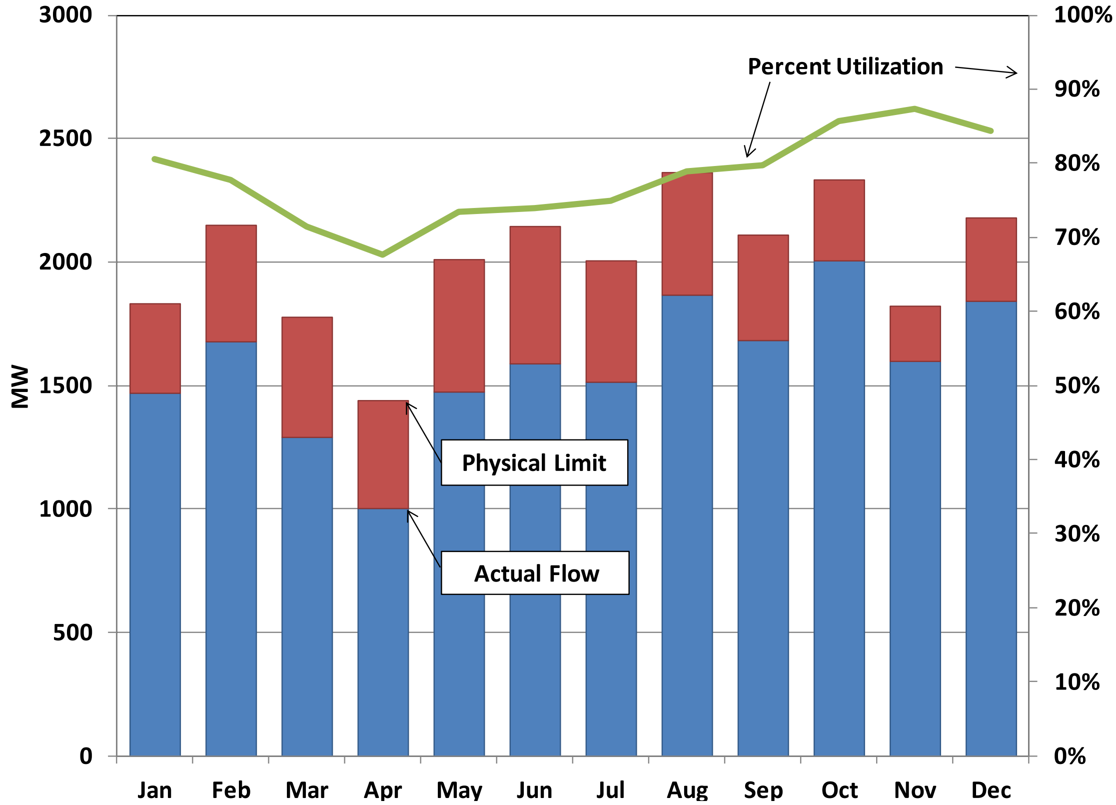


Two additional constraints on the list are also related to west zone wind generation, although in different directions. The Nicole to Oak Creek constraint is a small capacity 69 kV transmission line that typically overloads under high wind conditions, while due to its load serving nature, the Odessa North 138/69 kV transformer typically overloads under low wind conditions.

The second and third constraints shown in the figure are similar and reflect limitations on the amount of electricity that can be reliably imported into the Rio Grande Valley. This was most notable during the cold weather event of early February. Whereas system wide generation shortages were limited to February 2^{nd}, extremely high customer demands for electricity coupled with the extended planned outage of local generation led to shortages and resulting load curtailments in the Valley over the next two days. Constraints limiting imports to the Valley were active and not able to be resolved for a total of 13 hours during January and February.

To maximize the economic use of scarce transmission capacity, the ideal outcome would be for the actual transmission line flows to reach, but to not exceed the physical limits required to maintain reliable operations.

Utilization of the West to North Interface Constraint



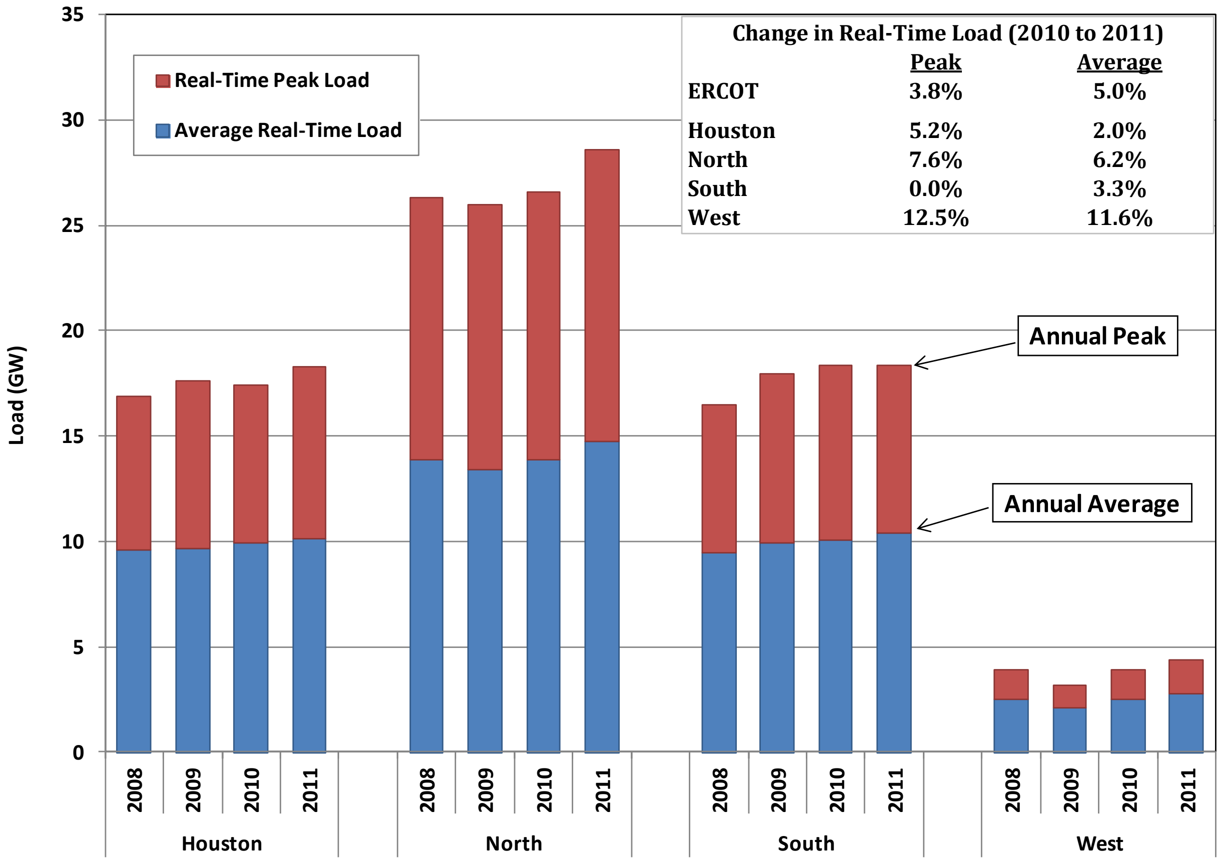
The figure above presents a summary of the utilization of the most active transmission constraint during 2011, the West to North interface. Its average utilization is determined by comparing the actual flow with the physical limit of the constraint for each real-time dispatch interval it was binding. Although there was significant variation throughout the year, the average physical limit was slightly less than 2,000 MW and the average actual flow during constrained intervals was approximately 1,500 MW. The average annual utilization of 76 percent compares favorably to 64 percent utilization experienced during the final months of the zonal market. Even more encouraging is the upward trend in utilization observed in the latter part of the year. This increase may be attributed to increased operator confidence that generators, specifically wind generators in this case, will reduce their output as expected when the constraint is active.

There should be opportunity for increased limits in the short term and even higher utilization of this constraint as ERCOT implements more sophisticated real-time analysis of this constraint, rather than relying on off-line studies. Over the long term, the physical limit will increase as CREZ transmission projects are completed.

E. Load and Generation

This figure below shows peak load and average load in each of the ERCOT zones from 2008 to 2011. In each zone, as in most electrical systems, peak demand significantly exceeds average demand. The North Zone is the largest zone (with about 39 percent of the total ERCOT load); the South and Houston Zones are comparable (27 percent) while the West Zone is the smallest (7 percent of the total ERCOT load). The figure also shows the annual non-coincident peak load for each zone. This is the highest load that occurred in a particular zone for one hour during the year; however, the peak can occur in different hours for different zones. As a result, the sum of the non-coincident peaks for the zones was greater than the annual ERCOT peak load.

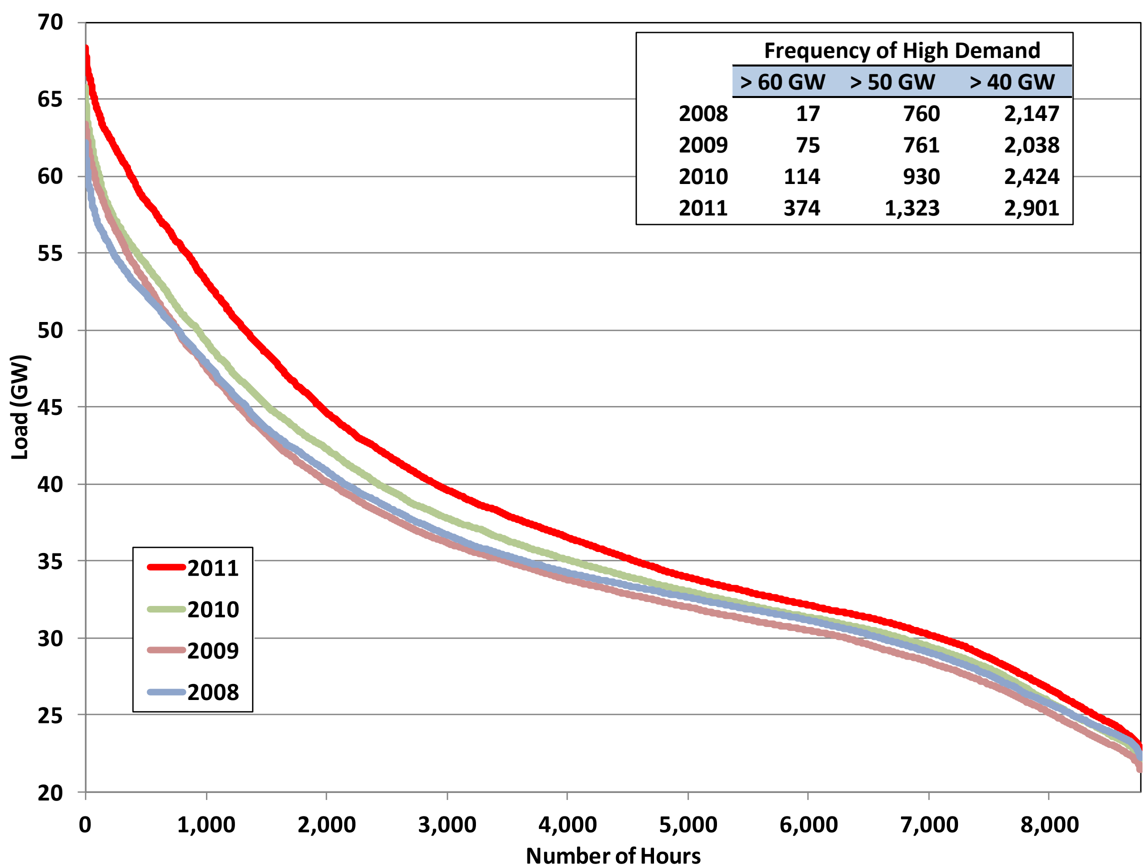
Annual Load Statistics by Zone



Total ERCOT load increased from 319 TWh in 2010 to 335 TWh in 2011, an increase of 5.0 percent or an average of approximately 1,800 MW every hour. Similarly, the ERCOT coincident peak hourly demand increased from 65,776 MW in 2010 to 68,379 MW, an increase of roughly 2,600 MW, or 4.0 percent.

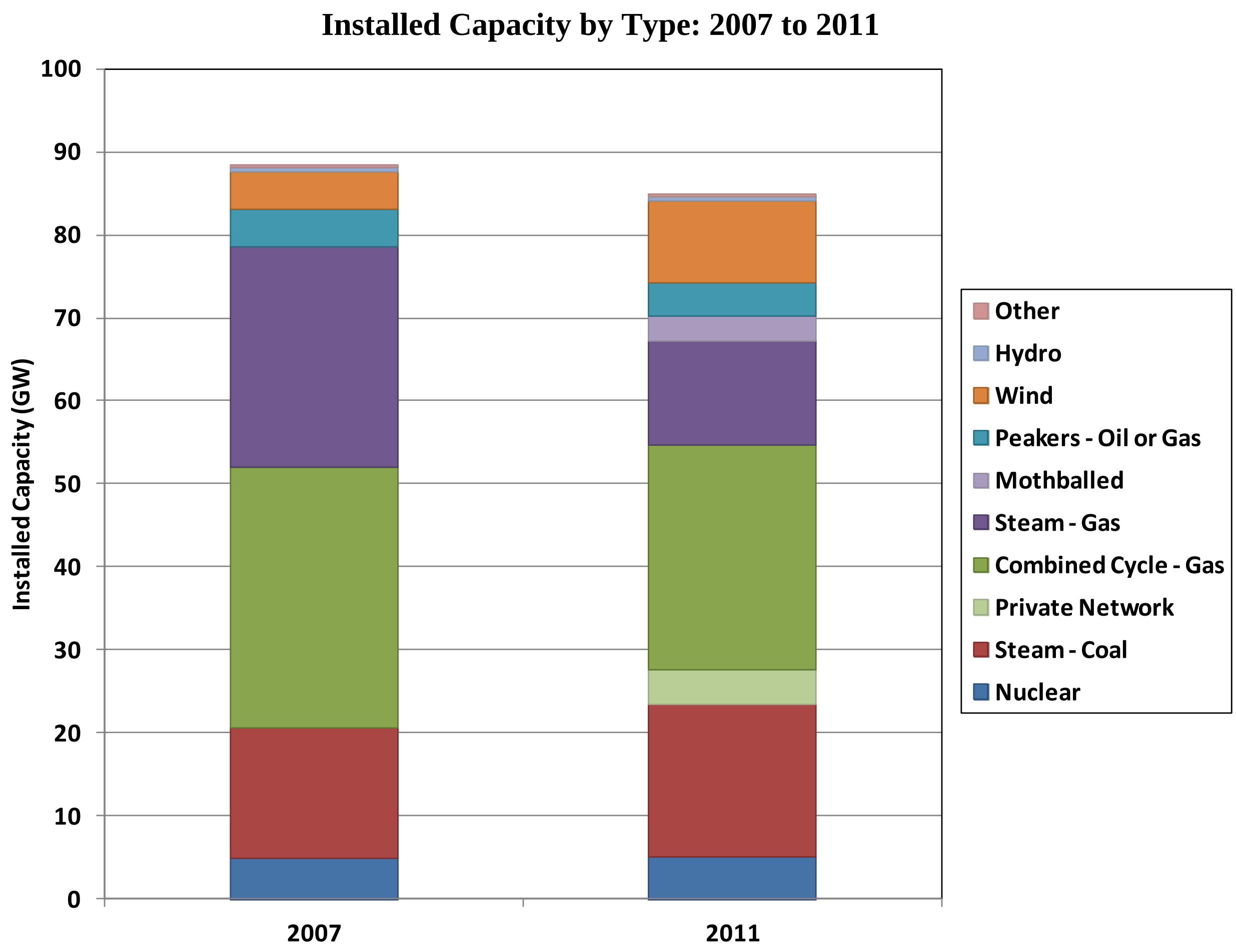
To provide a more detailed analysis of load at the hourly level, the next figure compares load duration curves for each year from 2008 to 2011. A load duration curve shows the number of hours (shown on the horizontal axis) that load exceeds a particular level (shown on the vertical axis). ERCOT has a fairly smooth load duration curve, typical of most electricity markets, with low to moderate electricity demand in most hours, and peak demand usually occurring during the late afternoon and early evening hours of days with exceptionally high temperatures.

Load Duration Curve – All hours



As shown in the figure above, the load duration curve for 2011 is significantly higher than in 2010 across all hours of the year. This is consistent with the aforementioned 5.0 percent load increase from 2010 to 2011.

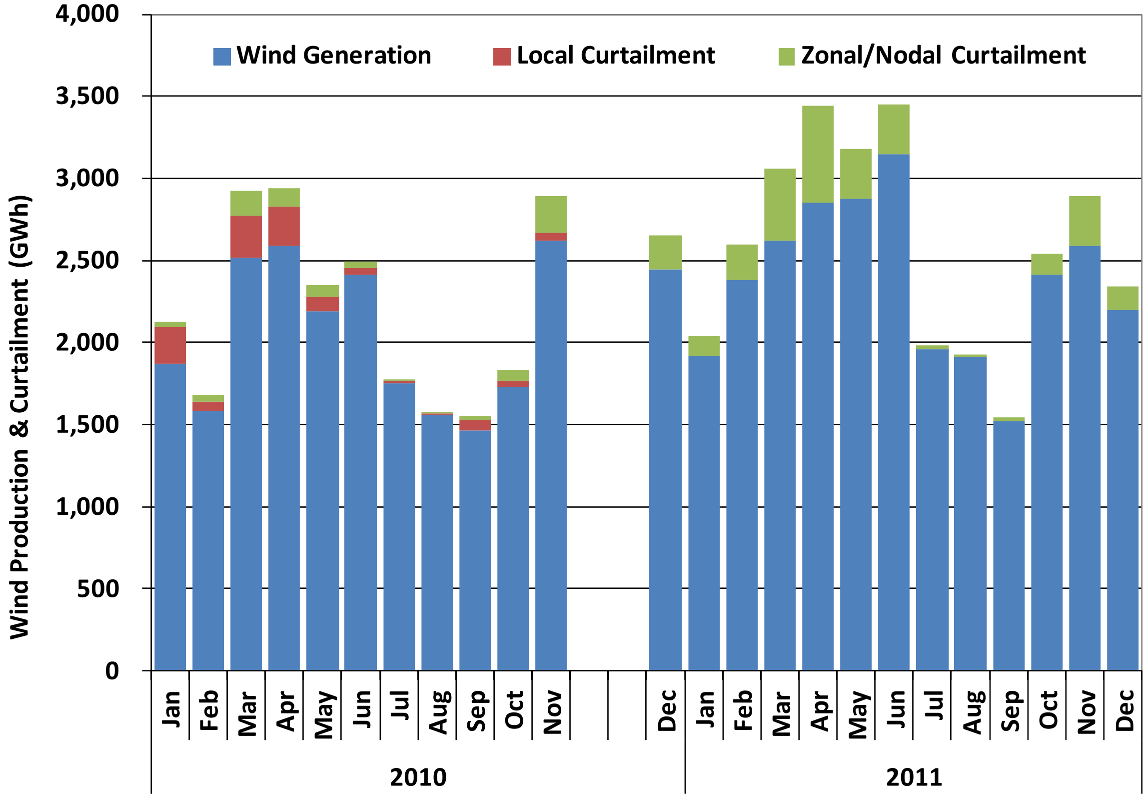
Although there were very few new units placed in service during 2011, by comparing the current mix of installed generation capacity to that in 2007, as shown in the figure below, the effects of longer term trends may be observed.



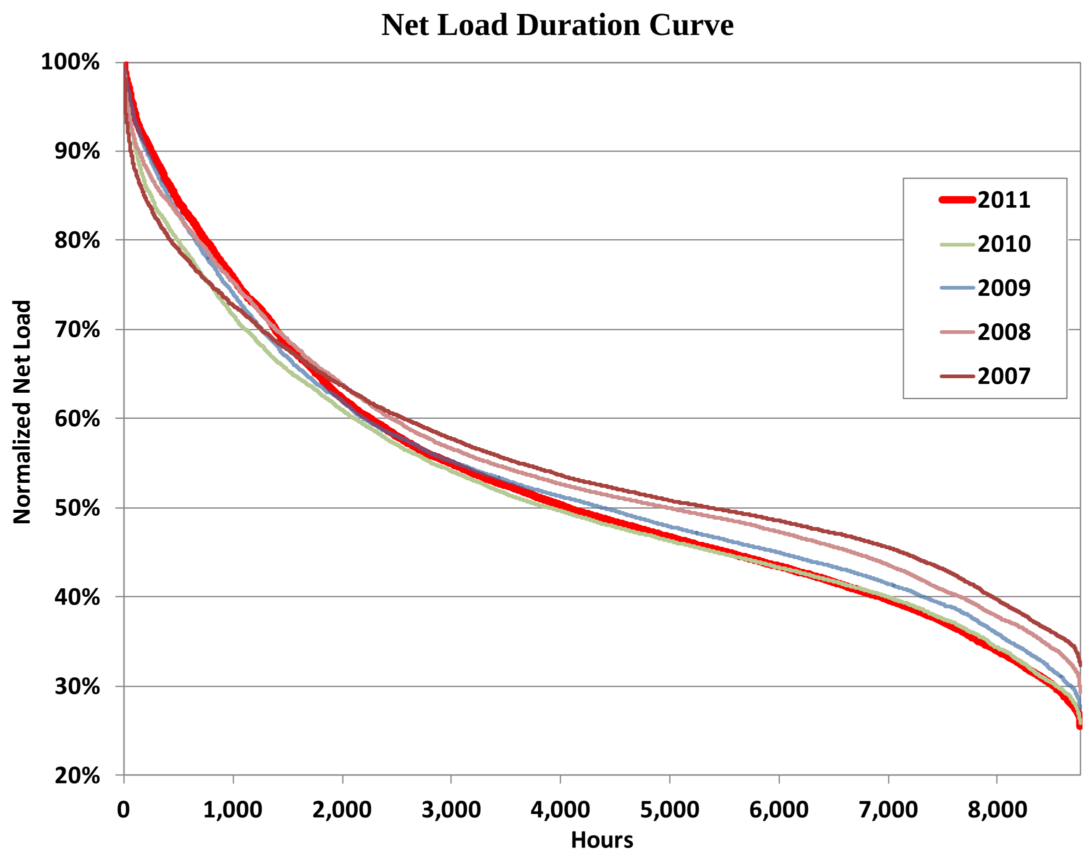
Over these five years wind and coal generation are the only two categories with increased capacity. However, the sizable additions in these two categories have been more than offset by retirements of natural gas fueled steam units, resulting in less installed capacity in 2011 than there was in 2007.

The next figure 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.

Wind Production and Curtailment



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. The figure below shows the net load duration curves for 2008 through 2011, normalized as a percent of peak load. 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.

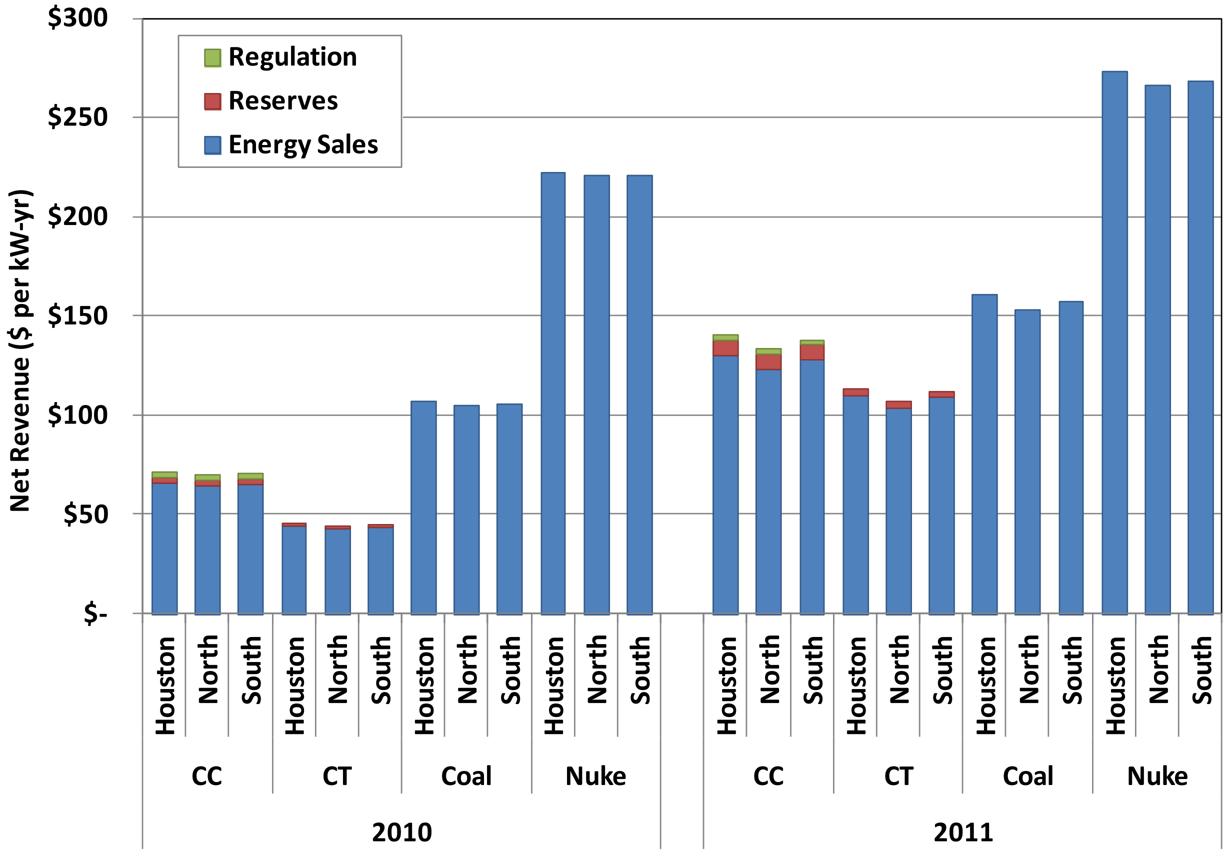


F. Resource Adequacy

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. 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.

The figure below 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. For coal and nuclear technologies, net revenue is calculated by assuming that the unit will produce at full output.

Estimated Net Revenue



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. Start-up costs and minimum running times and ramp restrictions, which can prevent the natural gas generators from profiting during brief price spikes, are not explicitly accounted for in the net revenue analysis. Despite these limitations, the net revenue analysis provides a useful summary of signals for investment in the wholesale market.

The figure above 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. 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 less than $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.

- For a new natural gas fueled combustion turbine, the estimated net revenue requirement is approximately $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, indicating that for the first time since 2008 that net revenues were sufficient to support new gas turbine generation.

- For a new natural gas fueled 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, again indicating that 2011 was the first time since 2008 that net revenues have been sufficient to support new combined cycle generation in ERCOT.

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.

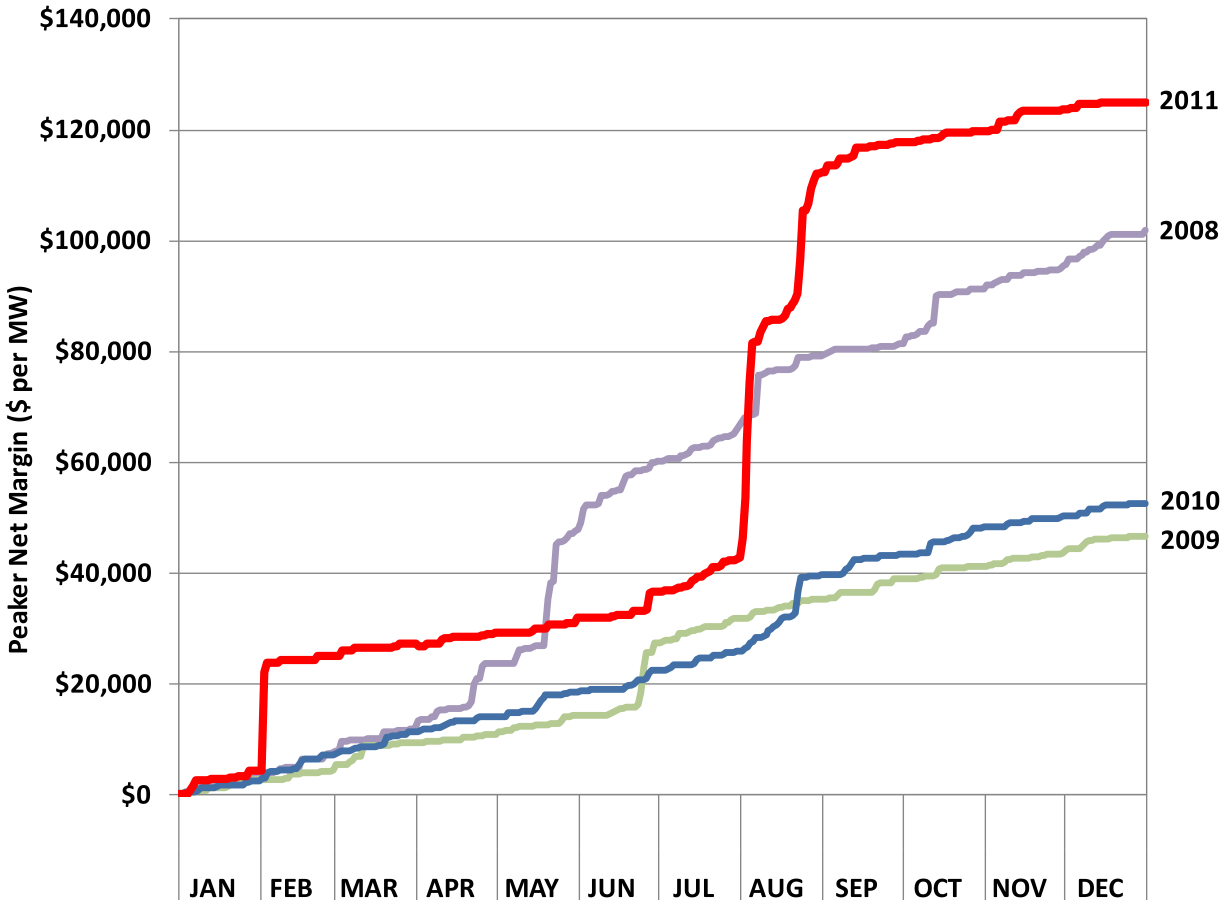
Scarcity Pricing Effectiveness

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, the system-wide offer cap is then reduced to the higher of $500 per MWh or 50 times the daily gas price index.

The next figure 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. 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 below and consistent with the previous findings in this section relating to net revenue, the PNM reached the level sufficient for new entry in 2011.

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. These improvements would address conditions that cause energy prices to understate the marginal costs of satisfying the real-time demand. In particular, real time energy prices do not fully reflect:

- The value of curtailed load when load resources are deployed;
- The value of reduced reliability when responsive reserves or non-spinning reserves have been converted to energy;
- The costs associated with starting and running the gas turbines (or other resources not dispatchable in the 5-minute energy dispatch) that were being deployed to meet demand.

After multiple protocol revisions are implemented in 2012, real-time energy price formation will be improved, but the non-spinning reserve deployment process remains 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 a "look ahead" dispatch functionality for the real-time market to produce an energy and ancillary services commitment and dispatch results that are co-optimized and recognize anticipated changes in system demands. 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.

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.

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, 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/or demand resources result in maintaining the minimum required installed reserve margin.

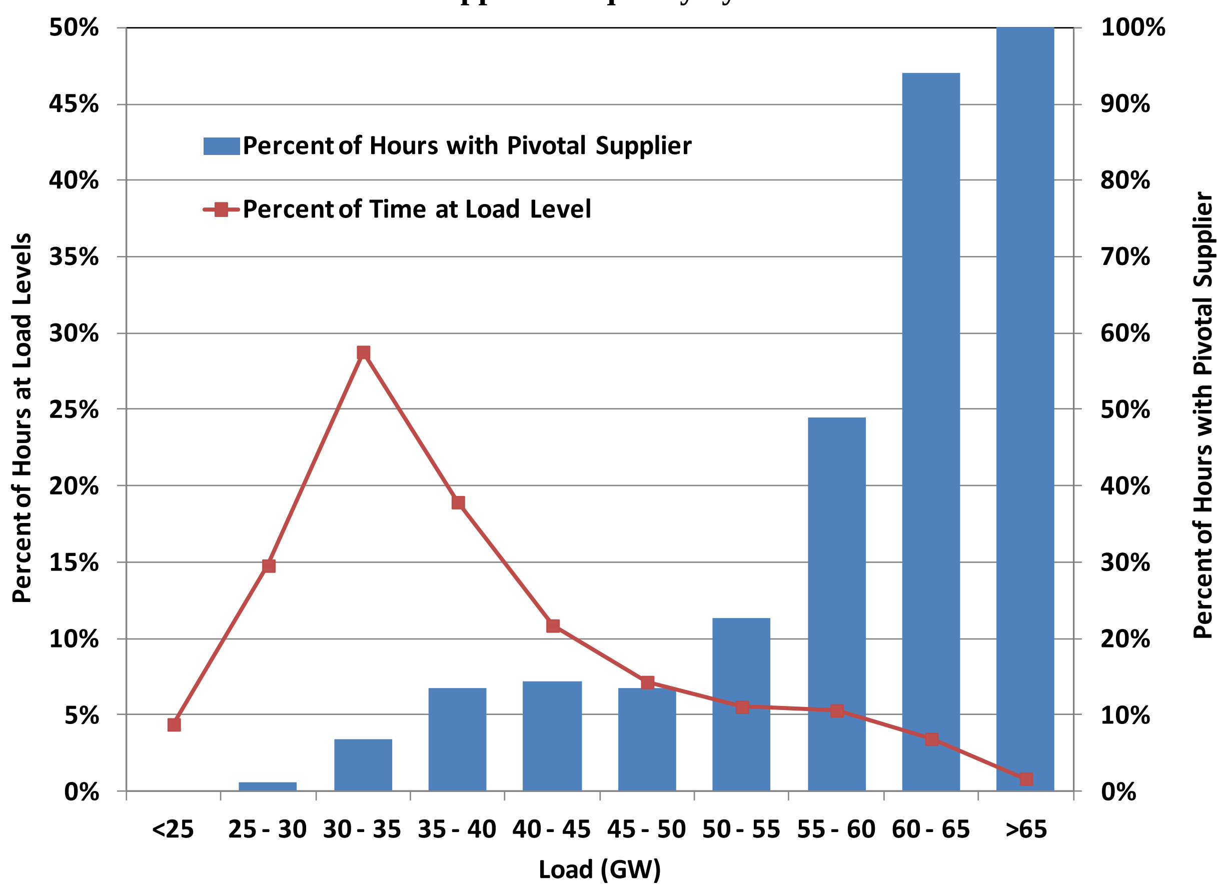
G. Analysis of Competitive Performance

The report evaluates market power from two perspectives, structural and behavioral. The Residual Demand Index ("RDI") is used to analyze market structure. 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 owned by other suppliers. When the RDI is greater than zero the largest supplier is pivotal; that is, 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 power. However, it does identify conditions under which a supplier would have the ability to raise prices significantly by withholding resources.

The figure below 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.

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.

The behavioral aspects of market power abuse are evaluated 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.

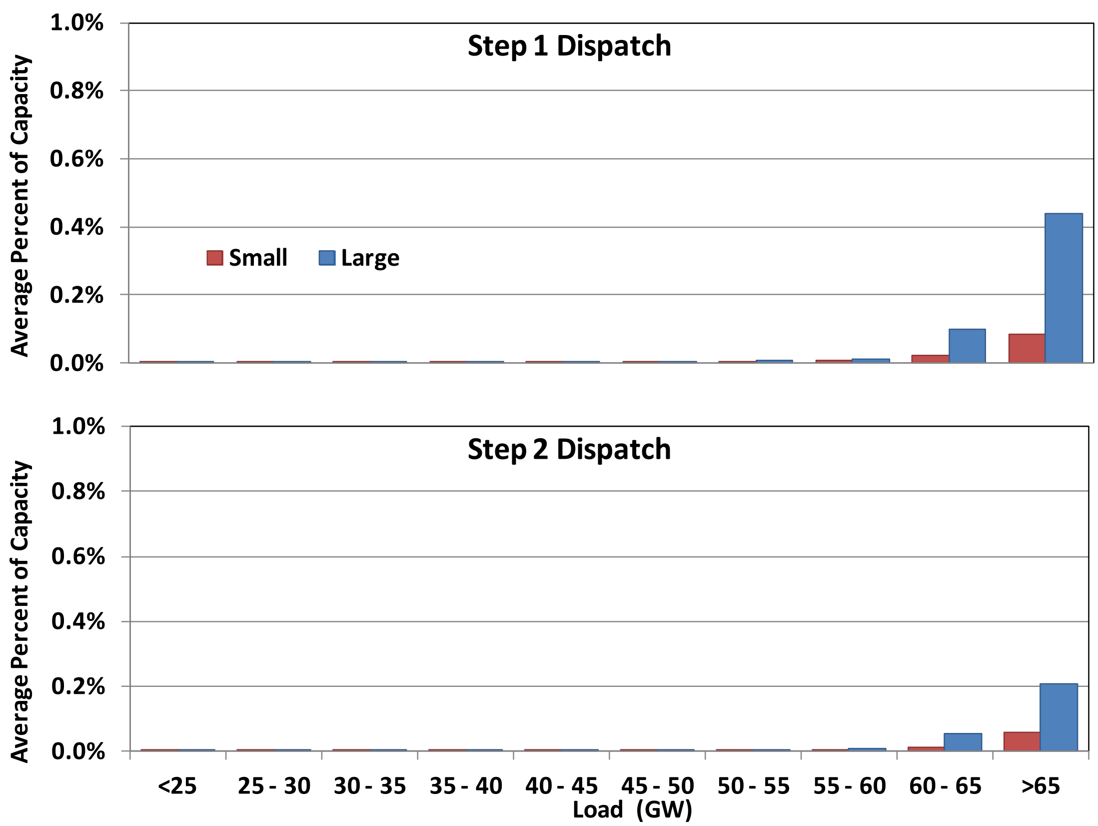
Resources are considered for inclusion in the output gap when they are committed and producing at less than full output. Energy not produced from committed resources is included in the output gap if the real-time energy price exceeds by at least $50 per MWh that unit's mitigated offer cap which serves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

The ultimate output gap is 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.

Incremental Output Gap by Load Level and Participant Size



The figure above shows the magnitude of the output gap to be very small, even at the highest load levels, for both steps in the dispatch process. These small quantities raise no competitive concerns. In summary, we find that the ERCOT nodal wholesale market performed competitively in 2011.

H. Nodal Market Performance and Recommendations

As discussed in prior ERCOT State of the Market Reports, implementation of the nodal market was expected to provide the following improvements:

- Fundamental improvements in ERCOT's ability to efficiently manage transmission congestion, which is one of the most important functions in electricity markets.
- The nodal market will enable all transmission congestion to be managed through market-based mechanisms
- The nodal market will provide better incentives to market participants, facilitate more efficient commitment and dispatch of generation, and improve ERCOT's operational control of the system.
- The use of unit-specific dispatch in the nodal market will allow ERCOT to more fully utilize generating resources than the zonal market, which frequently exhibited price spikes even when generating capacity was not fully utilized.
- The nodal market will allow ERCOT to increase the economic and reliable utilization of scarce transmission resources well beyond that attainable in the zonal market.
- The nodal market will significantly improve the ability to efficiently and reliably integrate the ever-growing quantities of intermittent resources, such as wind and solar generating facilities.
- The nodal market will produce price signals that better indicate where new generation is most needed (and where it is not) for managing congestion and maintaining reliability.

In the long-term, these enhancements to overall market efficiency should translate into substantial savings for consumers. This report reviews the first year of nodal market operations, highlights the areas of expected improvements that have been observed in the first year, documents areas of unanticipated outcomes during the nodal transition, and provides recommendations for future improvements to the nodal market.

Overall pricing outcomes from the nodal real-time market have met expectations for improved efficiency. The discussion of Figure 11, Figure 13, and Figure 14 on pages 12 and 14-15 describes how prices are much more appropriately correlated with load level in the nodal market than they were in the zonal market. Section V.B, Effectiveness of the Scarcity Pricing Mechanism , specifically at page 84, provides more details about the improved pricing during shortage conditions, now that scarcity pricing is no longer dependent upon the offers from participants with small generator fleets. The nodal market has also enabled the higher utilization of transmission facilities as described in the discussion of Figure 36, on page 45.

Three areas where the nodal market implementation led to unanticipated outcomes were identified and quickly resolved in 2011. The calculation of real-time settlement point prices every 15 minutes at resource node locations originally included weighting the price from each dispatch interval by the dispatch level (base point) of the resource. This led to price differences between locations when there was no transmission congestion. These price differences would have resulted in payments and charges to owners of Point-to-Point Obligations and Congestion Revenue Rights settled in real-time which were not supported by real-time congestion rent and would have required uplifted payments to support. The base point weighting factor was removed with the implementation of NPRR 326.

As described in Section III.A, Real-Time Constraints at page 46, transmission constraint and base point oscillations were observed during the spring of 2011. After ERCOT modified their constraint management software and started providing the curtailment flag to wind generators, as required under NPRR 285, there have been no more occurrences of constraint oscillation.

The last area of unanticipated outcomes has to do with the modeling of the transmission system and the impact that de-energized elements had on locational prices. Shortly after the implementation of the nodal market, it was determined that when particular generation resources were offline, according to the established pricing rules the real-time price at that location was set using a system-wide value. This created inconsistent pricing between the day-ahead and real-time markets, allowing participants to acquire certain Point-to-Point Obligations for low, or no cost in the day-ahead market and receive payment because there were real-time price differences. In February 2011, ERCOT improved their network model by adding hundreds of transmission system elements at 140 locations. This model improvement, combined with NPRR343 which precludes parties from buying Point-to-Point Obligations between electrically similar locations, has greatly reduced the potential for this type of inefficient trading activity. However, under certain combinations of transmission equipment outages similar price discrepancies can occur.

In conjunction with any market design changes that may result from the current PUCT proceedings related to resource adequacy, we recommend improvements to two aspects of the nodal market design.

1. We recommend a change to the automated mitigation procedures that are part of the real-time dispatch to eliminate the occurrences of over-mitigation we have observed. As more fully described in Section VI.C, Mitigation at page 107, we support introducing a test to determine whether a unit is either contributing to, or helping to resolve a transmission constraint and only subject the relieving units to mitigation.

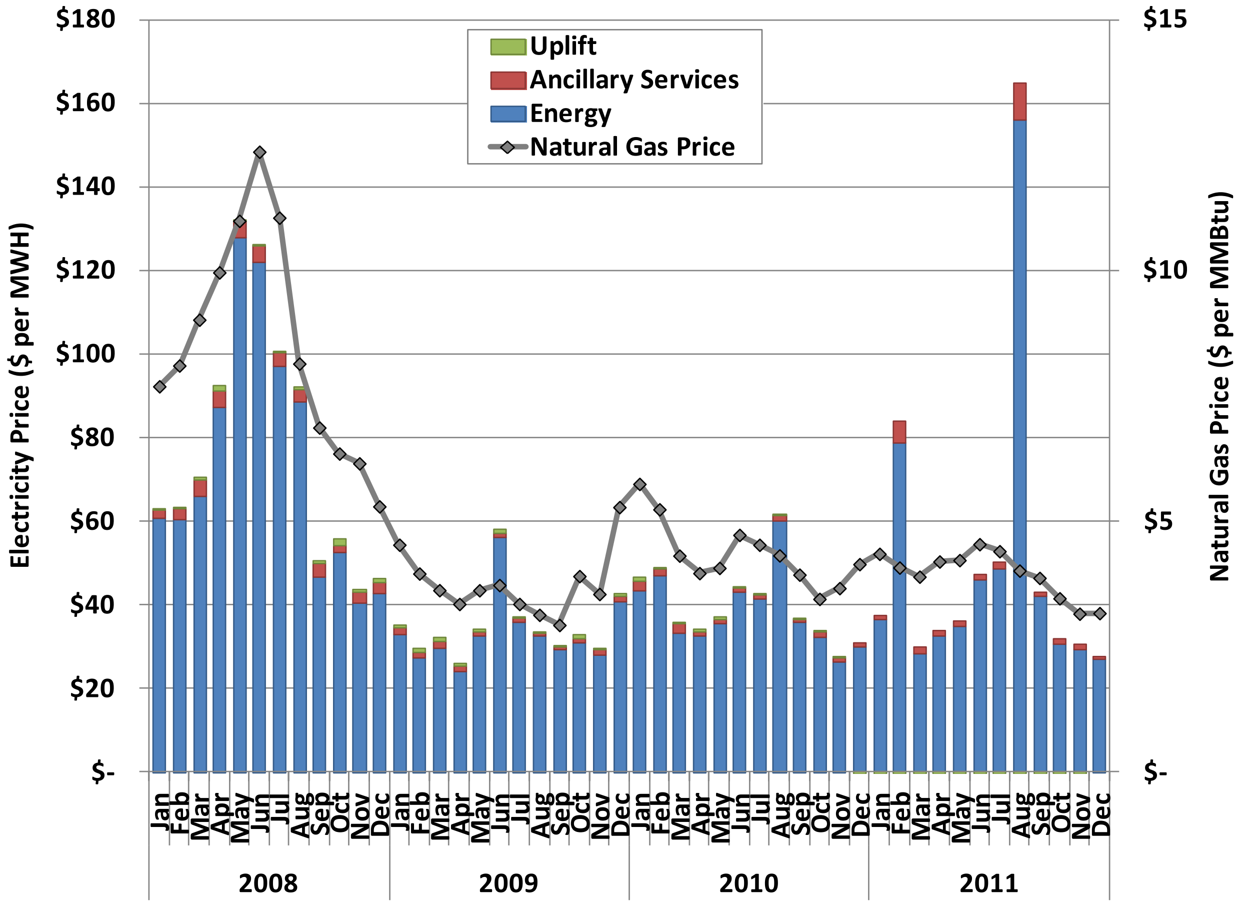
2. We recommend a change to the real-time market software to allow it to"look ahead" a sufficient amount of time to better commit load and generation resources that can be online within 30 minutes. More discussion of this topic can be found starting on page 86 in Section V.B, Effectiveness of the Scarcity Pricing Mechanism.

I. Review of Real-Time Market Outcomes

A. Real-Time Market Prices

Our first analysis evaluates the total cost of supplying energy to serve load in the ERCOT wholesale market. In addition to the costs of energy, loads incur costs associated with ancillary services and a variety of non-market based expenses referred to as "uplift". We have calculated an average all-in price of electricity for ERCOT that is intended to reflect wholesale energy costs as well as these additional costs.

Figure 1: Average All-in Price for Electricity in ERCOT



Energy, ancillary services and uplift costs are the three components in the all-in price of electricity. The ERCOT wide price is the load weighted average of the real-time market prices from all load zones. Prior to ERCOT's conversion to the nodal market in December 2010, energy costs were determined from the zonal balancing energy market. Ancillary services costs are estimated based on total system demand and prices in the ERCOT markets for regulation, responsive reserves, and non-spinning reserves. Uplift costs are assigned market-wide on a load-ratio share basis to pay for charges associated with additional reliability unit commitment and any reliability must run contracts.[1]

Figure 1 shows the monthly average all-in price for all of ERCOT from 2008 to 2011 and the associated natural gas price. With the noticeable exception of February and August last year, Figure 1 indicates that natural gas prices were a primary driver of the trends in electricity prices from 2008 to 2011. Again, this is not surprising given that natural gas is a widely-used fuel for the production of electricity in ERCOT, especially among generating units that most frequently set locational marginal prices in the nodal market. As discussed later, the high prices in February and August were the result of extreme weather conditions leading to generation scarcity.

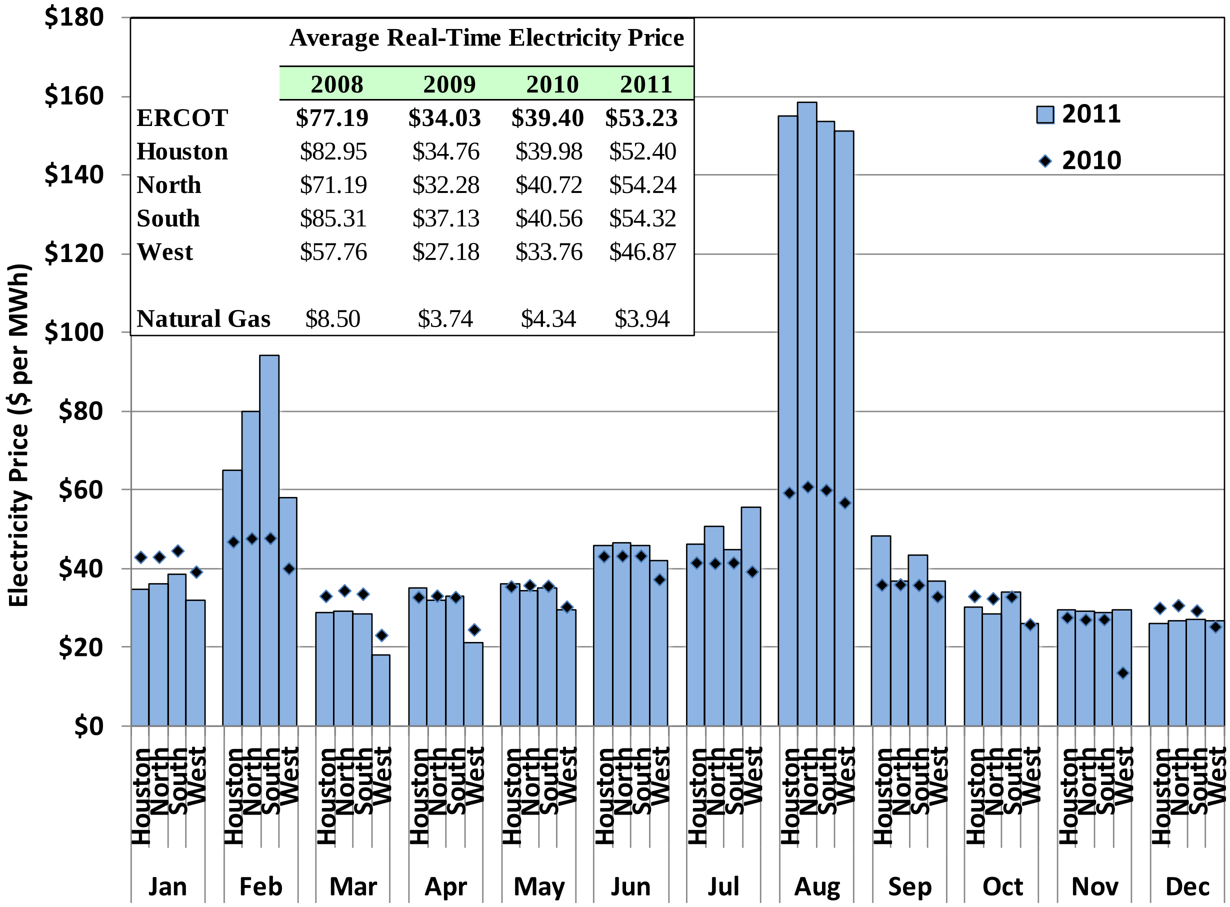
The largest component of the all-in cost of wholesale electricity is the energy cost, which is reflected by the locational marginal prices. As is typical in other wholesale markets, only a small share of the power produced in ERCOT is transacted in the spot market. However, the pricing outcomes in the real-time energy market are very important because they set the expectations for prices in the forward markets (including bilateral markets) where most transactions take place. Unless there are barriers preventing arbitrage of the prices between the spot and forward markets, the prices in the forward market should be directly related to the prices in the spot market (*i.e.*, the spot prices and forward prices should converge over the long-run). Hence, artificially low prices in the real-time energy market will translate to artificially-low forward prices. Likewise, price spikes in the real-time energy market will increase prices in the forward markets. This section evaluates and summarizes electricity prices in the real-time market during 2011.

To summarize the price levels during the past four years, Figure 2 shows the monthly load-weighted average prices in the four geographic ERCOT load zones. These prices are calculated by weighting the energy price for each interval and each zone by the total zonal load in that interval. Since December 2010 these prices were determined by the nodal real-time energy market. Prior prices were derived from the zonal balancing energy market. Load-weighted

[1] Prior to December 2010 uplift costs included charges for out-of-merit energy and capacity, replacement reserve services and any reliability must run contracts.

average prices are the most representative of what loads are likely to pay, assuming that real-time energy prices are, on average, generally consistent with bilateral contract prices.

Figure 2: Average Real-Time Energy Market Prices

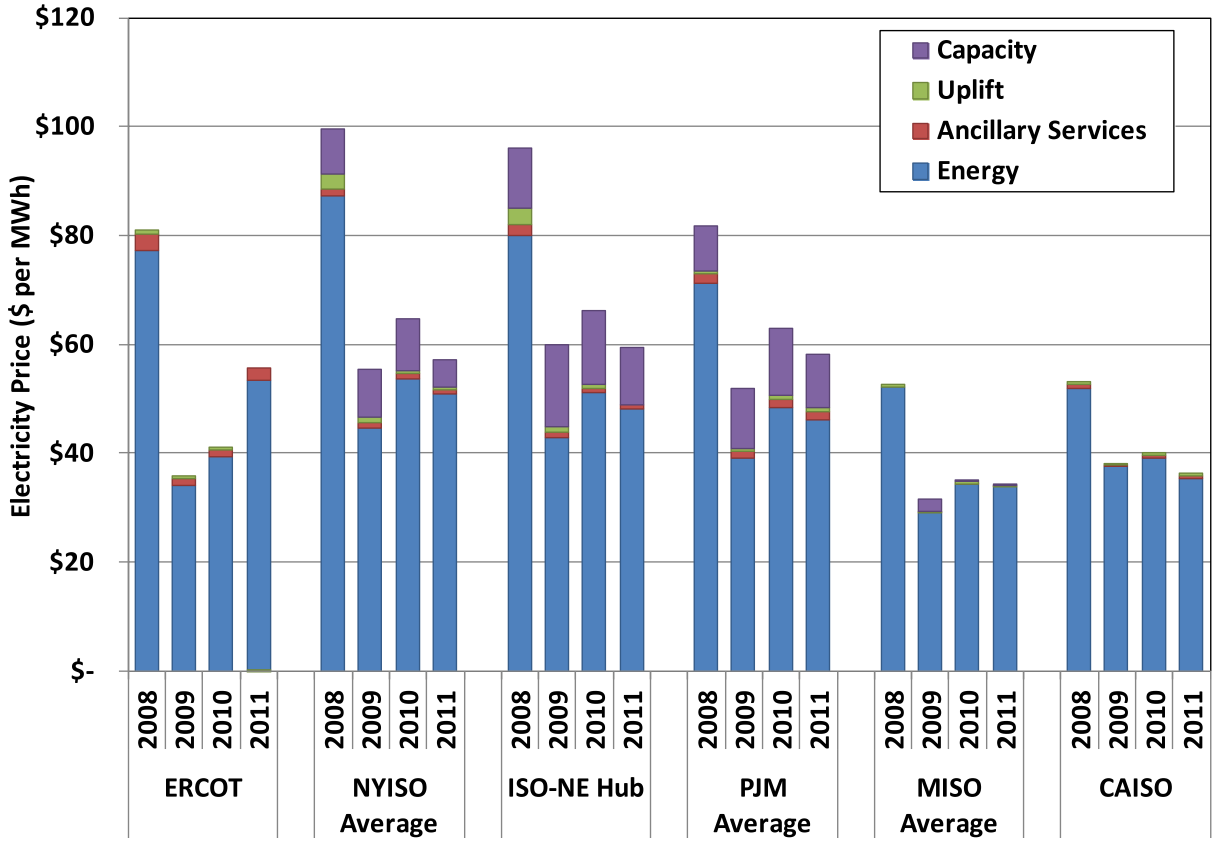


ERCOT average real-time market prices were 35 percent higher in 2011 than in 2010. The ERCOT-wide load-weighted average price was $53.23 per MWh in 2011 compared to $39.40 per MWh in 2010. February and August experienced the largest increases to real-time energy prices in 2011, averaging 67 and 160 percent higher than the prices in the same months in 2010. Price increases in both months were driven by extreme weather conditions which led to operating reserve deficiencies resulting in real-time energy prices reaching $3,000 per MWh for sustained periods of time.

The increase in real-time energy prices was partially offset by lower fuel prices in 2011. Natural gas prices decreased 9 percent in 2011, averaging $3.94 per MMBtu in 2011 compared to $4.34 per MMBtu in 2010. Although lower natural gas prices contributed to lower real-time energy prices in many hours, these reductions were smaller than the price effects of the shortages in February and August.

To provide additional perspective on the outcomes in the ERCOT market, our next analysis compares the all-in price metrics for ERCOT and other electricity markets. The following figure compares the all-in prices in ERCOT with other organized electricity markets in the U.S.: New York ISO, ISO New England, PJM, Midwest ISO, and California ISO.

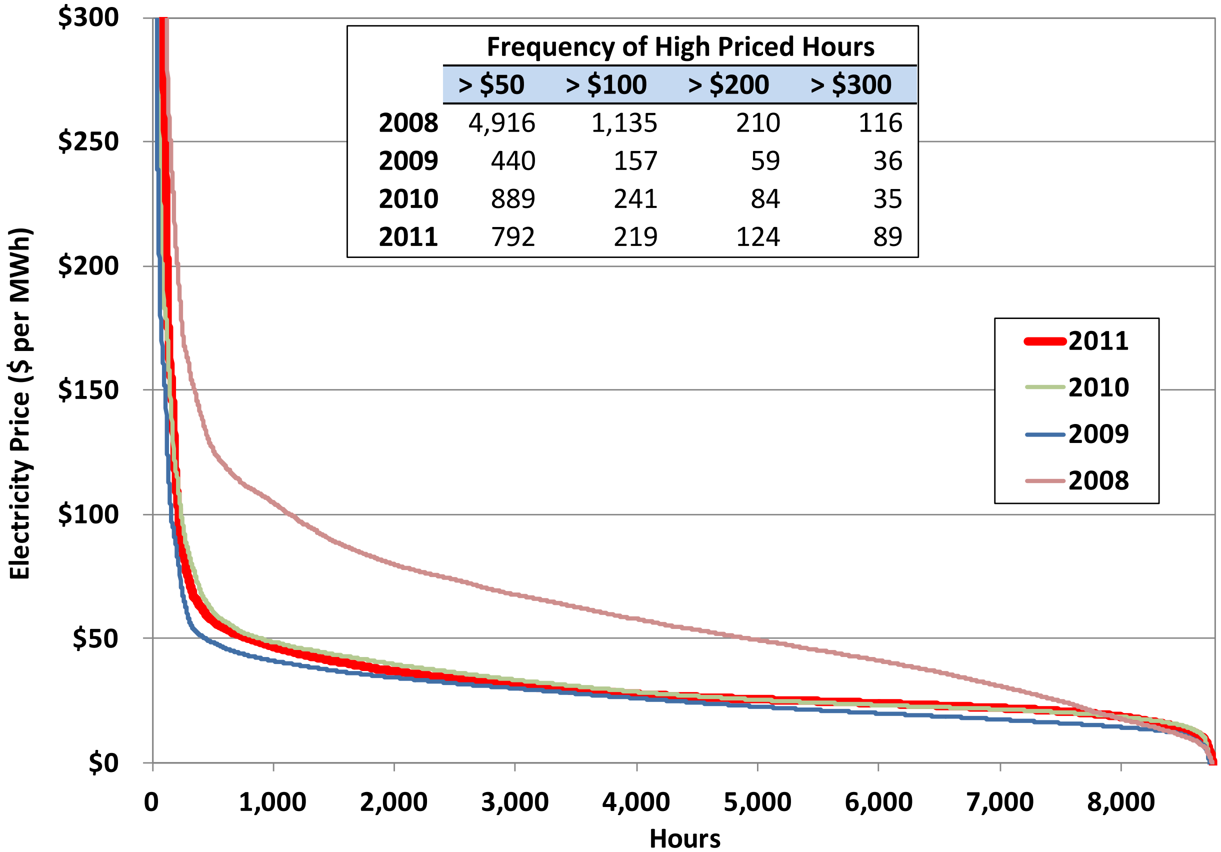
Figure 3: Comparison of All-in Prices across Markets



For each region, the figure reports the average cost (per MWh of load) for energy, ancillary services (reserves and regulation), capacity markets (if applicable), and uplift for economically out-of-merit resources. Figure 3 shows that ERCOT all-in prices in 2011 were on par with the all-in prices from the other markets with centralized capacity markets. As discussed in more detail in Section V.A, Net Revenue Analysis, after two years of inadequate prices signals, ERCOT energy prices in 2011 rose to levels to support much needed new supply.

Figure 4 presents price duration curves for ERCOT energy markets in each year from 2008 to 2011. A price duration curve indicates the number of hours (shown on the horizontal axis) that the price is at or above a certain level (shown on the vertical axis). The prices in this figure are the hourly load-weighted zonal balancing energy price for the zonal market and hourly load-weighted nodal settlement point price for the nodal market.[2]

Figure 4: ERCOT Price Duration Curve



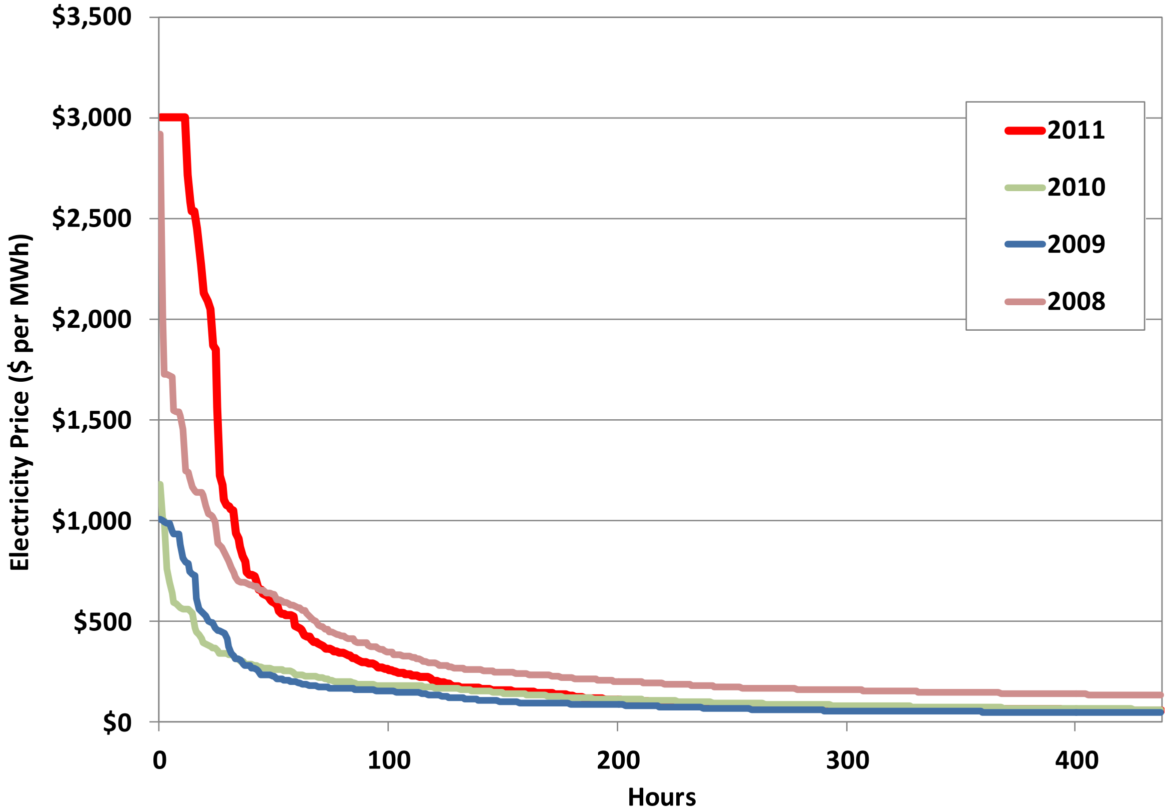
In Figure 4 we can see the impact of much higher natural gas prices experienced in 2008, leading to higher energy prices across the vast majority of hours in that year. In contrast, with similar levels of natural gas prices for the past three years, the price duration curves for 2009 – 2011 are remarkably close for most of the year.

To see the where the prices during 2011 were much different than in the previous two years, we present Figure 5, which compares prices for the highest five percent of hours. In 2011, energy prices for the top 100 hours were significantly higher than in the past two years. It is this small

2 ERCOT switched to a nodal market on December 1, 2010. The December nodal prices are included in the 2010 price duration curve.

number of very high priced hours which is the primary driver of higher average energy prices in 2011.

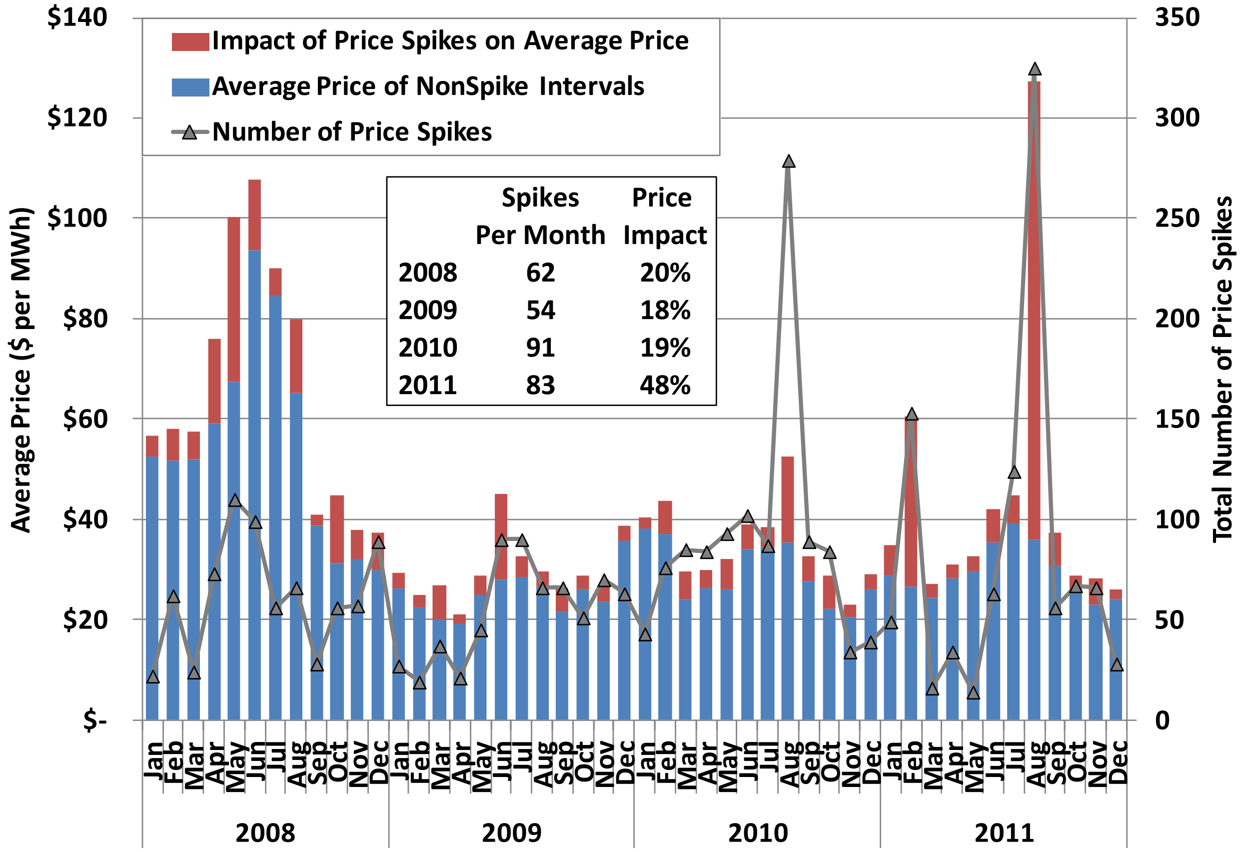
Figure 5: ERCOT Price Duration Curve – Top 5% of Hours



To better observe the effect of the highest-priced hours, the following analysis focuses on the frequency of price spikes in the real-time energy market since December 2010. Prior information was from the zonal balancing energy market. Figure 6 shows the average price and the number of price spikes in each month. For this analysis, price spikes are defined as intervals where the load-weighted average energy price in ERCOT is greater than 18 MMBtu per MWh times the prevailing natural gas price. Prices at this level should exceed the marginal costs of virtually all of the on-line generators in ERCOT.

The number of price spike intervals during 2011 was 83 per month, a decrease from the 91 per month in 2010. However, just looking at the average can be misleading. Comparing the monthly details of 2011 with 2010 we see that for most months there were much fewer price spike intervals in 2011, likely due to the improved efficiencies of the nodal market. The noticeable exceptions were the months of February and August.

Figure 6: Average Real-Time Energy Prices and Number of Price Spikes

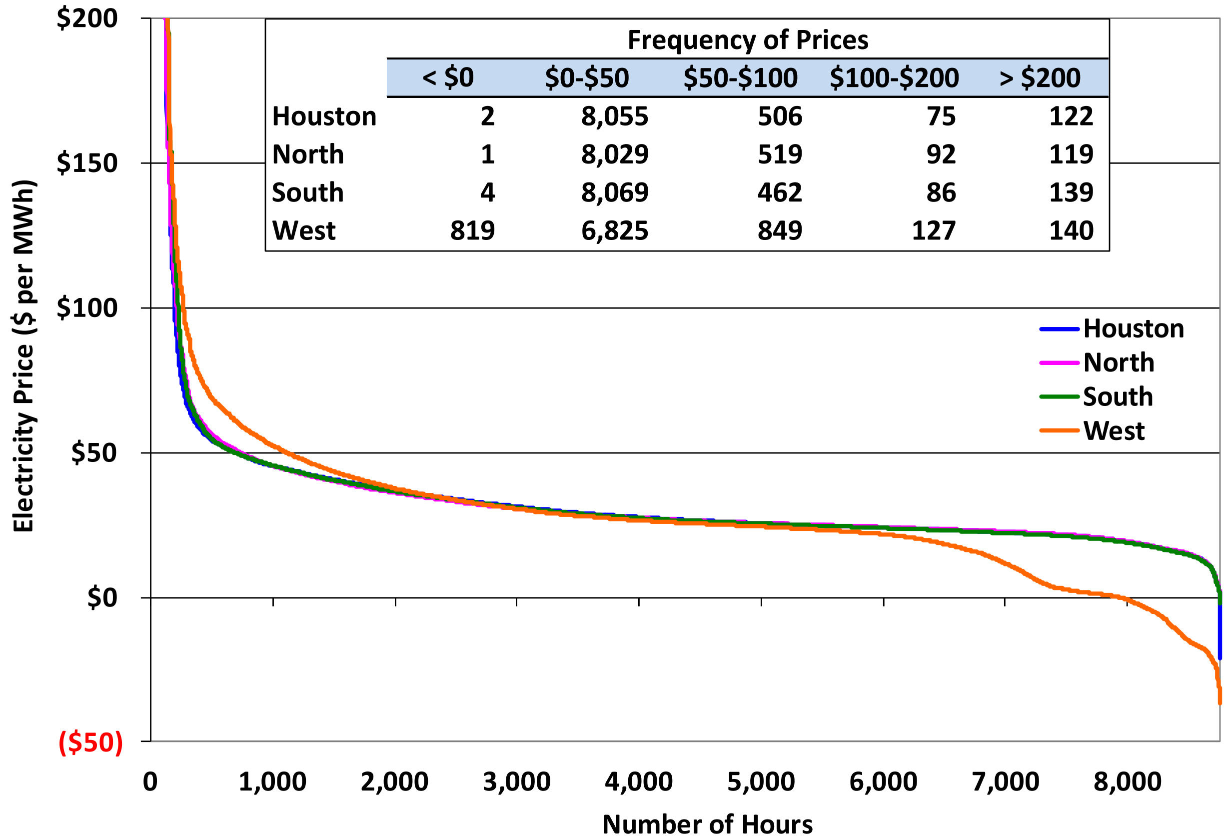


	Spikes Per Month	Price Impact
2008	62	20%
2009	54	18%
2010	91	19%
2011	83	48%

To measure the impact of these price spikes on average price levels, the figure also shows the average prices with and without the price spike intervals. The top portions of the stacked bars show the impact of price spikes on monthly average price levels. Prior to 2011, the impact grew with the frequency of the price spikes, averaging $10.71, $4.67 and $5.53 per MWh during 2008, 2009 and 2010, respectively. However, in 2011 the impact on average energy price was $14.09 per MWh, or 48 percent of the annual average price. This increased impact of the price spikes is a direct result of the improved mechanism for pricing real-time energy during scarcity, as discussed in more detail in Section V.B, Effectiveness of the Scarcity Pricing Mechanism.

To depict how real-time energy prices vary by hour in each zone, Figure 7 shows the hourly average price duration curve in 2011 for four ERCOT load zones. The Houston, North and South load zones had similar prices over the majority of hours. The price duration curve for the West Zone is noticeably different than the other zones, with more hours with prices greater than $50 per MWh and over 800 hours (9 percent of the time) when the average hourly price was less than zero. The relatively low prices in the West zone are generally caused by high wind output in the West that frequently results in severe congestion on transmission interfaces from the West zone to the other zones in ERCOT. The relatively higher prices in the West zone are caused by local transmission constraints that typically occur under low wind and high load conditions. Specifics about these transmission constraints are provided in Section III, Transmission and Congestion.

Figure 7: Zonal Price Duration Curves



	< $0	$0-$50	$50-$100	$100-$200	> $200
Houston	2	8,055	506	75	122
North	1	8,029	519	92	119
South	4	8,069	462	86	139
West	819	6,825	849	127	140

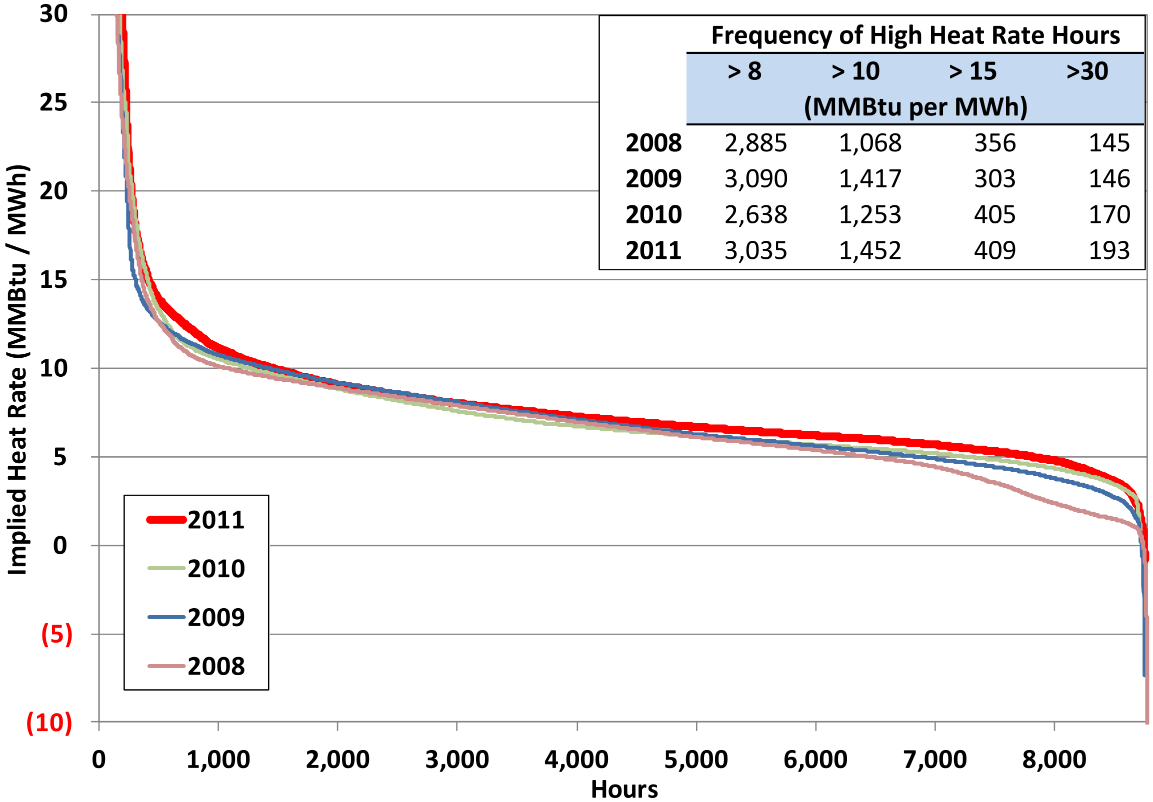
B. Real-Time Prices Adjusted for Fuel Price Changes

Although real-time electricity prices are driven to a large extent by changes in fuel prices, natural gas prices in particular, they are also influenced by other factors.

To clearly identify changes in electricity prices that are not driven by changes in natural gas prices, Figure 8 and Figure 9 show the load weighted, hourly average real-time energy price adjusted to remove the effect of natural gas price fluctuations. The first chart shows a duration curve where the real-time energy price is replaced by the marginal heat rate that would be implied if natural gas was always on the margin.[3]

[3] The *Implied Marginal Heat Rate* equals either the *Balancing Energy Price* (zonal) or the *Real-Time Energy Price* (nodal) divided by the *Natural Gas Price*. This methodology implicitly assumes that electricity prices

Figure 8: Implied Marginal Heat Rate Duration Curve – All hours

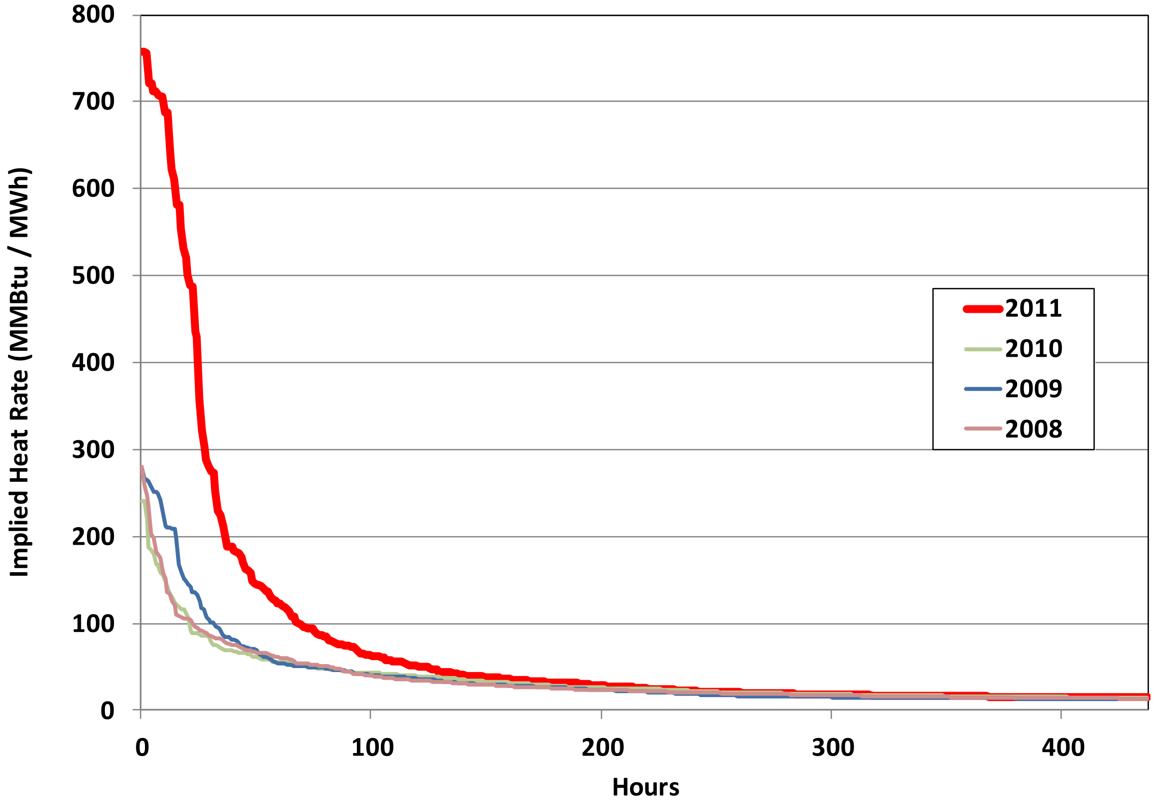


Frequency of High Heat Rate Hours	> 8	> 10	> 15	>30
	(MMBtu per MWh)			
2008	2,885	1,068	356	145
2009	3,090	1,417	303	146
2010	2,638	1,253	405	170
2011	3,035	1,452	409	193

The second chart shows the same duration curves for the five percent of hours in each year with the highest implied heat rate. Both figures show duration curves for the implied marginal heat rate for 2008 to 2011. Similar to Figure 4, Figure 8 shows that the implied marginal heat rates were relatively consistent across the majority of hours from 2008 to 2011. The implied heat rate during 2011 was somewhat higher for most hours, when compared to 2010. This can be explained by the much higher loads experienced throughout 2011. There were 193 hours during 2011 when the implied heat rate was greater than 30 MMBtu per MWh, compared to 145, 146, and 170 hours in 2008, 2009, and 2010, respectively. This indicates that there are price differences that are due to factors other than changes in natural gas prices.

move in direct proportion to changes in natural gas prices.

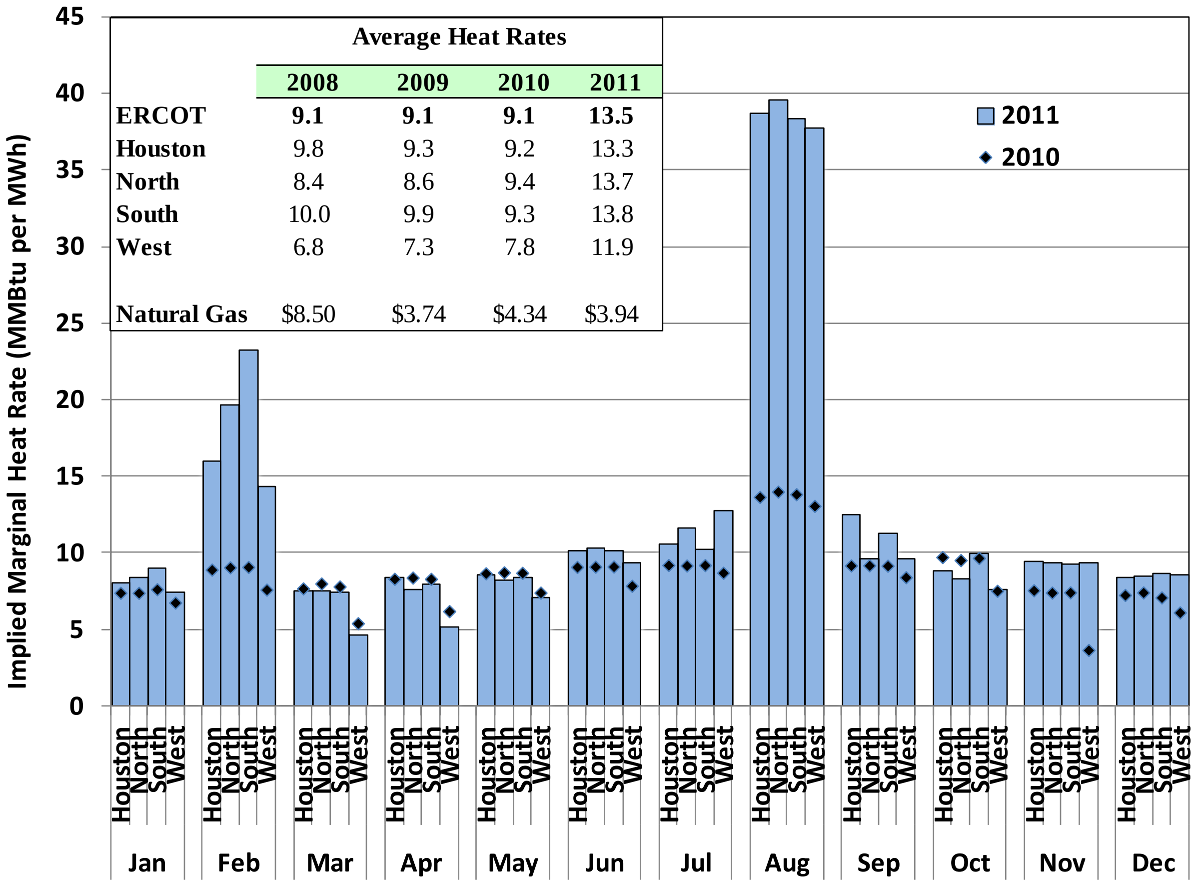
Figure 9: Implied Marginal Heat Rate Duration Curve – Top five percent of hours



The price differences that were apparent from Figure 4 in the highest-priced hours persist even after adjusting for natural gas prices. Figure 9 shows the implied marginal heat rates for the top five percent of hours in 2008 through 2011 and highlights that although the number of hours with high (greater than 30 MMBtu per MWh) implied heat rates did increase in 2011, the larger effect was due to the heights at which scarcity prices were set.

To further illustrate these differences, the next figure shows the implied marginal heat rates on a monthly basis in each of the ERCOT zones in 2010 and 2011, with annual average heat rate data for 2008 through 2011. This figure is the fuel price-adjusted version of Figure 2 in the prior sub-section. Adjusting for gas price influence, Figure 10 shows that the annual, system-wide average implied heat rate increased significantly after remaining constant for the previous three years.

Figure 10: Monthly Average Implied Heat Rates

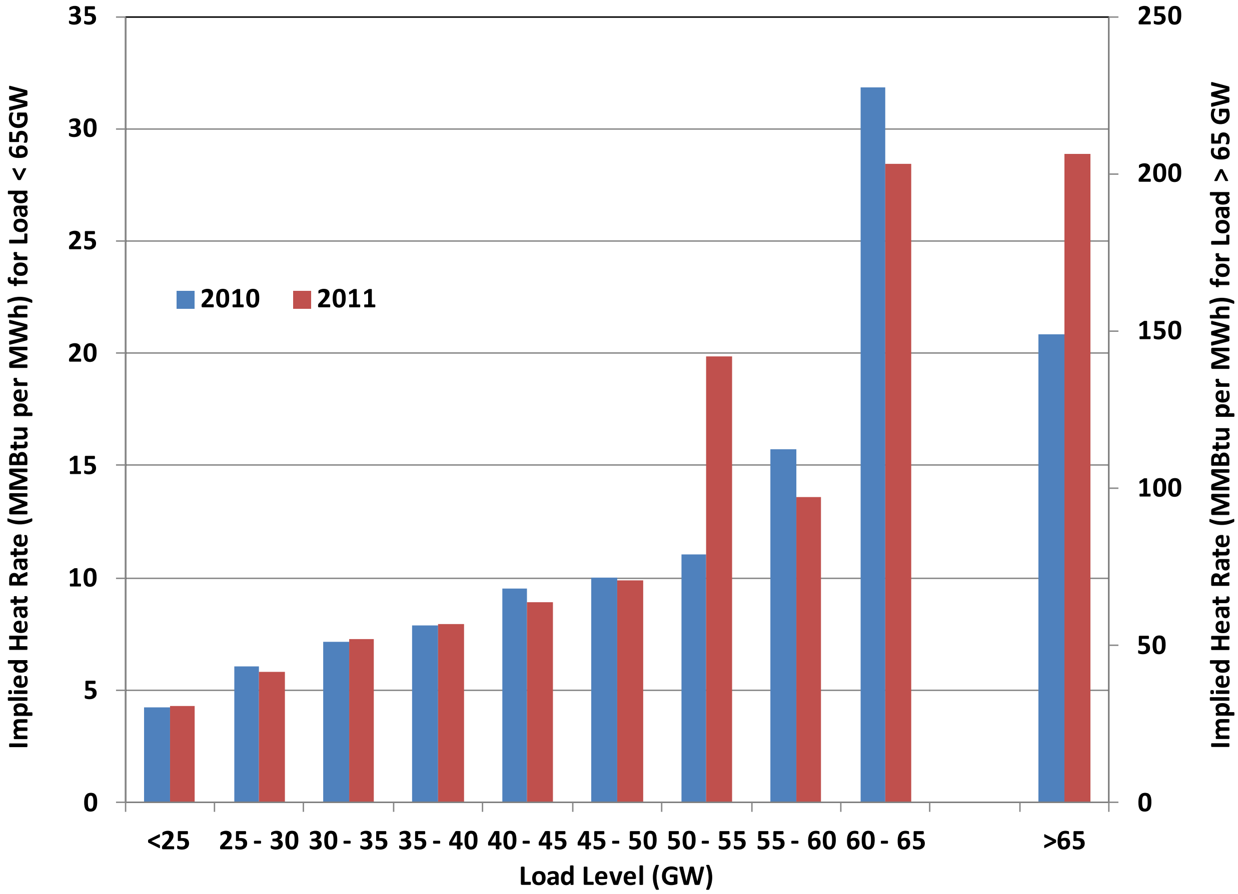


The monthly average implied heat rates in 2011 are generally consistent with 2010, with notable exceptions in February and August. Higher heat rates in February can be explained by the extended period when real-time prices were $3,000 per MWh due to extreme cold weather and the resulting unplanned outages of numerous generators. Extended hot, dry weather resulted in record system peak demands in August, and another extended period of energy prices reflecting scarcity conditions. The differences in the average annual implied heat rates observed at the zonal level can be attributed to the continued significant congestion related to wind generation exports from the West zone.

We conclude our examination of implied heat rates from the real-time energy market by evaluating them at various load levels. Figure 11 provides the average heat rate at various system load levels for 2011 and 2010.[4]

4 To appropriately compare twelve months of data under each market design, data labeled as 2010 in Figure 11 is from December 1, 2009 through November 30, 2010.

Figure 11: Heat Rate and Load Relationship

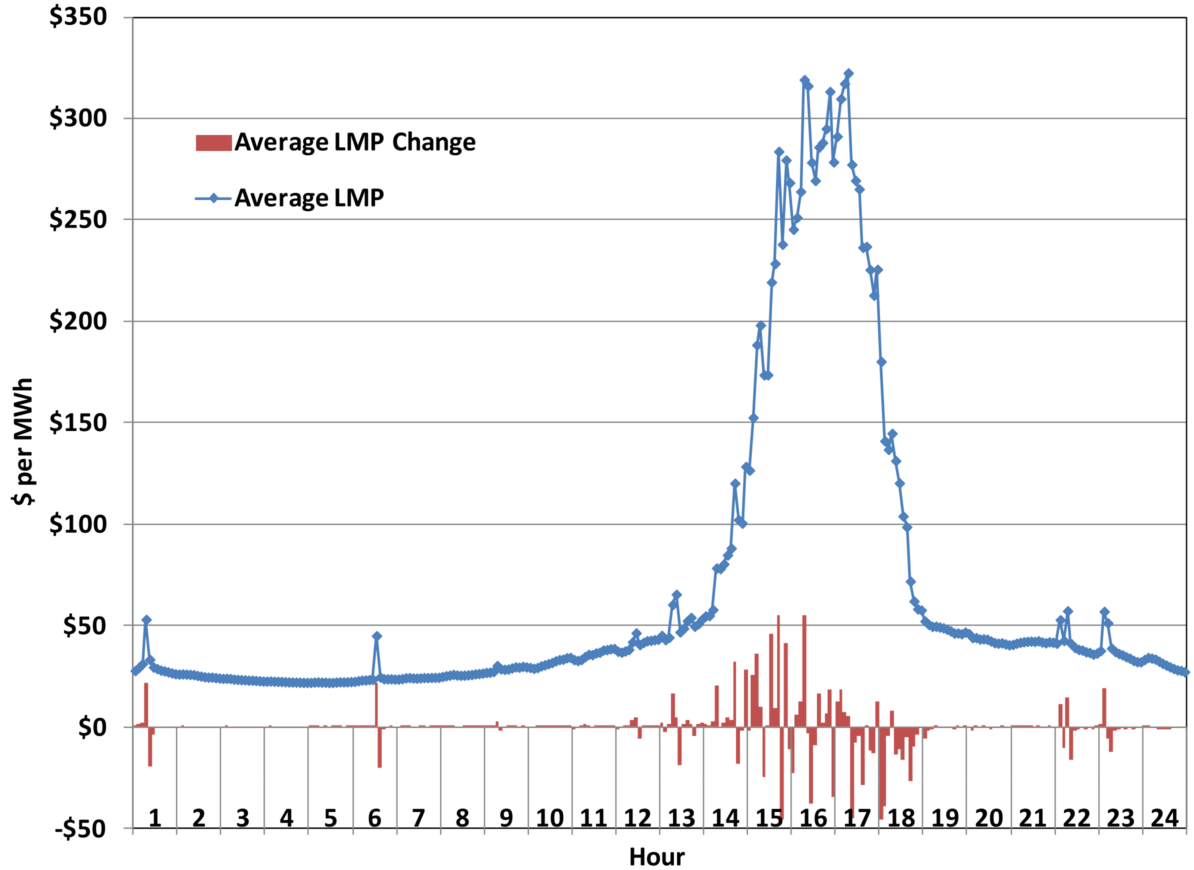


In a well performing market, a clear positive relationship between these two variables is expected since resources with higher marginal costs must be dispatched to serve higher loads. Although we do see a generally positive relationship, there is a noticeable disparity for loads between 50 and 55 GW. During the extreme cold weather event in early February, loads were at this level while prices reached $3,000 per MWh for a sustained period of time.
We also observe small reductions in heat rates for most load levels during 2011 compared to 2010, which we attribute to the enhanced efficiency of the nodal market.

C. Real-Time Price Volatility

Volatility in real-time wholesale electricity markets is expected because system load can change rapidly and the ability for supply to adjust can be restricted by physical limitations of the resources and the transmission network. Figure 12 presents a view of the price volatility experienced in ERCOT's real-time energy market during the summer months of May through August. Average five-minute LMPs are presented along with the magnitude of change in LMP for each five-minute interval. Outside of the hours from 15 to 18 (2:00 pm to 6:00 pm), short-term increases in average LMPs are typically caused by singular occurrences of high prices resulting from generator ramp rate limitations.

Figure 12: Real-Time Energy Price Volatility (May – August)



The average of the absolute value of changes in five-minute LMPs, expressed as a percentage of average LMP was approximately 6 percent for this period. To be able to compare with zonal market results, a similar percentage was calculated using 15 minute settlement point prices for the four geographic Load Zones.

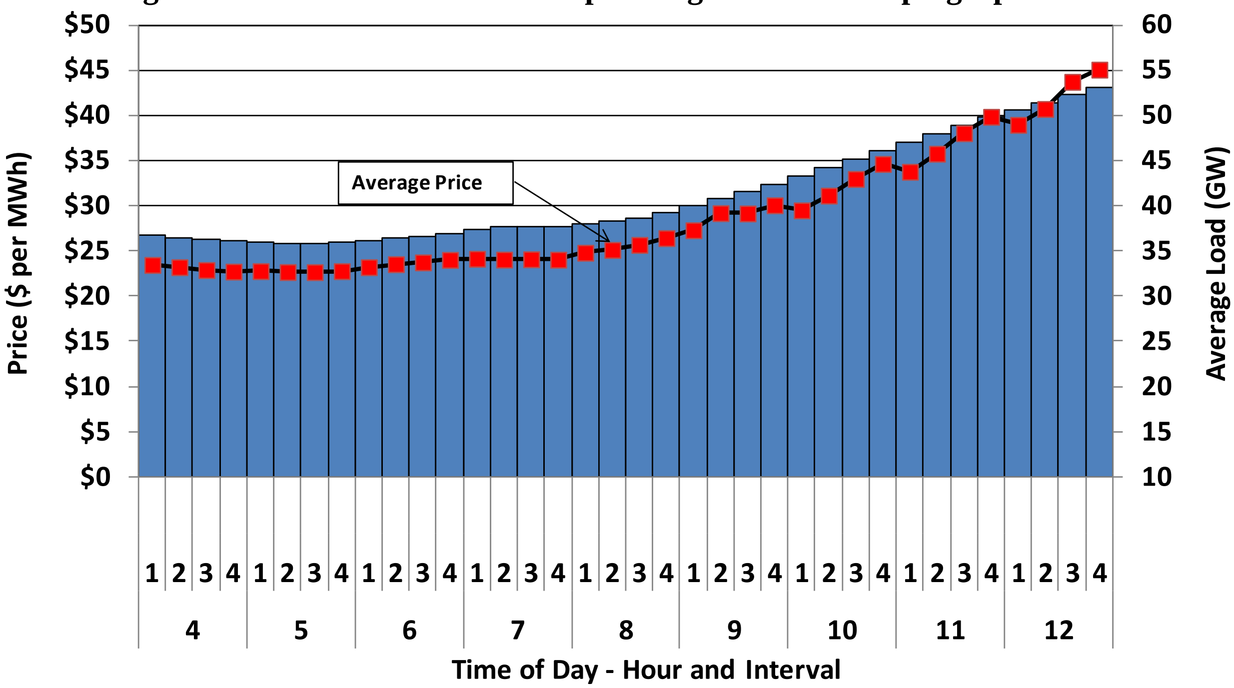
From the comparisons shown below in Table 1, implementation of the nodal market has resulted in less price volatility than experienced in the zonal market. Price volatility in the West zone has continued to be higher than in the other zones, which is expected given the very high penetration of variable output wind generation located in that area.

Table 1: Price Change as a Percent of Average Price

Load Zone	*2010 – Zonal*	*2011 – Nodal*
Houston	17.8%	14.0%
South	17.1	14.5
North	17.7	13.1
West	18.5	17.1

In well functioning markets we expect to observe a close correlation between price and load levels. This relationship was not observed under the zonal market design and was described repeatedly in prior annual reports[5].

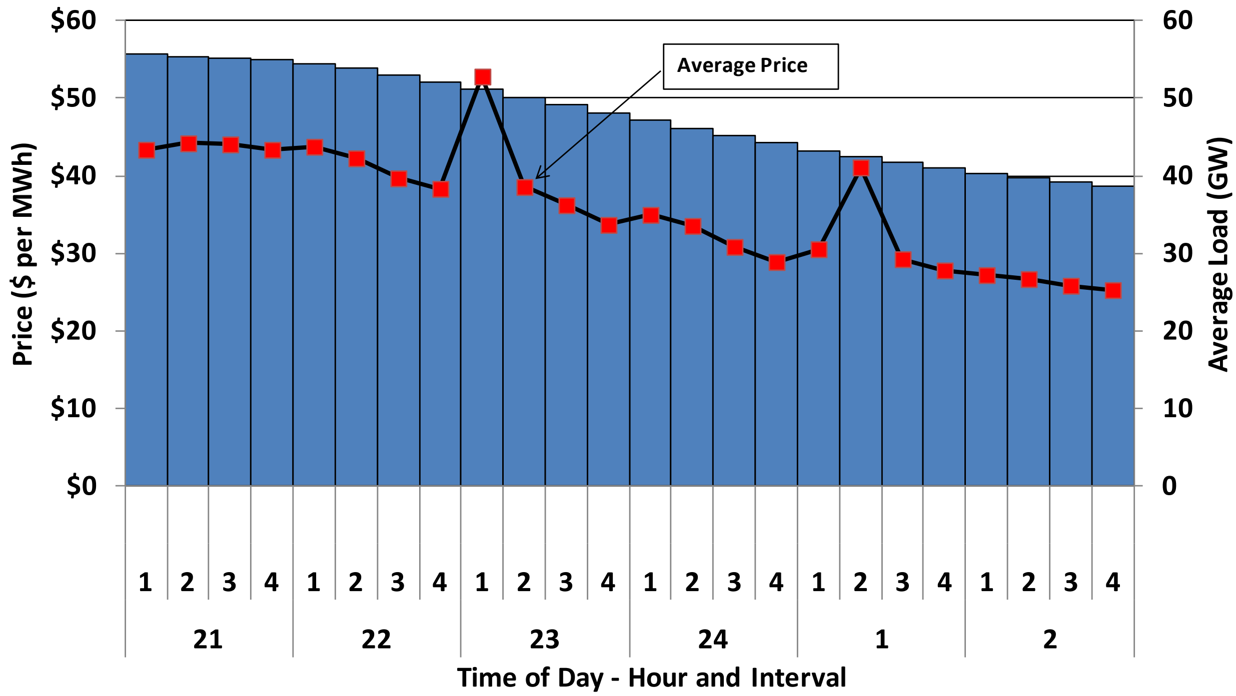
Figure 13: Price Load Relationship during Summer Ramping Up Hours



The relationship between average prices and average load levels during selected hours of the summer months are shown in Figure 13 and Figure 14. The periods shown in these two figures are times when there are typically large changes in load levels and associated generation ramping.

5 *See* 2009 ERCOT SOM Report at 21-28, 2008 ERCOT SOM Report at 21-28, and 2007 ERCOT SOM Report at 60-65.

Figure 14: Price Load Relationship during Summer Ramping Down Hours



The correlation between price and load is very high during the ramping up hours. This is the expected result when price formation is based on the cost of supply to meet the entire demand, rather than to meet a delta between total load and schedules, which was the case in the zonal market. The relationship between price and load during the ramping down hours exhibits discontinuities at 10:00 pm and just after midnight. These short term price increases are typically the result of prices rising in response to transitory generating unit ramp rate limitations in the aftermath of units turning off overnight. Even so, these price movements are much smaller, and less frequent than what was routinely observed in the zonal market.

D. February Cold Weather Event

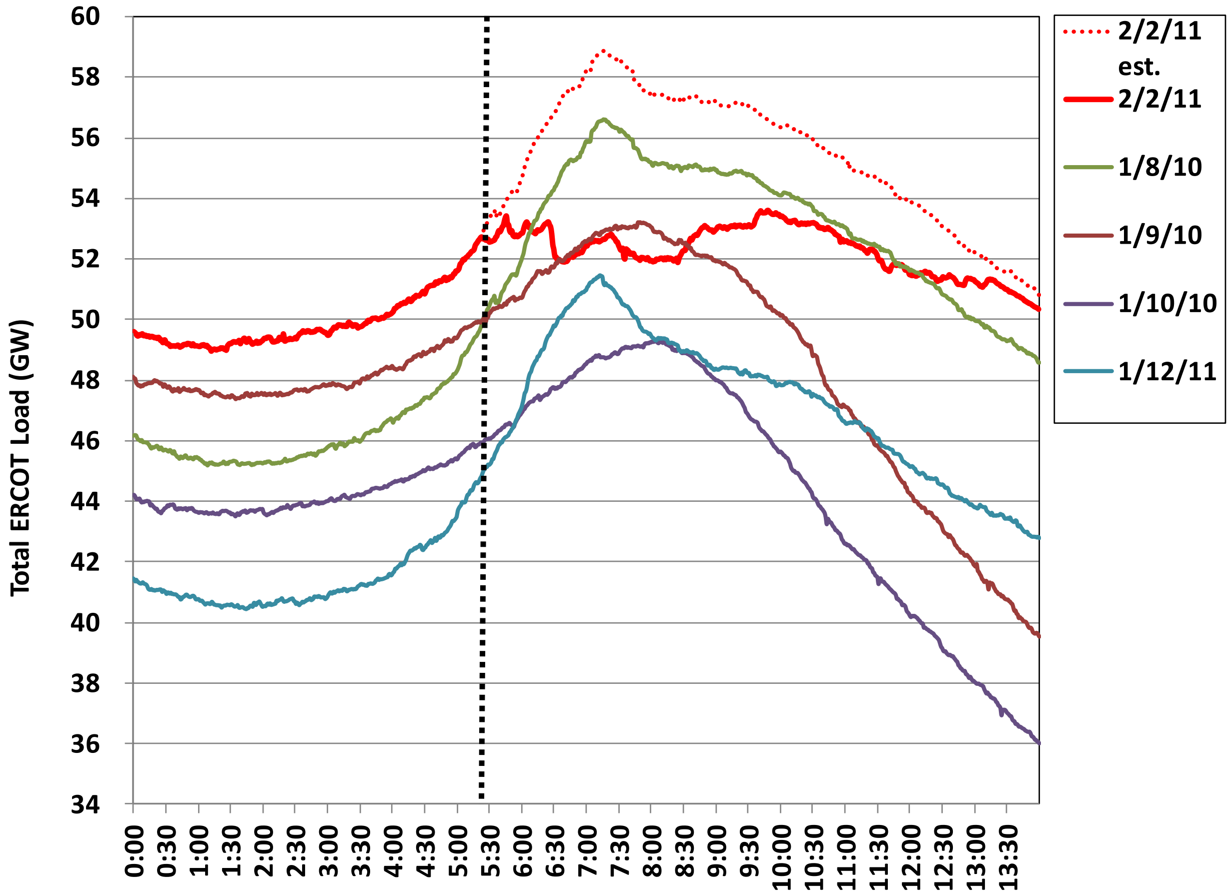
A significant operational challenge greeted the nascent nodal market in the early morning of February 2nd. The resulting market outcomes had a sizable effect on the overall annual results. This section more fully describes the specifics of that event.

In the early morning hours of February 2, 2011, the ERCOT region experienced extreme cold weather conditions, record electricity demand levels, and the loss of numerous electric generating facilities across the ERCOT region. These events combined to result in the deployment of load resources contracted to provide responsive reserve service and Emergency

Interruptible Load Service ("EILS") and culminated with 4,000 MW of firm load being shed for several hours.

Shown in Figure 15 are the five days through February 2, 2011 with the highest ERCOT electricity demand at the time just prior to the deployment of load resources. The demand for electricity in the early morning of February 2nd was 2,760 MW higher than on any other day in the history of the of the ERCOT region at this same time, and was experiencing a rapid rate of growth as is typical on such cold winter mornings. The demand curve for February 2, 2011 is noticeably distorted after 5:20 a.m. due to the various stages of load shedding that started at that time and remained in effect until just after 1:00 p.m., with the exception of approximately 470 MW of EILS deployments that remained in effect until approximately 10 a.m. on February 3rd.

Figure 15: Top 5 ERCOT Loads at 05:20 through Feb. 2, 2011

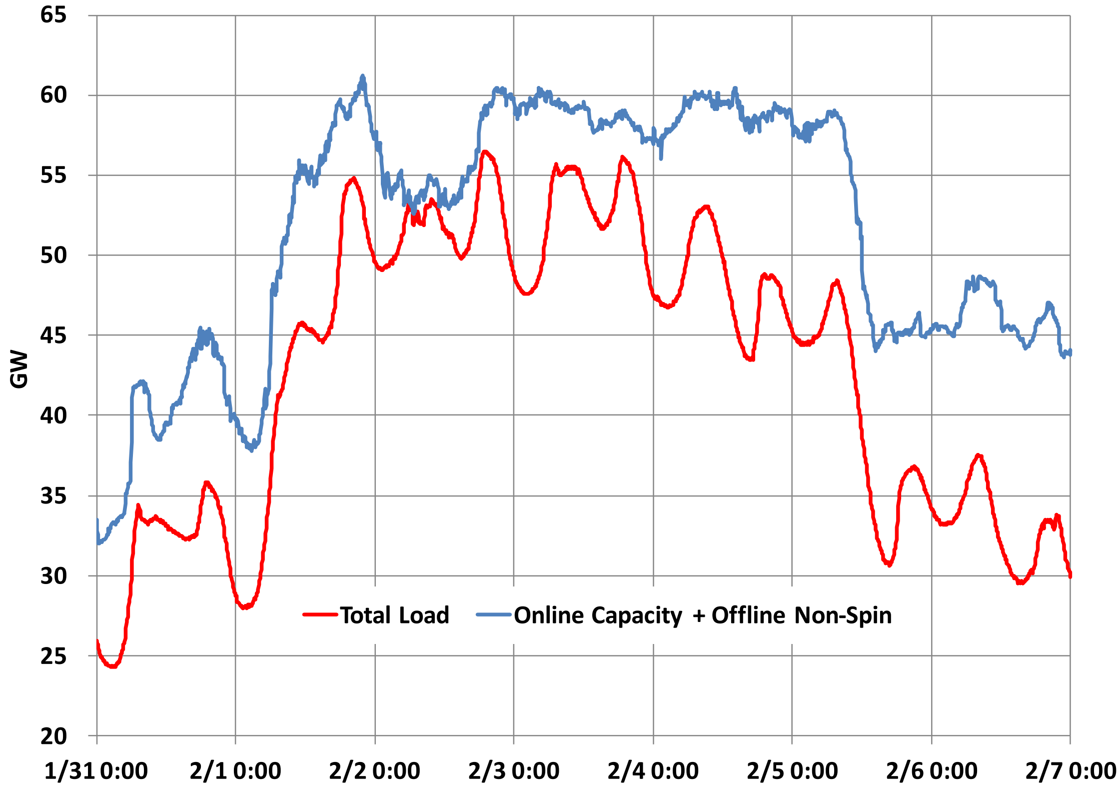


Also shown in Figure 15 is the estimated load that would have materialized on February 2nd absent any load curtailments, which indicates that, absent curtailments, the demand in the

ERCOT region would have approached 59,000 MW just after 7 a.m. This is almost 2,300 MW higher than the previous record instantaneous demand for electricity at this time of the day.

To provide additional perspective on the capacity limitations experienced on February 2nd, Figure 16 shows the available capacity (online capacity plus offline non-spinning reserves) and the ERCOT load for the seven days from January 31 through February 6, 2011.

Figure 16: Seven Day View of ERCOT Available Capacity and Load



The data in Figure 16 highlight the highly unusual and extremely narrow gap between available capacity and actual load that was experienced on the morning of February 2, 2011 relative to other days of similar and much lower load levels. These data also highlight the successful efforts to return substantial generating capacity to service prior to the record peak demand on the evening of February 2nd and to sustain the availability of that capacity for the high electricity demands experienced again on February 3rd.

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 that 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 operating reserves cannot be satisfied with the available resources, which will cause the system operator to take one or more of the following actions:

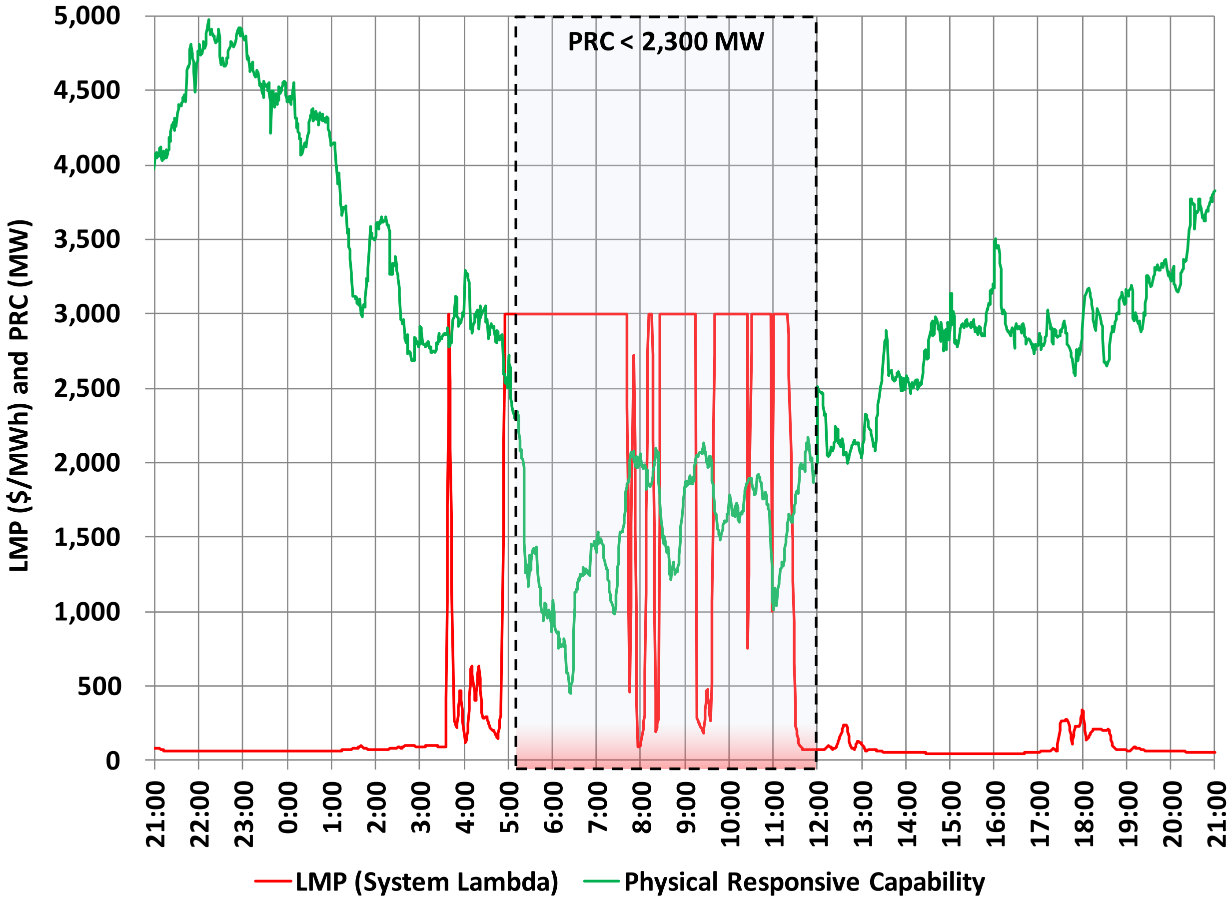
- Sacrifice a portion of the operating reserves by dispatching them for energy;
- Voluntarily curtail load through emergency demand response programs;
- Curtail exports or make emergency imports; or
- Involuntarily curtail load.

A market design that adheres to the pricing principles stated above will set prices that reflect each of these actions. When the market is in shortage, 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 reflected in energy prices to achieve efficient economic signals governing investment in generation, demand response and transmission.

During the morning of February 2, 2011, ERCOT operating reserve levels were reduced to perilously low levels for a sustained period of time. ERCOT's primary measure of overall operating reserves is Physical Responsive Reserve ("PRC"). ERCOT will remain in various levels of Energy Emergency Alert ("EEA") once PRC drops below 2,300 MW. Figure 17 shows the wholesale market prices and PRC from 21:00 on February 1 through 21:00 on February 2, 2011.

The data in Figure 17 show increased price volatility from 3:30 to 4:45 a.m. as system demand was increasing and generating units continued to be in various stages of tripping and starting. By 4:55 a.m., prices had reached a sustained level $3,000 per MWh, and PRC dropped below 2,300 MW by 5:10 a.m. PRC dropped to as low as 445 MW at 6:25 a.m., and remained consistently below the minimum 2,300 MW level until 12:00 p.m.

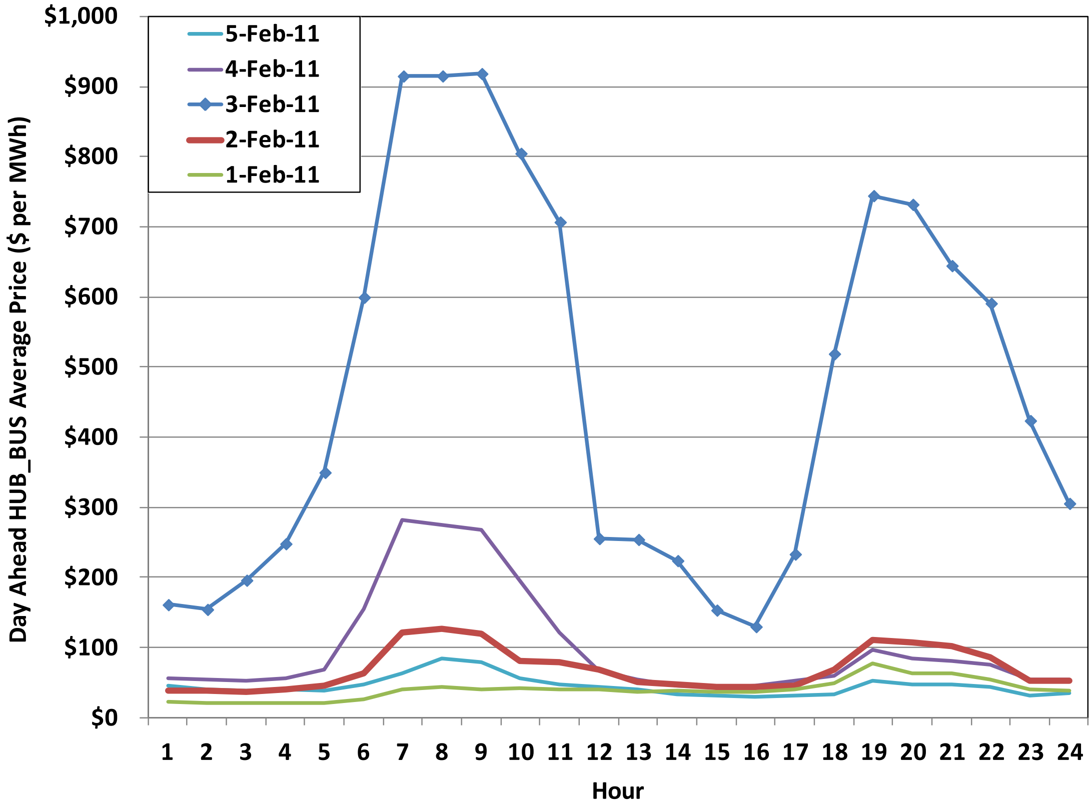
Figure 17: Prices and PRC (2/1/11 21:00 – 2/2/11 21:00)



These wholesale market pricing outcomes were consistent with the ERCOT energy-only market design. The wholesale market prices began communicating the degradation in system reliability as early as 3:30 a.m. By 4:55 a.m. – 15 minutes prior to the reduction of PRC below the minimum acceptable level of 2,300 MW and 50 minutes prior to the first stage of firm load shedding – prices were consistently communicating the rapidly deteriorating system reliability conditions. Finally, as load levels naturally reduced and reserve levels were restored, prices dropped back to levels typical of non-shortage conditions.

The secondary effect of the conditions during the morning of February 2, 2011 was the effect on the day-ahead market for February 3, 2011. Figure 18 shows the hourly average day-ahead market energy prices for February 1st through the 5th. Notable in Figure 18 is that, while somewhat higher than a typical day, the day-ahead prices for February 2nd are significantly lower than the real-time prices shown in Figure 17 for the same day. Figure 18 also shows that the day-ahead prices for February 3rd were substantially higher than on February 2nd and, in fact, represent the highest day-ahead market prices experienced since the implementation of the nodal market.

Figure 18: Average Hourly Day-Ahead Prices for Feb. 1-5, 2011



To better understand these day-ahead pricing outcomes for February 3rd requires a review of the day-ahead market function and timing. The ERCOT day-ahead market is not a mandatory market; rather, it is a voluntary market that consists of willing sellers that will be cleared for offers to sell energy at their offer price or higher and willing buyers that will be cleared for bids to buy at their bid price or lower. The day-ahead market is not limited to physical generation as sellers or physical load serving entities as buyers. In other words, any market participant – whether it has a physical position in the market or not – can participate in the day-ahead market and take a financial position against the real-time market. Because of the voluntary, financial nature of the day-ahead market, its outcomes are strongly driven by expectations of the real-time market performance for the following day.

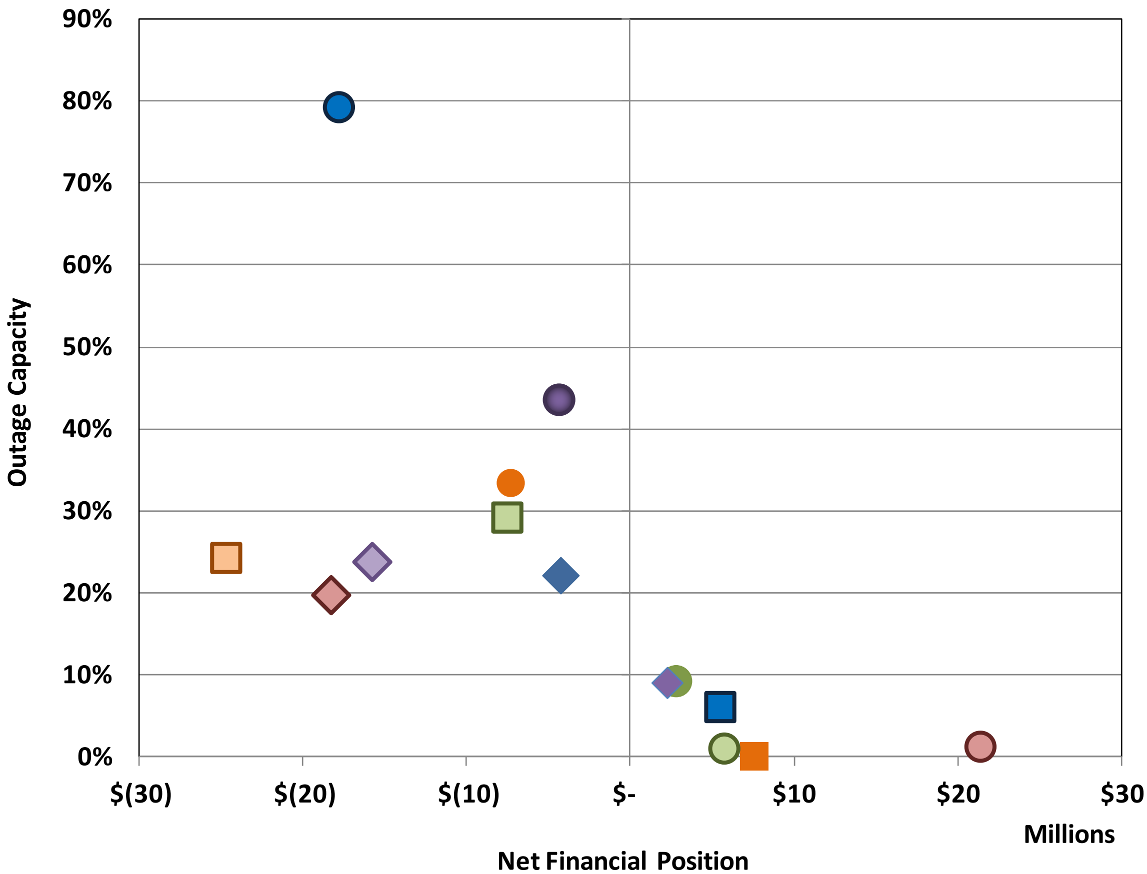
On this point, an understanding of the timing of the day-ahead market execution is critical. The day-ahead market opens for bid/offer submission at 6:00 a.m. on the day prior to the operating day, and the submission window closes at 10:00 a.m. Thus, for the February 3rd day-ahead market, the submission window opened at 6:00 a.m. and closed at 10:00 a.m. on February 2nd. Thus, at the time that bids/offers were submitted for the February 3rd day-ahead market, ERCOT was in the middle of the EEA level 3 events on February 2nd. Considerable uncertainty regarding generating unit availability and system conditions for February 3rd existed at that time, while the forecast called for continued arctic conditions across the state and record electricity demand was again forecast for the ERCOT region.

On a typical day, the day-ahead market results for February 3rd would give rise to market performance concerns, just as the real-time results on February 2nd would also raise concerns on a typical day. However, the real-time system conditions on February 2nd were far from typical, with the market outcomes reflecting the underlying system reliability conditions, consistent with the energy-only market design. Likewise, the day-ahead market outcomes for February 3rd were driven by the highly atypical uncertainties and risks facing both the supply and demand sides that existed at the time the day-ahead market submissions occurred, and the results are not unexpected given those considerations. Notably, while the day-ahead prices for February 3rd averaged $465.64 per MWh, day-ahead prices for February 4th and 5th averaged $99.56 and $44.68 per MWh, respectively, as the weather moderated resulting in decreased electricity demands and generation resources previously experiencing outages were returned to service. Although near-record electricity demand levels were again experienced on February 3rd, a substantial number of generating units that were forced out of service on February 2nd were able to return by the morning of February 3rd. Real-time prices on February 3rd averaged approximately $112 per MWh, which is higher than a typical day but much lower than the day-ahead prices for that day. Overall, we find that the real-time and day-ahead wholesale markets for February 2nd and 3rd operated efficiently given the system conditions and the outcomes are consistent with the ERCOT energy-only wholesale market design.

Although a wide range of actions were undertaken by generation resource owners in preparation for the extreme weather conditions, it is clear from the unprecedented loss of generation capacity on the morning of February 2nd that many of these preparatory efforts were unsuccessful. This experience will serve to produce lessons learned and specific areas for improvement in the areas of generation resource weatherization and coordinated extreme weather planning. Overall, although the scope and magnitude of the generating unit outages on February 2nd was absolutely unprecedented, we do not find any evidence that indicates that any of the outages were the result of physical withholding.

Another measure to provide additional insight related to this finding is the relative profitability of market participants during these events and how it correlates with unit outages. Although an assessment of profitability in isolation is insufficient to draw conclusions related to market manipulation or market power, increased profitability is the primary motive associated with resource withholding strategies. Hence, a negative correlation between resource outages and profitability would provide increased confidence in the finding that the outages were not the result of market manipulation strategies or market power abuses.

Figure 19: Generation Availability and Net Financial Position on Feb. 2, 2011



Real-time market prices on the morning of February 2nd were at or near the system-wide cap of $3,000 per MWh due to the short-supply conditions existing during the EEA event. Figure 19 shows the relationship between wholesale market profitability on February 2nd and availability of generation during the morning of February 2nd for market participants representing the largest fleets of generating resources. The data in Figure 19 show that those market participants who were able to operate their generation fleet at greater than 90% availability during the morning of February 2nd were financially successful that day. In contrast, market participants affected by significant generation outages found themselves unprofitable that day.[6]

Day-ahead market prices for February 3rd were also affected by the conditions on February 2nd and were substantially higher than normal levels. Although some market participants that lost money on February 2nd were able to recover much of their lost generating capacity and financial losses on February 3rd, none of the market participants that lost significant generating capacity and were unprofitable on February 2nd had financial gains on Feb. 3rd that significantly exceeded their losses on February 2nd.

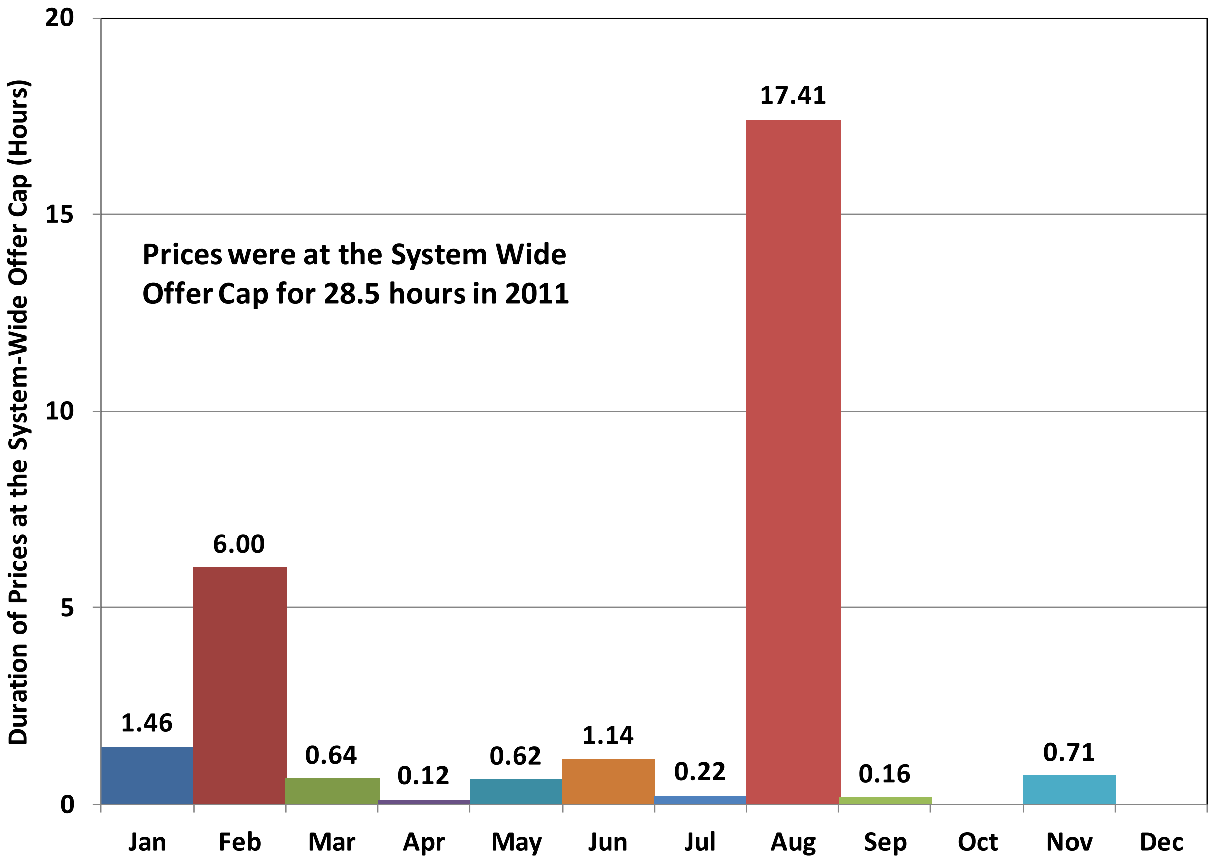
E. August Weather Conditions and Shortages

The summer of 2011 will be remembered as the hottest and driest on record in ERCOT. These extreme weather conditions led to record high demand for electricity during August. There were 50 hours in 2011 with electricity demands that exceeded the highest hourly demand that occurred in 2010. More details of the demand for electricity in ERCOT are provided in Section IV.A, ERCOT Loads in 2011.

During these high demand conditions there is an increased likelihood that the available generation capacity is not sufficient to meet customer demands for electricity and maintain the required reliability reserves. As more fully described later in Section V.B, Effectiveness of the Scarcity Pricing Mechanism, the nodal market causes energy prices to rise toward the system-wide offer cap as available operating reserves approach minimum required levels to reflect the degradation in system reliability. Figure 20 shows the aggregated amount of time represented by all dispatch intervals where the real-time energy price was at the system-wide offer cap, displayed by month. Of the 28.5 hours of the annual total time at the system-wide offer cap, more than 17 hours (60 percent) occurred during August.

6 The data in Figure 19 do not include market participants without physical generation resources or market participants operating only wind generation resources or relatively small fleets of non-wind generation resources. Outage capacity does not include planned outages.

Figure 20: Duration of Prices at the System Wide Offer Cap

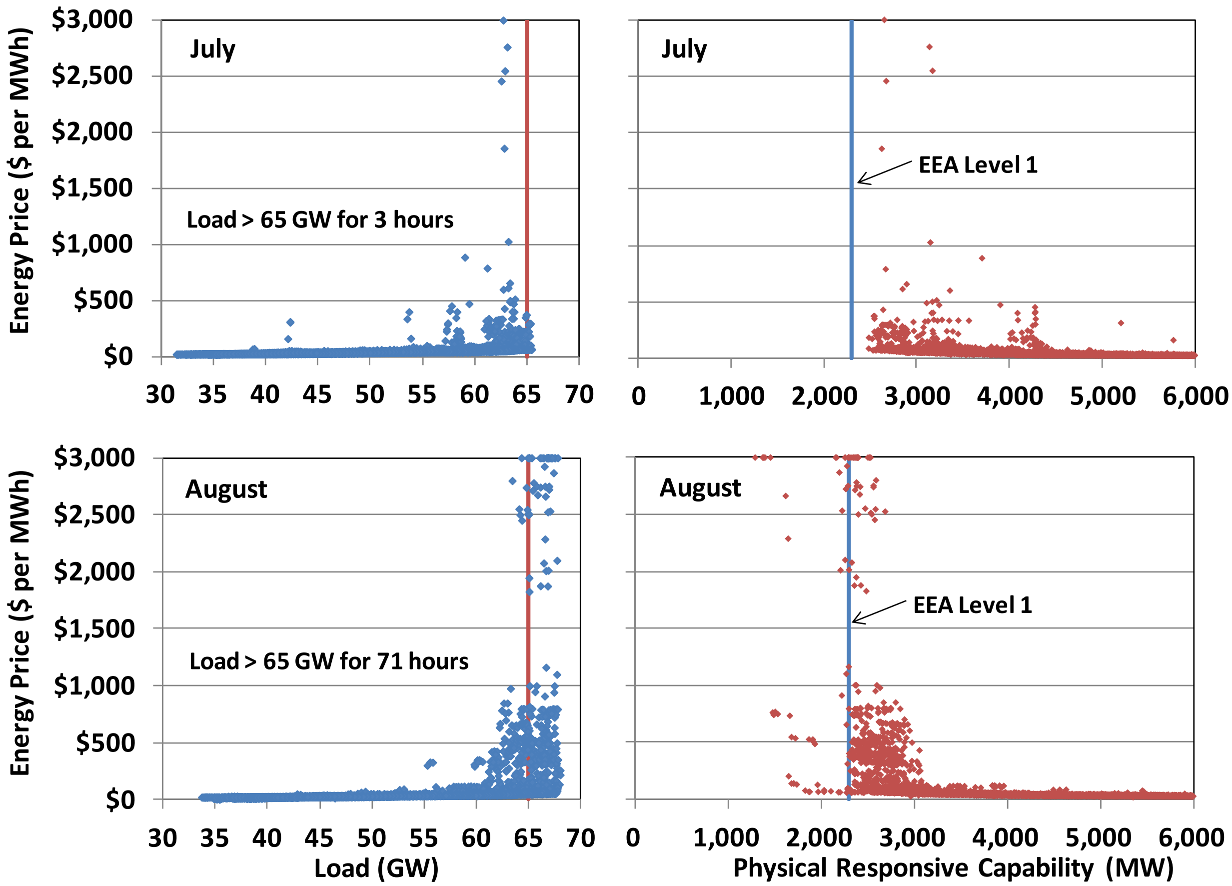


The next figure provides a more detailed comparison of load, required reserve levels, and prices in July and August. The weather in ERCOT was extremely hot and dry during both months, but there were very few dispatch intervals when real-time energy prices reached the system-wide offer cap in July compared to the relatively high frequency it occurred in August. Although the weather may have been similar, there were significant differences in load and available operating reserve levels, resulting in much higher prices in August.

Shown on the left side of Figure 21 is the relationship between real-time energy price and load level for each dispatch interval during the months of July and August. ERCOT loads were greater than 65 GW for three hours in July, whereas load levels exceeded 65 GW for 71 hours in August. As previously discussed, a strong positive correlation between higher load and higher prices is expected in a well functioning energy market. We observe such a relationship between higher prices and higher loads in both months. With overall higher loads and more frequent occurrences of very low operating reserves in August, higher energy prices are expected.

Although load levels are strong predictors of energy prices, an even more important predictor is the level of operating reserves. Simply put, operating reserves are the difference between the total capacity of operating resources and the current load level. As load level increases against a fixed quantity of operating capacity, the amount of operating reserves diminishes. The minimum required operating reserves prior to the declaration of Energy Emergency Alert Level 1 by ERCOT is 2,300 MW. As the available operating reserves approach the minimum required amount, energy prices should rise toward the system-wide offer cap to reflect the degradation in system reliability.

Figure 21: Load, Reserves and Prices: July and August



On the right side of Figure 21 are data showing the relationship between real-time energy prices and the quantity of available operating reserves for each dispatch interval during the months of July and August. This figure shows a strong correlation between diminishing operating reserves and rising prices. In July available operating reserves were generally maintained well above minimum levels, and there were only 0.22 hours where the energy price reached $3,000 per MWh. In contrast, there were numerous dispatch intervals in August when the minimum operating reserve level was approached or breached, with 17.4 hours where prices reached $3,000 per MWh. However, there are also a substantial number of dispatch intervals where operating reserves are below minimum requirements with prices well below the level that would be reflective of the reduced state of reliability. In Section V.C, Demand Response Capability we provide an example explaining why this can occur and offer a recommendation for improvement.

II. Review of Day-Ahead Market Outcomes

One of the fundamental improvements brought about by the implementation of ERCOT's nodal market design is the establishment of a centralized day-ahead market, which allows participants to make financially binding forward purchases and sales of power for delivery in real-time. Offers to sell can take the form of either a three-part supply offer, which allow sellers to reflect the unique financial and operational characteristics of a specific generation resource, or an energy only offer, which is location specific but is not associated with a generation resource. Bids to buy are also location specific. Ancillary services are also procured as part of the day-ahead market clearing. The third type of transaction included in the day-ahead market is bids to buy point to point ("PTP") Obligations, which allow parties to hedge the incremental cost of congestion between day-ahead and real-time operations.

With the exception of the acquisition of ancillary service capacity, the day-ahead market is a financial market. Although all bids and offers are evaluated in the context of their ability to reliably flow on the transmission network, there are no operational obligations resulting from the day-ahead market clearing. These transactions are made for a variety of reasons, including satisfying the participant's own supply, managing risk by hedging the participant's exposure to the real-time market, or arbitraging with the real-time markets. For example, load serving entities can insure against volatility in the real-time market by purchasing in the day-ahead market. Finally, the day-ahead market plays a critical role in coordinating generator commitments. For all of these reasons, the performance of the day-ahead market is essential.

In this section we review energy pricing outcomes from the day-ahead market and compare their convergence with real-time energy prices. We will also review the volume of activity in the day-ahead market, including a discussion of Point to Point Obligations. We conclude this section with a review of the ancillary service markets.

A. Day-Ahead Market Prices

One indicator of market performance is the extent to which forward and real-time spot prices converge over time. Forward prices will converge with real-time prices when two main conditions are in place: a) there are low barriers to shifting purchases and sales between the forward and real-time markets; and b) sufficient information is available to market participants to allow them to develop accurate expectations of future real-time prices. When these conditions are met, market participants can be expected to arbitrage predictable differences between forward prices and real-time spot prices by increasing net purchases in the lower-priced market and increasing net sales in the higher-priced market. These actions will tend to improve the convergence of forward and real-time prices.

In this section, we evaluate the price convergence between the day-ahead and real-time markets. This analysis reveals whether persistent and predictable differences exist between day-ahead and real-time prices, which participants should arbitrage over the long-term. To measure the short-term deviations between real-time and day-ahead prices, we also calculate the average of the absolute value of the difference between the day-ahead and real-time price on a daily basis.

This measure captures the volatility of the daily price differences, which may be large even if the day-ahead and real-time energy prices are the same on average. For instance, if day-ahead prices are $70 per MWh on two consecutive days while real-time prices are $40 per MWh and $100 per MWh on the two days, the absolute price difference between the day-ahead market and the real-time market would be $30 per MWh on both days, while the difference in average prices would be $0 per MWh.

Day-ahead prices averaged $46 per MWh in 2011 compared to an average of $43 per MWh for real-time prices.[7] This slight premium is consistent with expectations due to the much higher volatility of real-time prices. Risk is lower for loads purchasing in the day-ahead and higher for generators selling day-ahead. The higher risk for generators is associated with the potential of having a forced outage and buying back energy at real-time prices. This may explain why the highest premiums occurred during the highest priced months. Overall, the day-ahead premiums were very similar to the differences observed in 2009 and 2010.[8] The average absolute difference between day-ahead and real-time prices was $24.50 per MWh in 2011; much higher than in the previous two years where the average absolute difference was $12.25 and $12.37 in 2010 and 2009, respectively. This large increase was the result of the significant periods of very

---

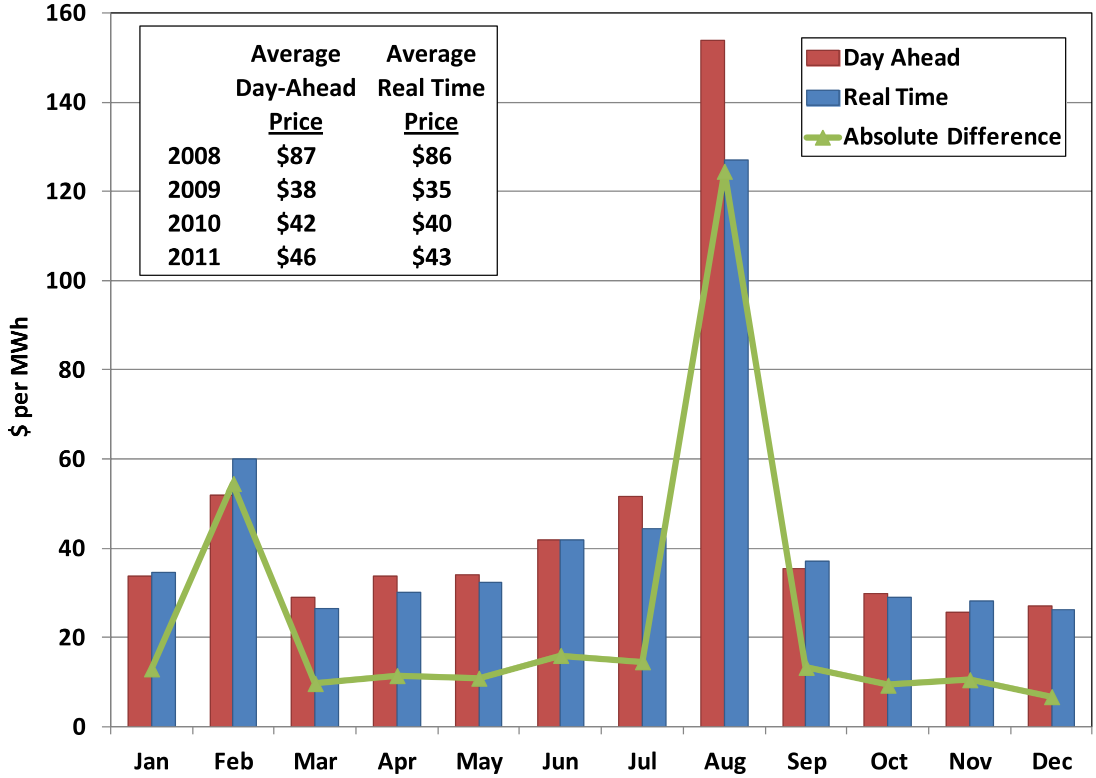
7 These values are simple averages, rather than load-weighted averages presented in Figure 1and Figure 2.

8 In 2009 and 2010 under the zonal market the comparison was made between on-peak forward prices and prices for the same on-peak period in the balancing energy market.

high real-time prices during February and August. Removing the contribution from these two months reduces the average absolute difference to $11.49 per MWh in 2011.

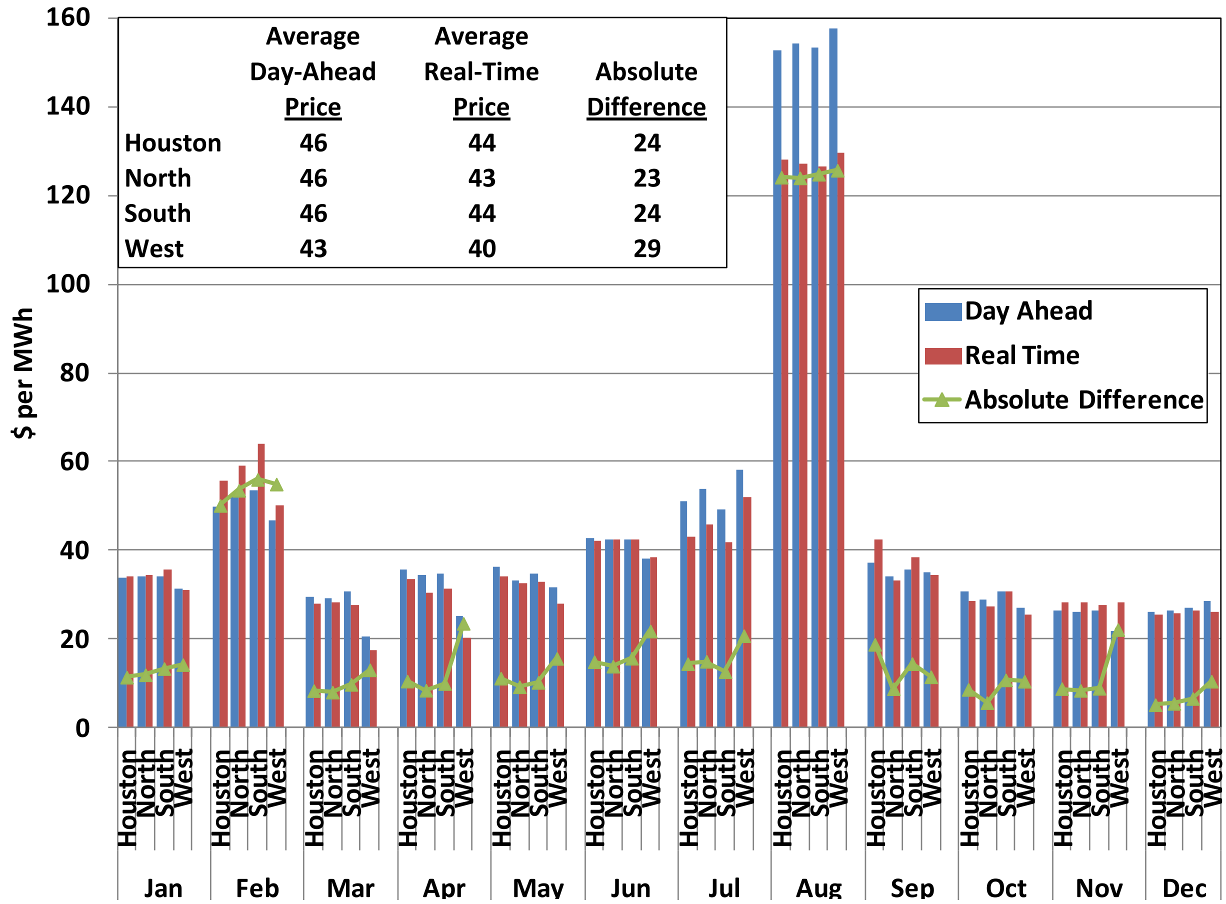
Figure 22 shows the price convergence between the day-ahead and real-time market, summarized by month.

Figure 22: Convergence between Forward and Real-Time Energy Prices



Below, in Figure 23 monthly day-ahead and real-time prices are shown for each of the geographic load zones. Of note is the difference in the west zone data compared to the other regions. The higher volatility in west zone pricing is likely associated with the uncertainty of forecasting wind generation output and the resulting price levels between day-ahead and real-time.

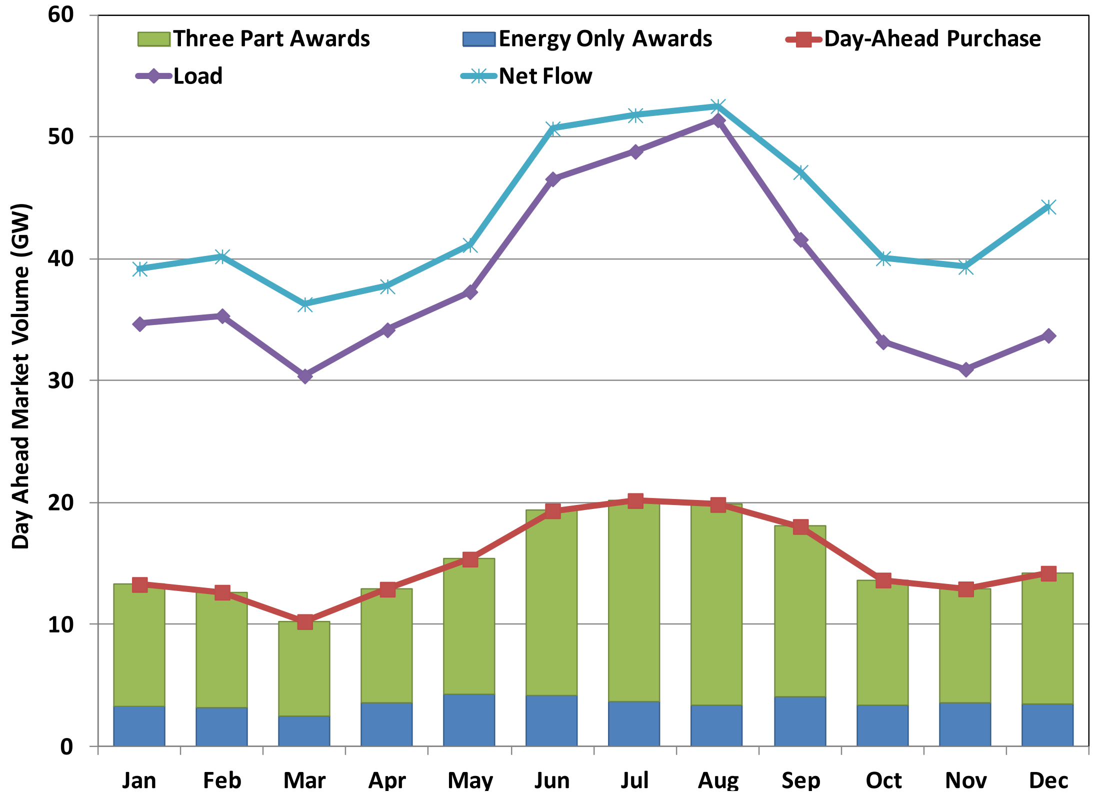
Figure 23: Day-Ahead and Real-Time Prices by Zone



B. Day-Ahead Market Volumes

Our next analysis summarizes the volume of day-ahead market activity by month. In Figure 24, we find that day-ahead purchases are approximately 40 percent of real-time load. These energy purchases are met through a combination of generator specific and virtual offers.

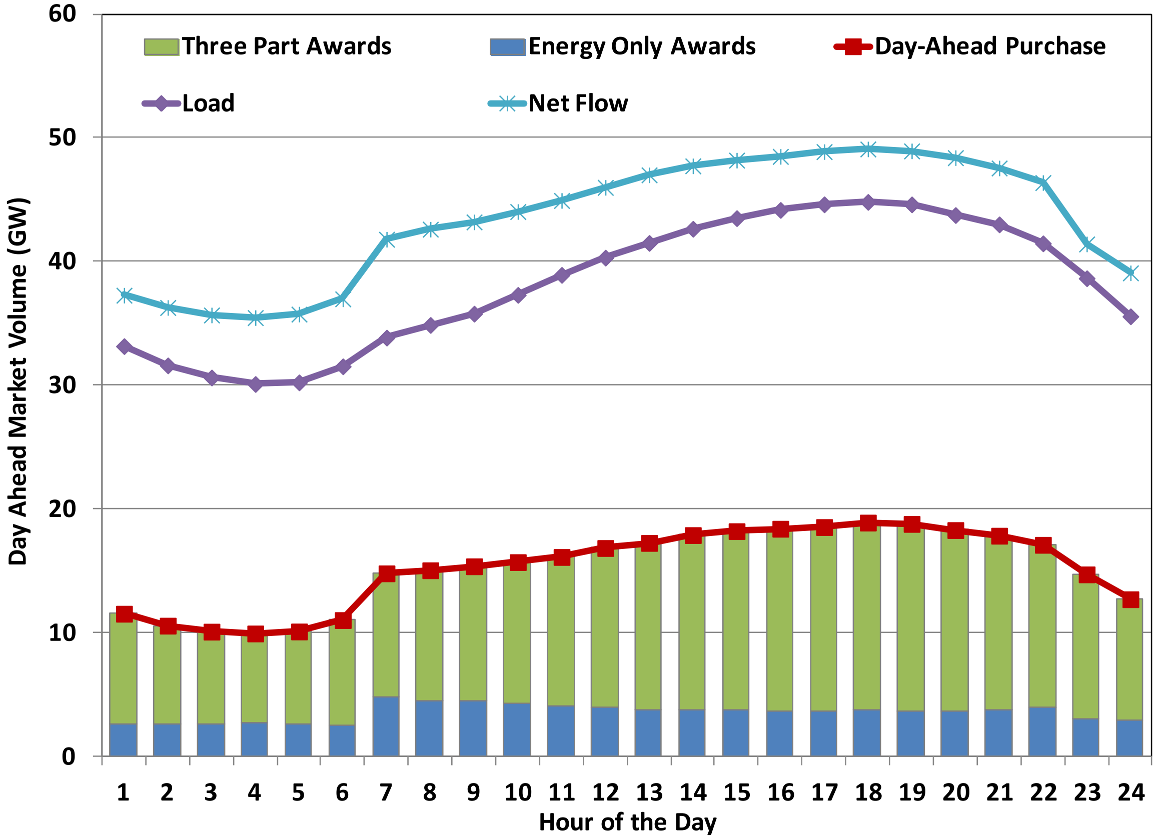
Figure 24: Volume of Day-Ahead Market Activity by Month



By adding in the effects of net energy flows associated with purchases of PTP Obligations, we find that on average total volumes transacted in the day-ahead market are greater than real-time load.

Figure 25 presents the same day-ahead market activity data summarized by hour of the day. In this figure the volume of day-ahead market transactions is disproportionate with load levels between the hours of 6 and 22. Since these times align with common bilateral transaction terms, it appears that market participants are using the day-ahead market to trade around those positions.

Figure 25: Volume of Day-Ahead Market Activity by Hour



C. Point to Point Obligations

Purchases of Point to Point ("PTP) Obligations comprise a significant portion of day-ahead market activity. These instruments are similar to, and can be used to complement Congestion Revenue Rights (CRRs). CRRs, as more fully described in Section III.C, are acquired via monthly and annual auctions and allocations. CRRs accrue value to their owner based on locational price differences as determined by the day-ahead market. PTP Obligations are instruments that are purchased as part of the day-ahead market and accrue value to their owner based on real-time locational price differences. By acquiring a PTP Obligation using the proceeds due to them from their CRR holding, a market participant may be described as rolling their hedge to real-time.

In this subsection we provide additional details about the volume and profitability of these PTP Obligations.

Figure 26: PTP Obligation Volume

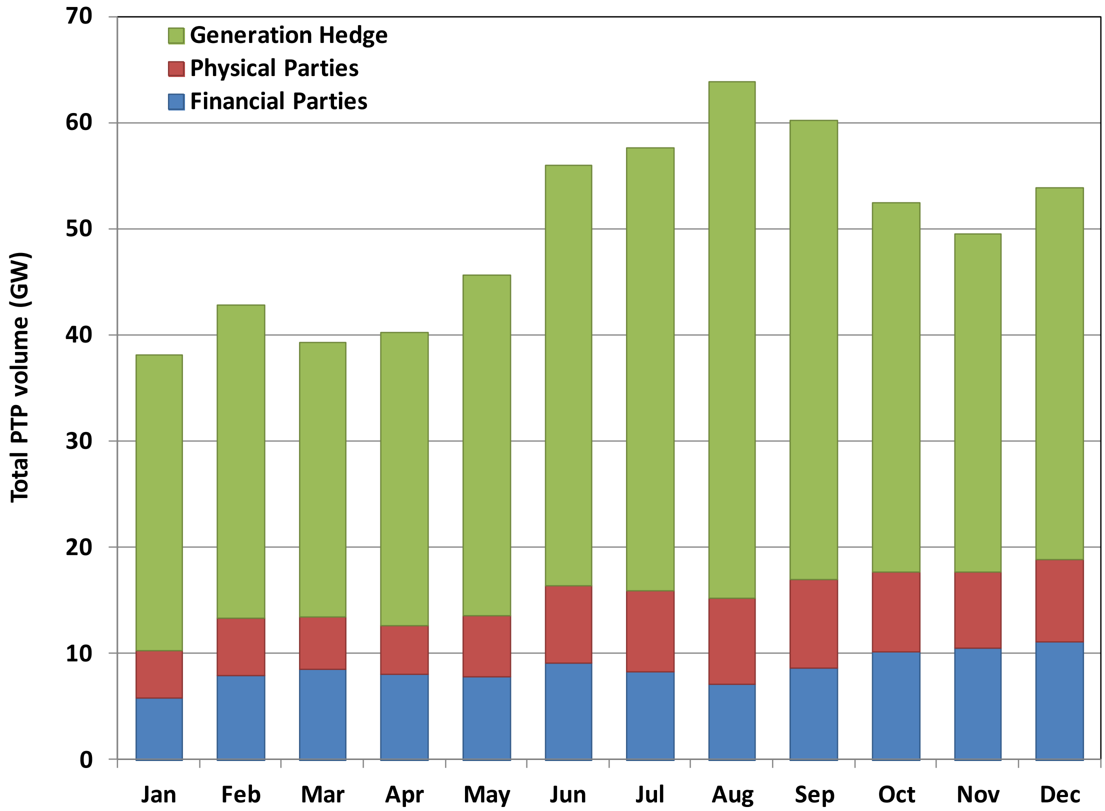
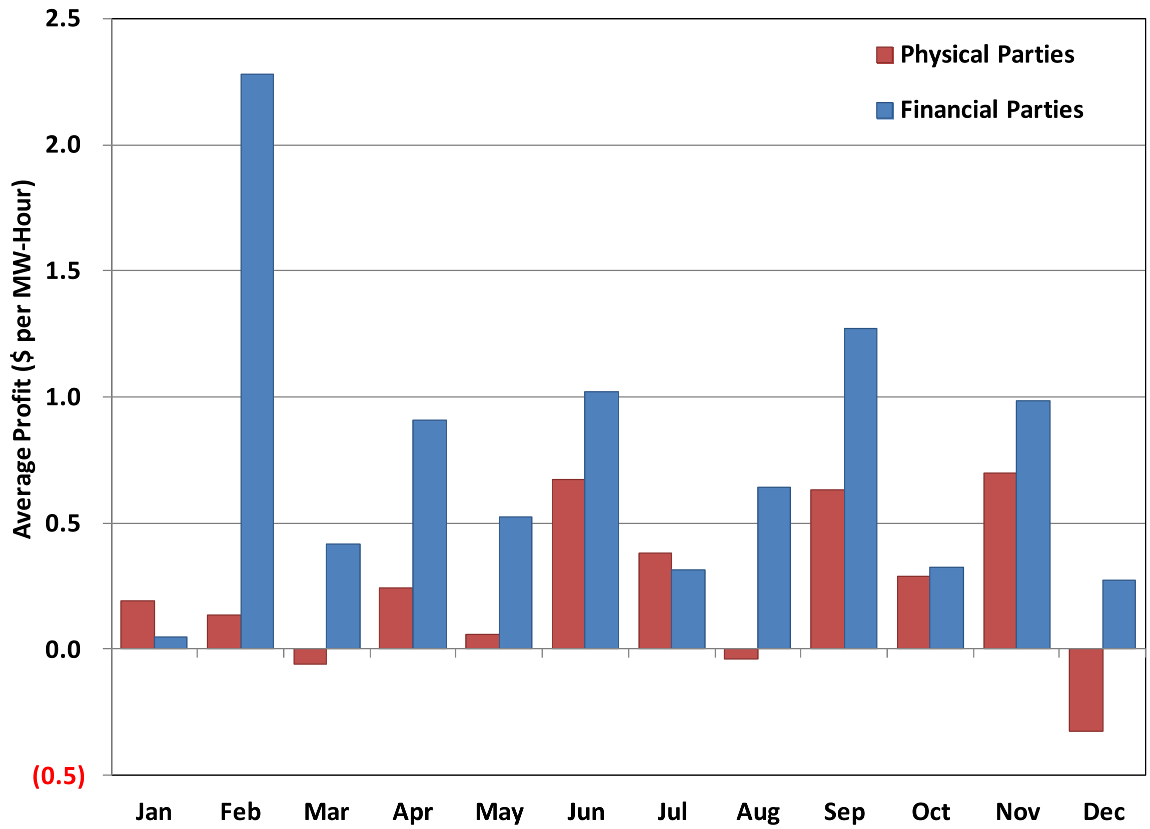


Figure 26 presents the total volume of PTP Obligation purchases divided into three categories. Compared to the previous two figures which showed net flows associated with PTP Obligations, in this figure we examine the total volume. For all PTP Obligations that source at a generator location, we attribute capacity up to the actual generator output as a generator hedge. From the figure above we see that this comprised most of the volume of PTP Obligations purchased. The remaining volumes of PTP Obligations are not directly linked to a physical position and are assumed to be purchased primarily to profit from anticipated price differences between two locations. We further separate this arbitrage activity by type of market participant. Physical parties are those that have actual real-time load or generation, whereas financial parties have neither. We find that the arbitrage activity is fairly evenly split between physical and financial parties, and further, the volume of arbitrage activity steadily increased throughout the year.

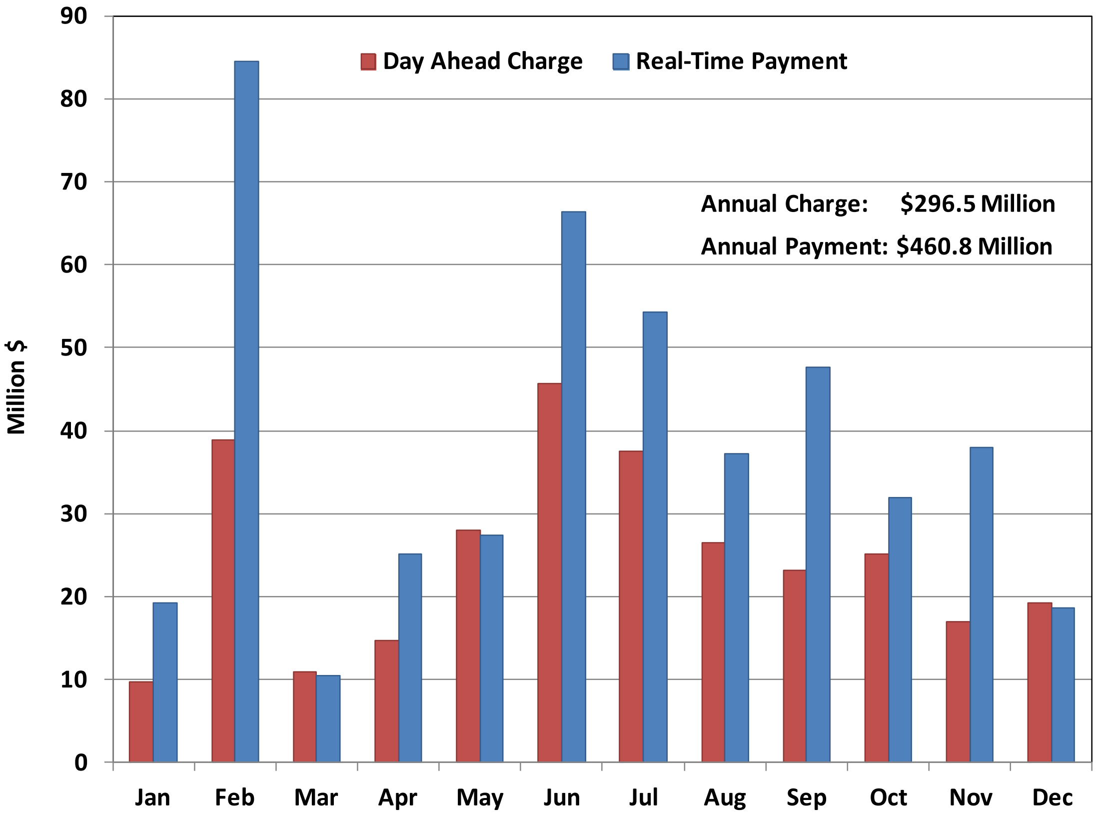
To the extent the price difference between the source and sink of a PTP Obligation is greater in real-time than it was day-ahead price, the owner will have made a profit. Conversely, if the price difference does not materialize in real-time, the PTP Obligation may be unprofitable. We compare the profitability of PTP Obligation holdings by the two types of participants in Figure 27.

Figure 27: Average Profitability of PTP Obligation



From the figure above we can infer different motivations between the two types of participants. Because financial participants have no real-time load or generation they have no other exposure to real-time prices. If a financial participant is not making a profit on their PTP Obligations there is no reason for them to buy any. In fact, their profit seeking action of buying PTP Obligations between points where congestion is expected helps make the day-ahead market converge with real-time market outcomes. On the other hand, physical participants do have exposure to real-time prices. It is reasonable to expect that this type of participant is most interested in limiting that exposure by using PTP Obligations as a hedge.

Figure 28: Point to Point Obligation Charges and Payments



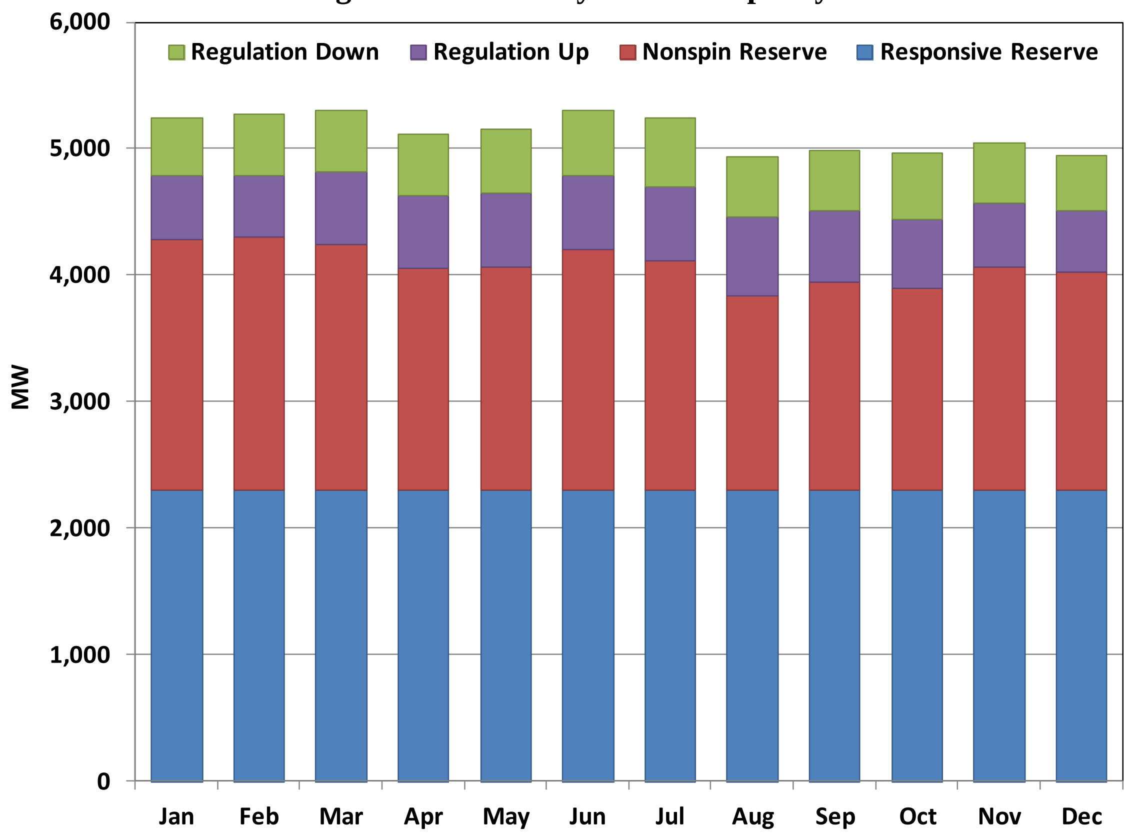
To conclude our analysis of PTP Obligations, in Figure 28 we compare the total amount paid for these instruments day-ahead, with the total amount received by their holders in real-time. In most months owners received, in aggregate, more in payments for their PTP Obligations than they paid to acquire them. This occurs when real-time congestion is greater than what occurred in the day-ahead. The payments made to PTP Obligation owners come from real-time congestion rent. We assess the sufficiency of real-time congestion rent to cover both PTP Obligations and payments to owners of CRRs who elect to receive payment based on real-time prices in Section III.C, Congestion Rights Market at page 52.

D. Ancillary Services Market

The primary ancillary services are up regulation, down regulation, responsive reserves, and non-spinning reserves. Market participants may self-schedule ancillary services or purchase their required ancillary services through the ERCOT markets. This section reviews the results of the ancillary services markets in 2011.

In general, the purpose of responsive and non-spinning reserves is to protect the system against unforeseen contingencies (*e.g.*, unplanned generator outages, load forecast error, wind forecast error), rather than for meeting normal load fluctuations. ERCOT procures responsive reserves to ensure that the system frequency can quickly be restored to appropriate levels after a sudden, unplanned outage of generation capacity. Non-spinning reserves are provided from slower responding generation capacity, and can be deployed alone, or to restore responsive reserve capacity. Regulation reserves are capacity that responds every four seconds, either increasing or decreasing as necessary to fill the gap between energy deployments and actual system load. Figure 29 displays the quantities of each ancillary service procured each month.

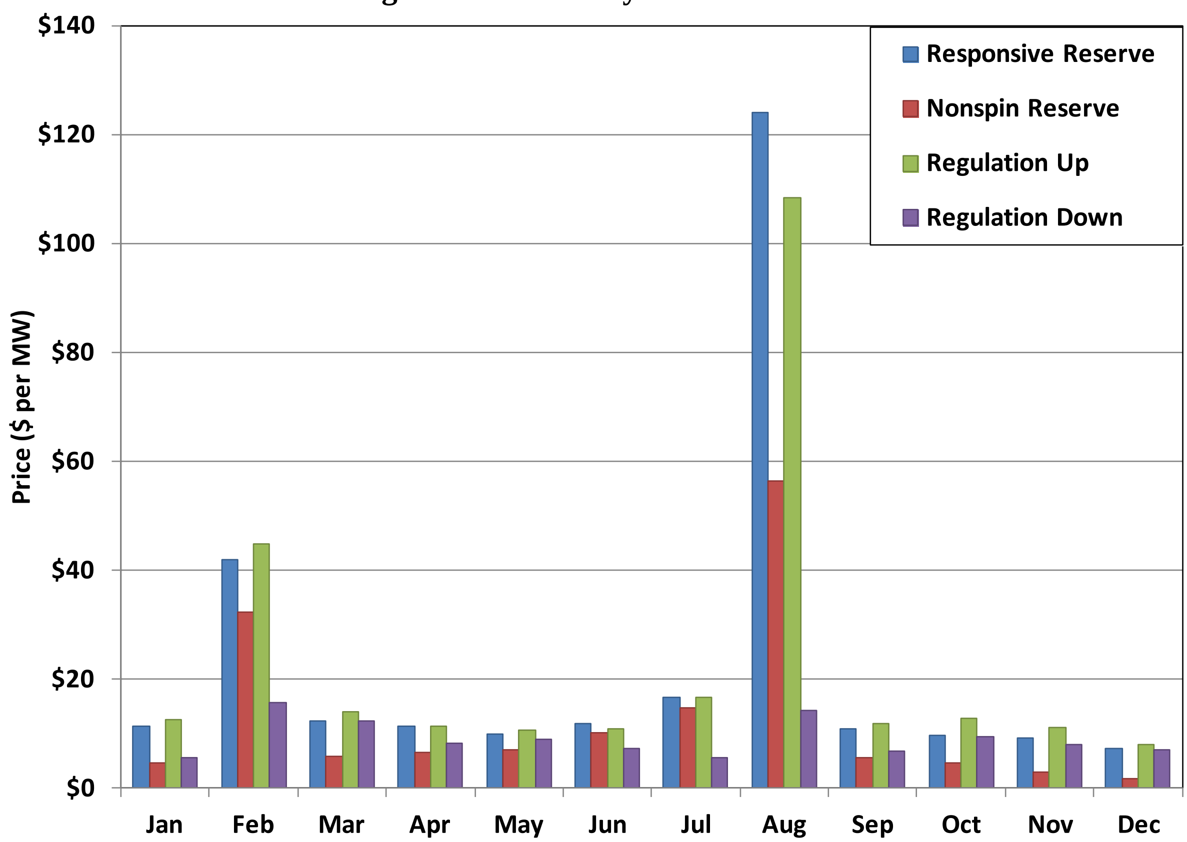
Figure 29: Ancillary Service Capacity



One significant change under the nodal market is that deployments of energy occur more frequently, typically every five minutes. The more frequent deployment of energy has meant that less regulation capacity is required under the nodal market design. Even with the greater quantity of regulation capacity procured in the zonal market, ERCOT operators would resort to issuing out-of-merit instructions to one of the larger generation fleets to provide additional capacity to manage gaps between regulation and balancing energy deployments. This activity occurred 190 times during the last eleven months of the zonal market, typically occurring during periods with large changes in wind generation output. Since the outset of the nodal market there has not been a need to supplement market based deployments by calling on a single market participant, even with increased wind generation ramping.

Another change under the nodal market is that ancillary service offers are co-optimized as part of the day-ahead market clearing. This means that market participants no longer have to include their expectations of forgone energy sales in their ancillary services capacity offers. However, because clearing prices for ancillary services capacity will explicitly account for the value of energy, there is a much higher correlation between ancillary services prices and real-time energy prices. As shown in Figure 30, clearing prices for ancillary services rose quite high during February and August, corresponding with the energy prices in those months.

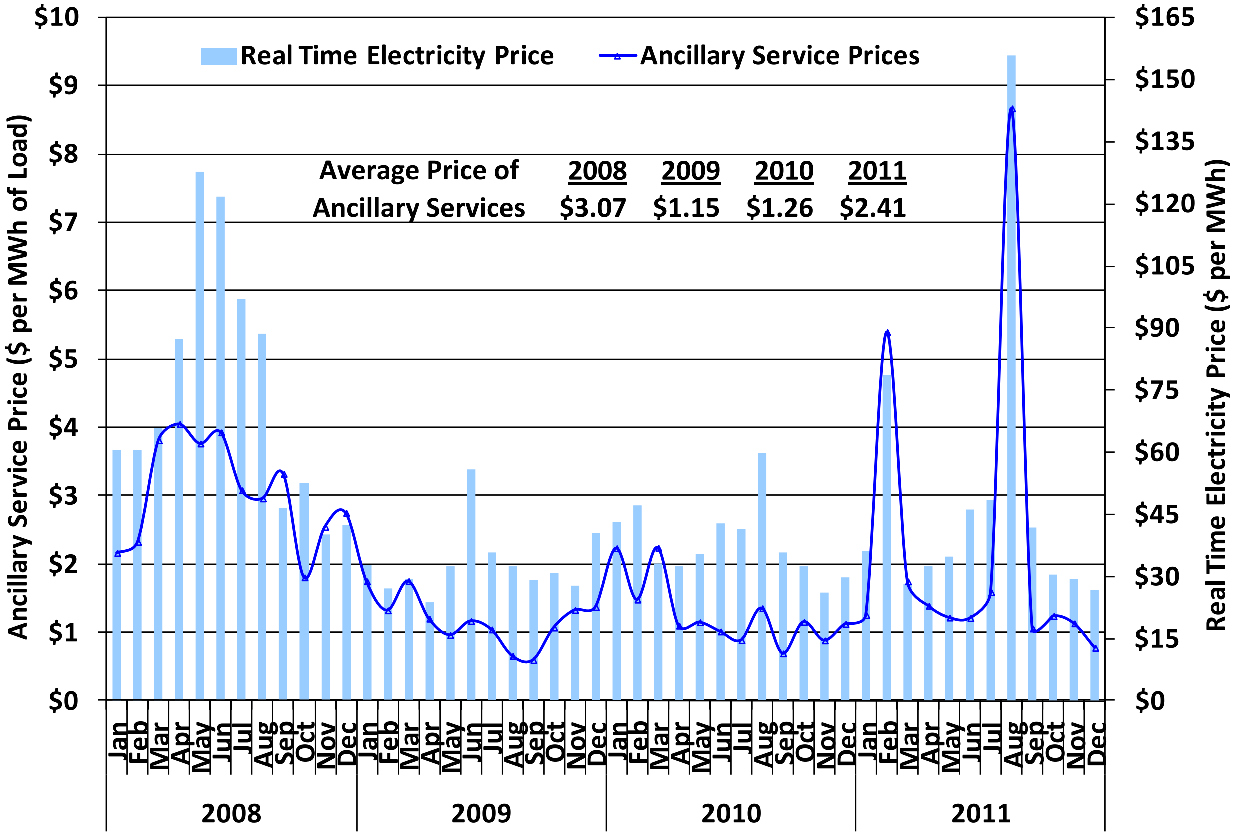
Figure 30: Ancillary Service Prices



In contrast to the previous data that show the individual ancillary service capacity prices, Figure 31 shows the monthly total ancillary service costs per MWh of ERCOT load and the average real-time energy price for 2008 through 2011. Figure 31 shows that total ancillary service costs are generally correlated with real-time energy price movements, which, as previously discussed, are highly correlated with natural gas price movements.

The average ancillary service cost per MWh of load increased to $2.41 per MWh in 2011 compared to $1.26 per MWh in 2010, an increase of 91 percent. Total ancillary service costs increased from 3.2 percent of the load-weighted average energy price in 2010 to 4.5 percent in 2011.

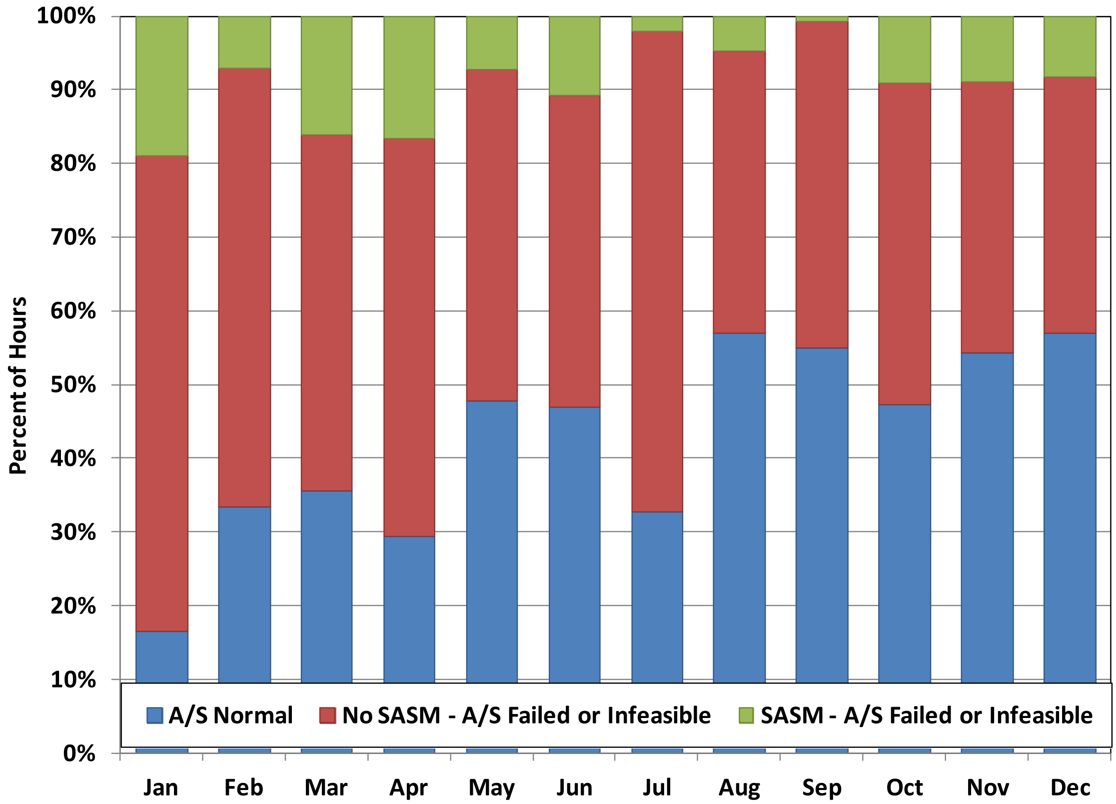
Figure 31: Ancillary Service Costs per MWh of Load



Ancillary Service capacity is procured as part of the day-ahead market clearing. Between the time it is procured and the time that it is required to be provided events can occur which make this capacity unavailable to the system. Transmission constraints can arise which make the capacity undeliverable. Outages or limitations at a generating unit can lead to failures to provide. When either of these situations occurs ERCOT may open a supplemental ancillary services market (SASM) to procure replacement capacity[9].

Figure 32 presents a summary of the frequency with which A/S capacity was not able to be provided and the number of times that a SASM was opened. The percent of time that capacity procured in the day-ahead was actually able to provide the service in the hour it was procured for was less than 20 percent of the time at the beginning of the year, increasing to more than 50 percent by the end of 2011. Even though in more than 40 percent of the hours there were deficiencies in A/S deliveries, SASMs were opened to procure replacement capacity only a fraction of the time.

Figure 32: Frequency of SASM Clearing



In Table 2 below, we provide an annual summary of the frequency and quantity of ancillary service deficiency, where deficiency is defined as either the failure to provide or undeliverability.

9 ERCOT may also open a SASM if they change their ancillary service plan. This did not occur during 2011.

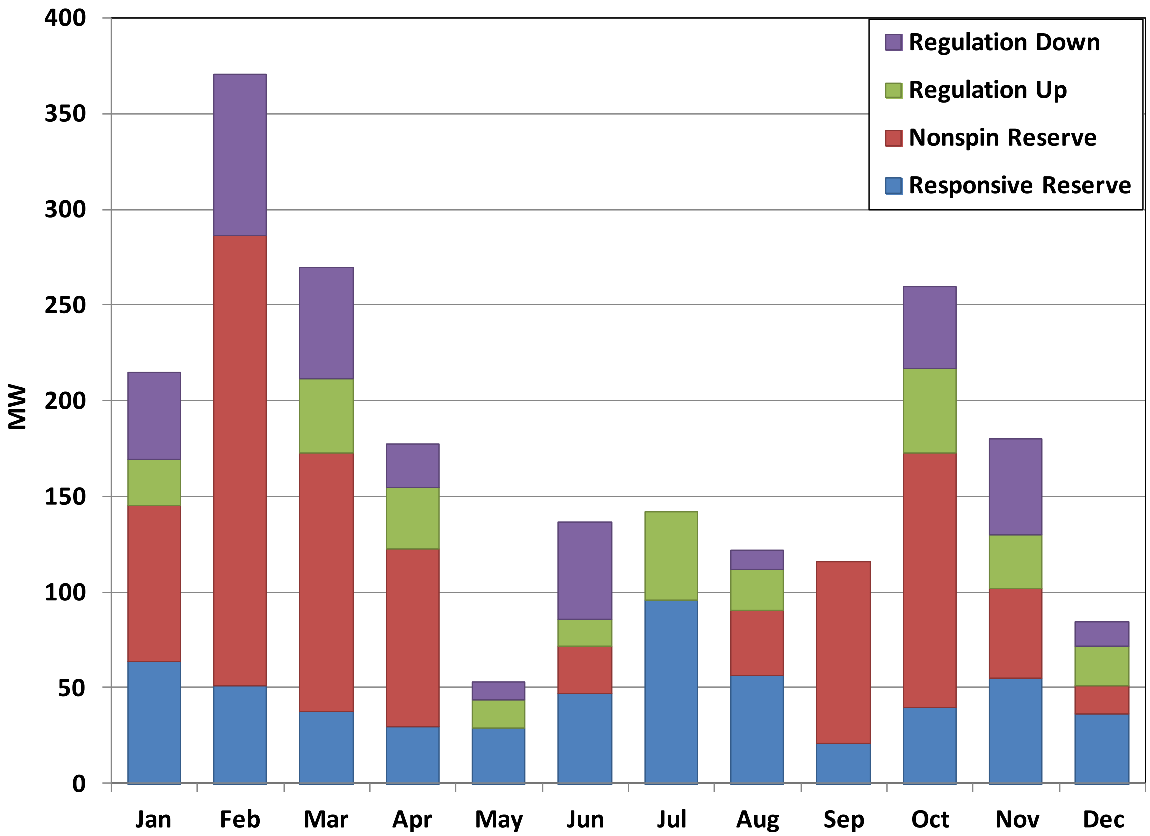
Table 2: Ancillary Service Deficiency

Service	*Hours Deficient*	*Mean Deficiency (MW)*	*Median Deficiency (MW)*
Responsive Reserve	4053	39	20
Up Regulation	1222	27	20
Down Regulation	1235	22	11
Non-Spin Reserve	1254	90	39

Responsive Reserve service was deficient most frequently. As was the case for all ancillary services, the overwhelming majority of time, 4003 out of 4053 hours, the Responsive Reserve deficiency was due to the resource failing to provide, not because of a transmission constraint.

The next analysis, shown in Figure 33 summarizes the average quantity of each service that was procured via SASM.

Figure 33: Ancillary Service Quantities Procured in SASM



III. TRANSMISSION AND CONGESTION

One of the most important functions of any electricity market is to manage the flows of power over the transmission network by limiting additional power flows over transmission facilities when they reach their operating limits. The action taken to ensure operating limits are not violated is called congestion management. The effect of congestion management is to change generator(s) output level so as to reduce the amount of electricity flowing on the transmission line nearing its operating limit. Because the transmission system is operated such that it can withstand the unexpected outage of any element at any time, congestion management actions most often occur when a transmission element is expected to be overloaded if a particular unexpected outage (contingency) were to occur. Congestion leads to higher costs due to lower cost generation being reduced and higher priced generation increased. Different prices at different nodes are the result. With the change to ERCOT's nodal market, the decision about which generator(s) will vary their output is based on generating unit specific offer curves.

Previous annual reports described at great length the inconsistencies and resulting inefficiencies in the bifurcated way in which congestion was managed under the zonal market design. Although zonal congestion management instructions were bid-based, because all generators located within in a zone were assumed to have the same ability to affect the flows across a zonal constraint, the result was significant operational inefficiency and uncertainty. All other constraints were managed by paying generators to either increase or decrease their output from their scheduled level. Because the money to make these payments to generators was collected from all loads in ERCOT, generators had no incentive to take into account the state of the transmission system when scheduling their output.

The nodal market provides many improvements, including unit-specific offers and shift factors, simultaneous resolution of all transmission congestion, actual output instead of schedule-based dispatch, and 5-minute instead of 15-minute dispatch, among others. These changes have helped to increase the economic and reliable utilization of scarce transmission resources beyond that experienced in the zonal market, and in so doing, also dispatch the most efficient resources available to reliably serve demand.

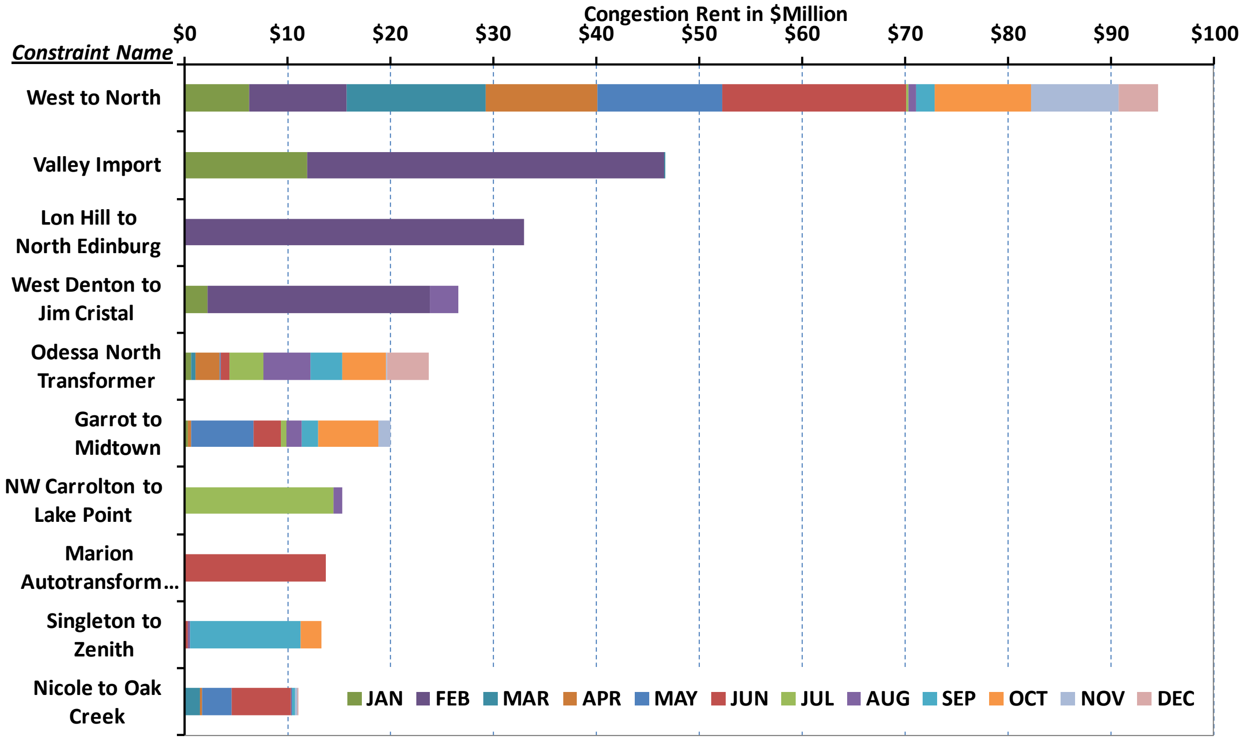
In this section of the report we will start with a review the costs and frequency of transmission congestion in both the day-ahead and real-time markets. We will then provide a review the activity in the congestion rights market.

A. Real-Time Constraints

We begin our review by examining the real-time constraints with the highest financial impact. In all there were more than 300 different constraints active at some point during 2011. The median financial impact, as measured by congestion rent was approximately $300,000.

Figure 34 displays the ten most highly valued real-time constraints as measured by congestion rent and indicates that the West to North interface constraint was by far the most highly valued during 2011.

Figure 34: Top Ten Real-Time Constraints



This constraint is very similar to the competitively significant constraint that existed since the inception of ERCOT's zonal market. Through the years it has been a major impediment to delivering all the wind generation located and produced in the western reaches of ERCOT to the load centers. This constraint was active at some point during every month of 2011.

Two additional constraints on the list are also related to west zone wind generation, although in different directions. The Nicole to Oak Creek constraint is a small capacity 69 kV transmission line that typically overloads under high wind conditions, while due to its load serving nature, the Odessa North 138/69 kV transformer typically overloads under low wind conditions.

The second and third constraints shown in Figure 34 are similar and reflect limitations on the amount of electricity that can be reliably imported into the Rio Grande Valley. This was most notable during the cold weather event of early February. Whereas system wide generation shortages were limited to February 2nd, extremely high customer demands for electricity coupled with the extended planned outage of local generation led to shortages and resulting load curtailments in the Valley over the next two days. Constraints limiting imports to the Valley were active and not able to be resolved for a total of 13 hours during January and February.

When a constraint becomes irresolvable, ERCOT's dispatch software can find no combination of generators to dispatch in a manner such that the flows on the transmission line(s) of concern are below where needed to operate reliably. In these situations, offers from generators are not setting locational prices. Prices are set based on predefined rules, which since there are no supply options for clearing, should reflect the value of reduced reliability for demand. In the case of these constraints related to Valley imports, the shadow price, exceeded $4,500 per MW for the entire time they were not able to be resolved. The effects of these high prices at Valley locations were felt across the entire South Load Zone through high load zone prices.

Although the pricing outcomes were as designed, in the aftermath of this high priced congestion event, ERCOT stakeholders revisited the pricing parameters for irresolvable constraints. The outcome of this effort implemented a set of changes for pricing this constraint specifically and all irresolvable constraints in general. Specifically, due to the radial nature of the Valley Import constraint, its shadow price when irresolvable was lowered to $2,000 per MW. To address the situation more generally, a regional peaker net margin mechanism was introduced such that once local price increases accumulate to a predefined threshold due to an irresolvable constraint; the shadow price of that constraint would drop. The IMM supported this compromise which was finalized late in the year and went into effect for 2012.

One other constraint of note shown in Figure 34 is the Garrot to Midtown 138 kV line, which limits flows within the Houston area. The transmission system upgrade necessary to resolve this longstanding, well known constraint was delayed well past its original planned in-service date. It is scheduled for completion in 2012.

The remaining constraints on the list are all fairly short duration, high impact constraints, reflecting limitations on the ability to import power to a major load center. The Carrolton Northwest to Lake Point 138 kV line and the West Denton to Jim Cristal 138 kV constraints were related to serving load in the DFW area. Singleton to Zenith limited imports to Houston from the north, while the Marion autotransformer constraint was due to an equipment deration that limited electricity flows into San Antonio.

Figure 35: Most Frequent Real-Time Constraints

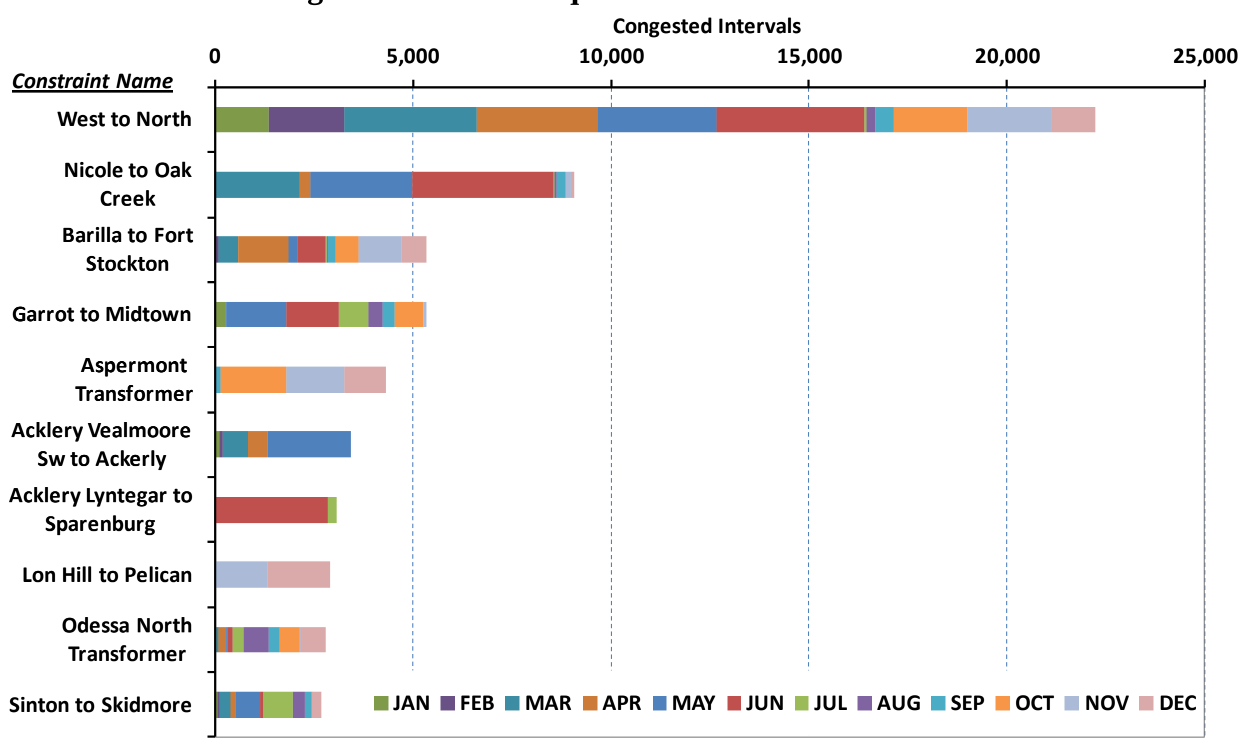
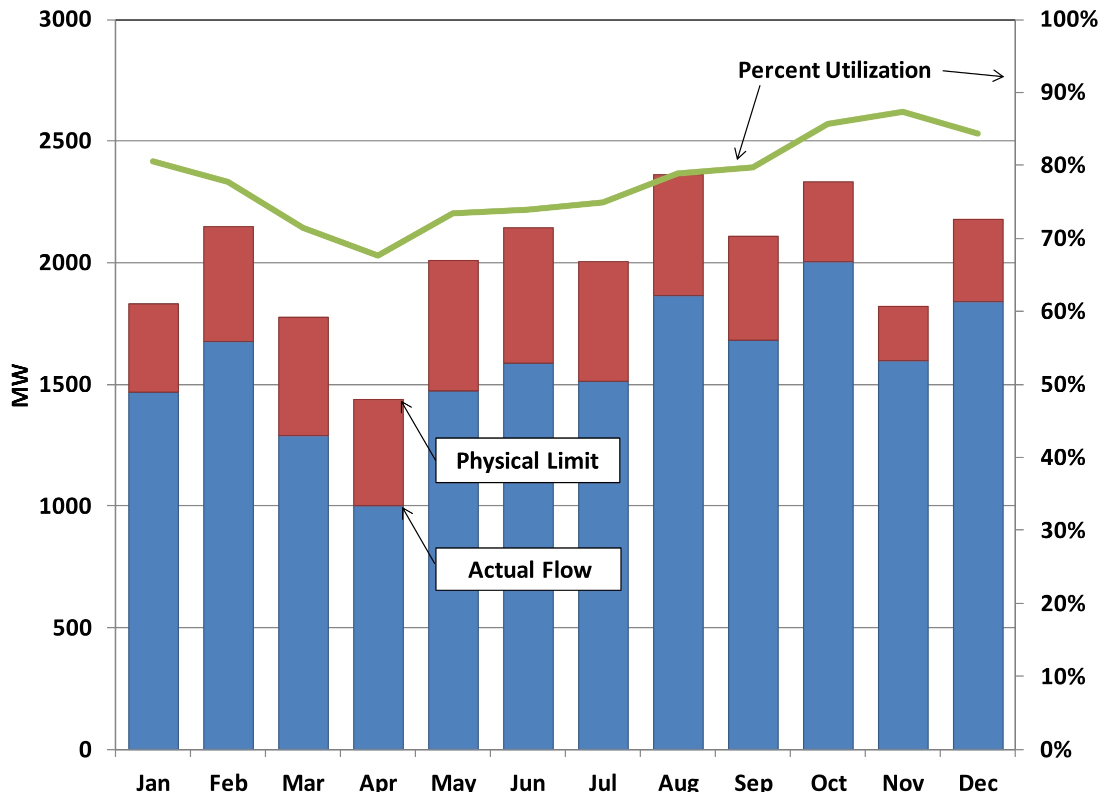


Figure 35 presents a slightly different set of real-time constraints. These are the most frequently occurring. With the exception of Garrot to Midtown, described previously, all are related to wind generation. The Odessa North 138/69 kV transformer typically overloads under low wind and high local load conditions. The Lon Hill to Pelican 138 kV line was a limitation affecting a specific coastal wind generator. The other seven frequently occurring constraints are all related to high west zone wind. Again, the West to North interface constraint tops the list as the most frequently occurring constraint in 2011. To put it in context, the West to North constraint was binding more than 20 percent of the time in 2011.

To maximize the economic use of scarce transmission capacity, the ideal outcome would be for the actual transmission line flows to reach, but to not exceed the physical limits required to maintain reliable operations. Figure 36 presents a summary of the utilization of the most active transmission constraint during 2011, the West to North interface. By comparing the actual flow with the physical limit of the constraint for each dispatch interval it was binding, we can compute its average utilization.

Figure 36: Utilization of the West to North Interface Constraint



Although there was significant variation throughout the year, the average physical limit was slightly less than 2,000 MW and the average actual flow during constrained intervals was approximately 1,500 MW. The average annual utilization of 76 percent compares favorably to 64 percent utilization experienced during the final months of the zonal market. Even more encouraging is the upward trend in utilization observed in the latter part of the year. This increase may be attributed to increased operator confidence that generators, specifically wind generators in this case, will reduce their output as expected when the constraint is active.

There should be opportunity for increased limits in the short term and even higher utilization of this constraint as ERCOT implements more sophisticated real-time analysis of this constraint, rather than relying on off-line studies. Over the long term, the physical limit will increase as CREZ transmission projects are completed.

Although much improved, congestion management in the nodal market has not been perfect. During the spring of 2011, unexpected levels of base point and price oscillations were observed related to congestion management. The initial efforts to resolve the issue were focused on wind generation issues and managing their curtailment related to the West to North constraint. Wind generators, like all generators, are expected to continuously telemeter to ERCOT the maximum capacity output their generator can sustain ("HSL"). When wind units are not curtailed, their HSL is their current output. When wind units are curtailed, their HSL should be the maximum output they could generate if they were not being curtailed.

One of the first changes made to the nodal market systems, implemented in May 2011, was the introduction of a curtailment flag sent by ERCOT to wind generators every interval. This flag replaced the practice which required wind generators to artificially freeze their HSLs for up to 5 minutes prior to each execution of the dispatch software. Although holding the HSLs constant allowed wind generators to compare their received dispatch instructions to determine whether they had been curtailed, using stale data as input to the dispatch software created congestion management challenges.

Although the implementation of the wind curtailment flag was expected to greatly improve management of the West to North constraint, this was not the only constraint where oscillation issues were observed. Investigation by the IMM and ERCOT staff determined that oscillations were being created because ERCOT systems were using two assessments from different moments in time to calculate the generation dispatch required to resolve a constraint. The delay between the two assessments at times exceeded 5 minutes. After identifying this misalignment in constraint management calculations, ERCOT was able to resolve the issue by the end of June.

Since that time, the rare observation of constraint related oscillation has been attributable to a wind generator specific telemetry issue.

B. Day-Ahead Constraints

In this section we review transmission constraints from the day-ahead market. To the extent the model of the transmission system used for the day-ahead market matches the real-time transmission system, and assuming market participants transact in the DAM similarly to how they transact in real-time, we would expect to see the same transmission constraints appear in the day-ahead market as actually occurred during real-time.

Figure 37: Top Ten Day-Ahead Constraints

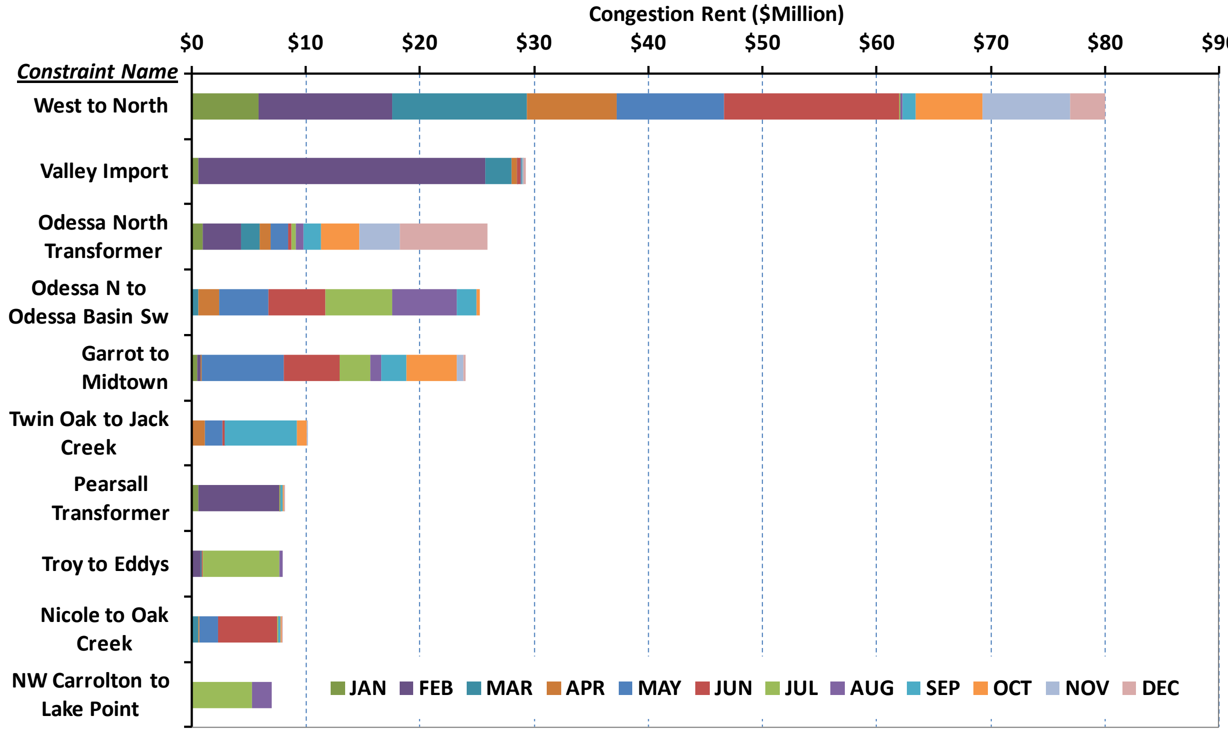


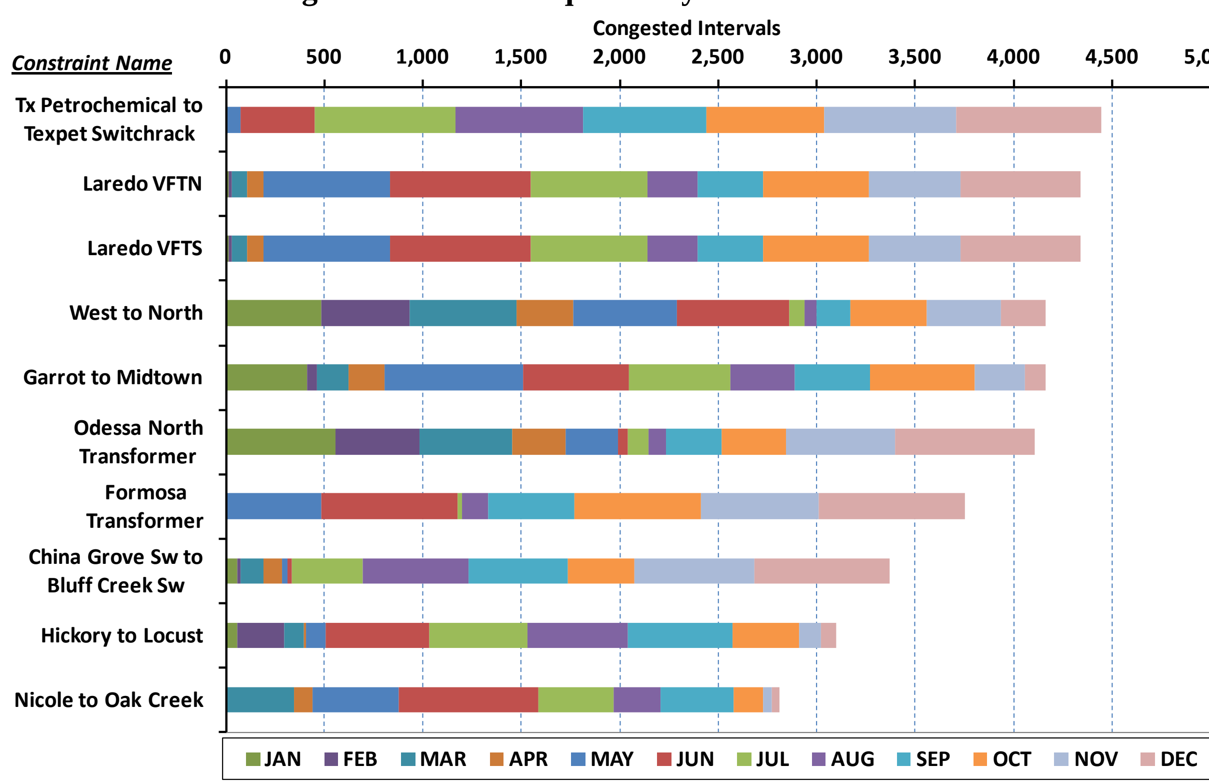
Figure 37 presents the top ten constraints from the day-ahead market, ranked by their congestion rent. As it was in real-time, the West to North constraint has, by far, the highest financial impact. The Valley import constraint, as it was in real-time, was second on the list. The next two constraints are located in Odessa. The limit at the Odessa North transformer is the same constraint which appears in real-time, while the limit on the Odessa North to Odessa Basin Switch 69 kV line is very similar. Three more high impact constraints from real-time also make appear on the day-ahead list. They are the Garrot to Midtown 138 kV line in Houston, the

Nicole to Oak Creek 69 kV line in West Texas, and the Carrollton Northwest to Lakepoint 138 kV line in the Dallas area.

The last three constraints rounding out the list are the Twin Oak to Jack Creek 345 kV line, the Pearsall 138/69 kV Transformer and the Troys to Eddys 69 kV line. Twin Oak to Jack Creek is a constraint that limits imports to Houston from the North. The Pearsall transformer constraint was related to the high impact Valley import congestion early in the year. The Troy to Eddys constraint appeared primarily in July. Although it wasn't specifically a high impact constraint during real-time, there were other similar constraints activated in the Temple / Waco area.

In our final analysis of this section we review the most frequently occurring day-ahead constraints shown in Figure 38. This list includes the now familiar, West to North, Garrot to Midtown, Odessa transformer, and Nicole to Oak Creek constraints.

Figure 38: Most Frequent Day-Ahead Constraints



However, four constraints appearing on the list, including the top three, are constraints that would not occur in real-time. The two Laredo VFT constraints appear frequently as day-ahead constraints, but in real-time operations all transactions with Mexico using these transformers are scheduled using a separate process which would strictly limit their volume.

The Texas Petrochemical to Texpet Switchrack 69 kV line (PR_PRS33) and the Formosa 138/69 kV transformer are both Private Use Network ("PUN") facilities. Though they are represented in ERCOT's transmission network model, because they are privately owned transmission facilities and the action necessary to relieve the constraint would be to redispatch PUN generation, they are not constraints that ERCOT should activate in real-time.

The final two constraints on the list are the China Grove Switch to Bluff Creek Switch 138 kV in west Texas and the Hickory to Locust 69 kV line in the City of Denton.

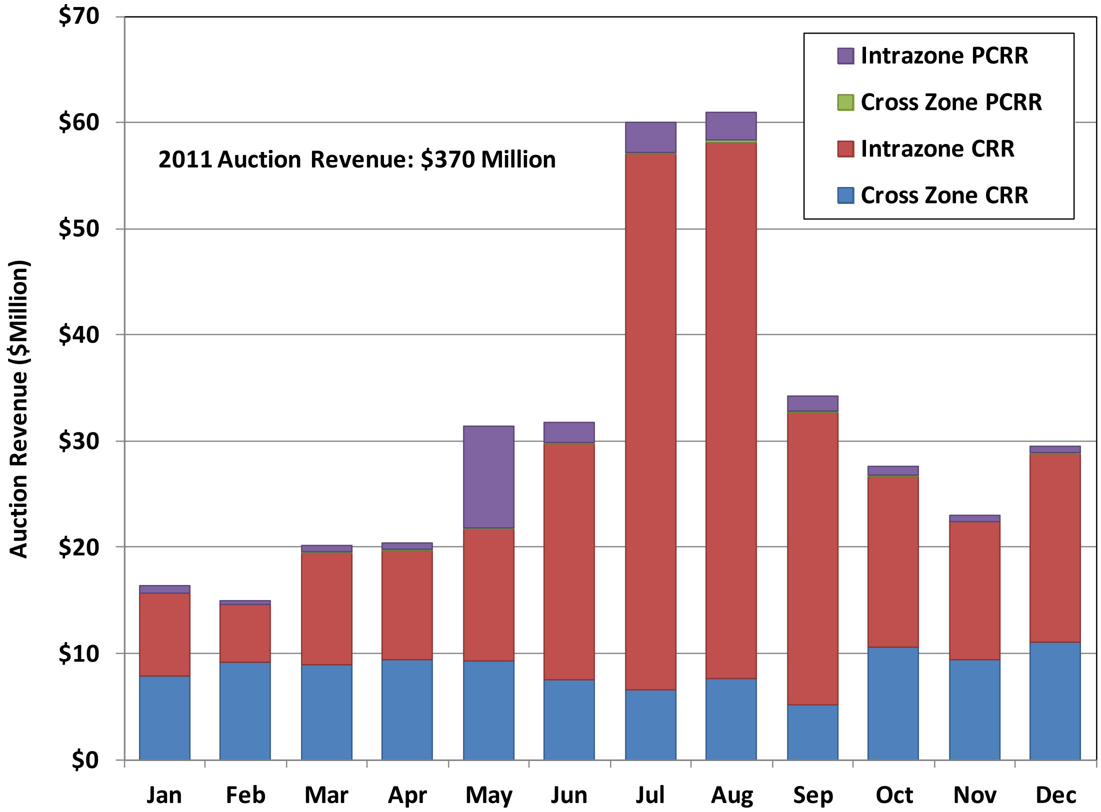
C. Congestion Rights Market

Congestion can be significant from an economic perspective, compelling the dispatch of higher-cost resources because power produced by lower-cost resources cannot be delivered due to transmission constraint(s). Under the nodal market design, one means by which ERCOT market participants can hedge these price differences is by acquiring Congestion Revenue Rights ("CRRs") between any two settlement points. In the zonal market only the costs associated with managing congestion on the commercially significant constraints could be hedged. All other congestion costs were uplifted to all loads.

CRRs are acquired by annual and monthly auctions while Pre-assigned Congestion Revenue Rights ("PCRRs") are allocated to certain participants based on their historical patterns of transmission usage. Parties receiving PCRRs pay only a fraction of the auction value of a CRR between the same source and sink. Both CRRs and PCRRs entitle the holder to payments corresponding to the difference in locational prices of the source and sink.

Figure 39 summarizes the revenues collected by ERCOT in each month for all CRRs, both auctioned and allocated. These revenues are distributed to loads in one of two ways. Revenues from cross zone CRRs are allocated to loads ERCOT wide. Revenues from CRRs that have their source and sink in the same geographic zone are allocated to loads within that zone. This method of revenue allocation provides a disproportionate share of CRR auction revenues to loads located in the West zone. In 2011, CRRs with both their source and sink in the West zone accounted for 25 percent of CRR Auction revenues. This share of revenue was allocated to West zone loads, which accounted for only 7 percent of the ERCOT total.

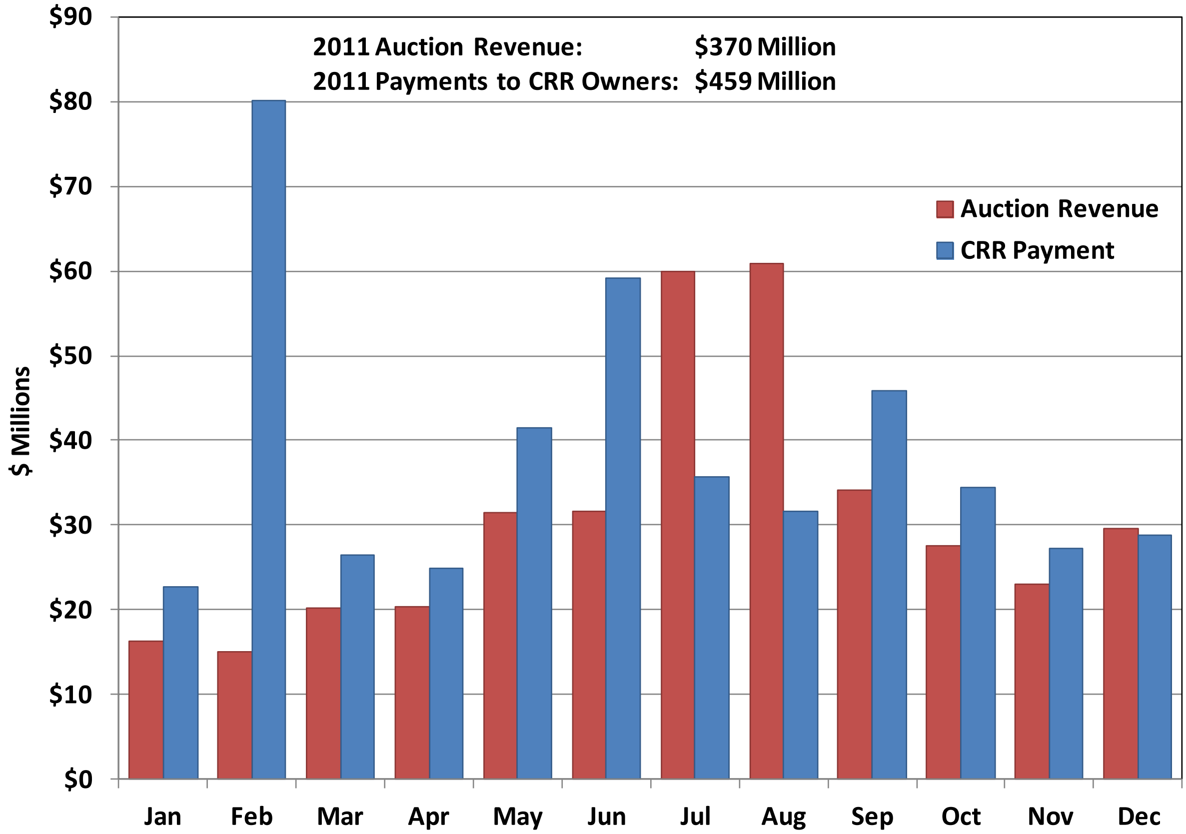
Figure 39: CRR Auction Revenue



Next, in Figure 40 we examine the value CRR owners (in aggregate) received compared to the price they paid to acquire the CRRs. Although results for individual participants and specific source/sink combinations varied, we find that in most months participants did not over pay in the auction. Across the entire year, participants spent $370 million to procure CRRs and received $459 million.

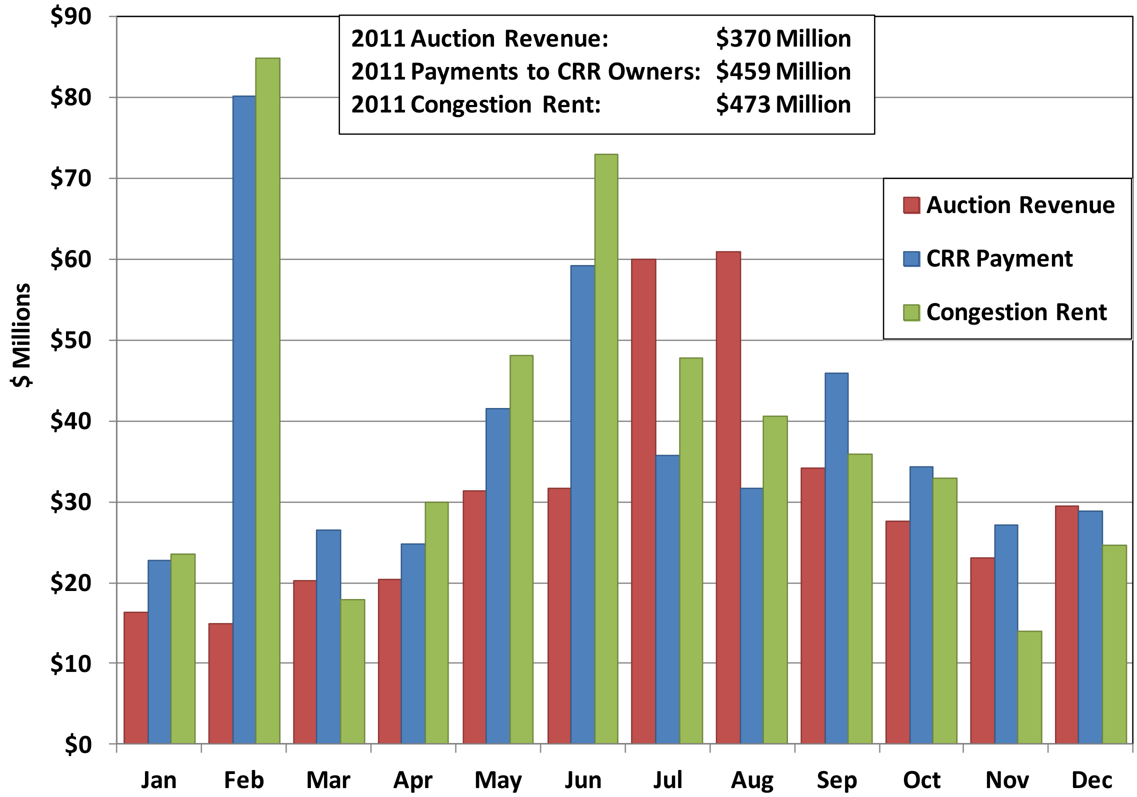
The two months where participants significantly overpaid to acquire CRRs were July and August. The amount paid to procure CRRs in July and August seems reasonable since it was very similar to the payments received by CRR owners in June. The smaller payment to CRR owners in July and August may have been a result of ERCOT operators choosing not to activate particular transmission constraints at times when the system was experiencing, or was anticipated to experience scarcity conditions.

Figure 40: CRR Auction Revenue and Payment Received



This may seem like an appropriate action for ERCOT operators to take so as to not reduce the amount of capacity available to the system. However, pricing parameters in ERCOT's dispatch software are set such that if there is indeed a scarcity situation, the software will allow transmission constraints to be violated if capacity is needed to meet power balance requirements. Because the dispatch software will effectively prioritize between managing transmission constraints and balancing supply and demand, we recommend that ERCOT reconsider its practice of deactivating transmission constraints in the dispatch software during peak demand conditions.

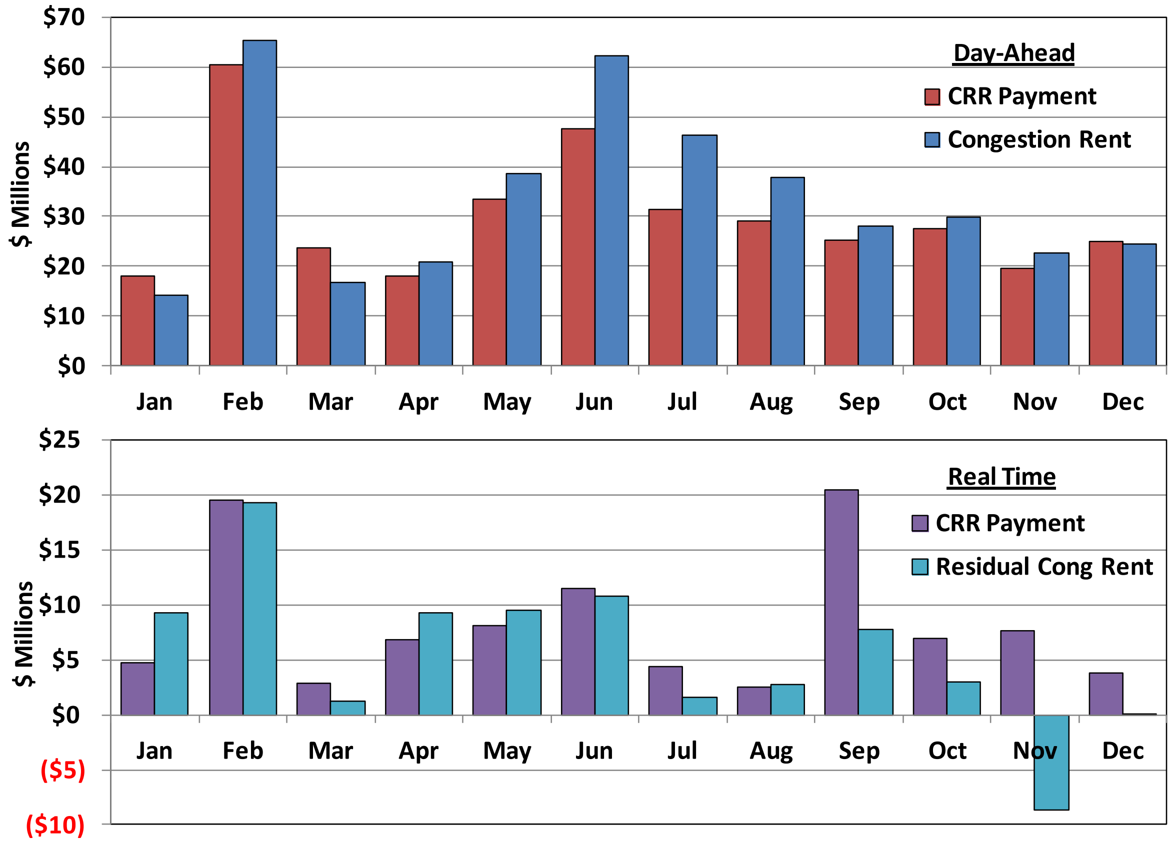
Figure 41: CRR Auction Revenue, Payments and Congestion Rent



In our next look at aggregated CRR positions, we add congestion rent to the picture. Simply put, congestion rent is the difference between the total costs that loads pay and the total revenue that generators receive. Congestion rent creates the source of funds used to make payments to CRR owners. Figure 41 presents all three values for each month of 2011. For the year, congestion rent totaled $473 million and payments to CRR owners were $459 million. However, in March and September through December congestion rent was less than payments to CRR owners.

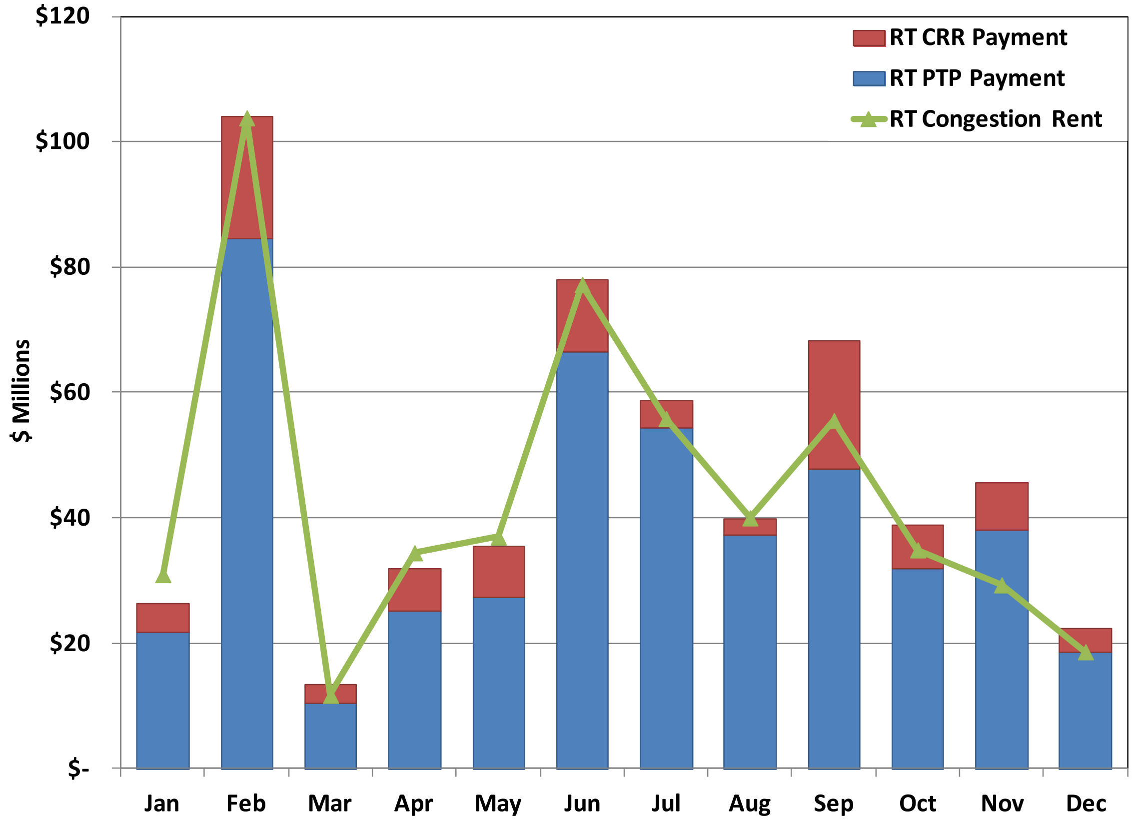
We further analyze the relationship between congestion rent and payments to CRR owners by separating the impacts of CRRs that are settled based on day-ahead prices from the subset of CRRs that are paid based on real-time prices.

Figure 42: Day-Ahead and Real-Time Congestion Payments and Rent



The top portion of Figure 42 displays the comparison of day-ahead congestion rent to payments received by CRR owners. Congestion rent is larger than payments in most months, and for the year rent is $407 million compared to $359 million that was paid to CRR owners. The bottom portion of Figure 42 presents a different view. For this analysis we have assumed that all PTP Obligations have been fully funded from real-time congestion rent and any residual real-time congestion rent is available to fund payments to the subset of CRR owners that have elected to have their CRRs be settled based on real-time prices. With this assumption there was $66 million in residual congestion rent available to fund real-time CRR payments of $99 Million. Hence, real-time congestion rent is insufficient to fund all PTP Obligations and CRRs being settled in real-time. The next figure shows this explicitly.

Figure 43: Real-Time Congestion Payments



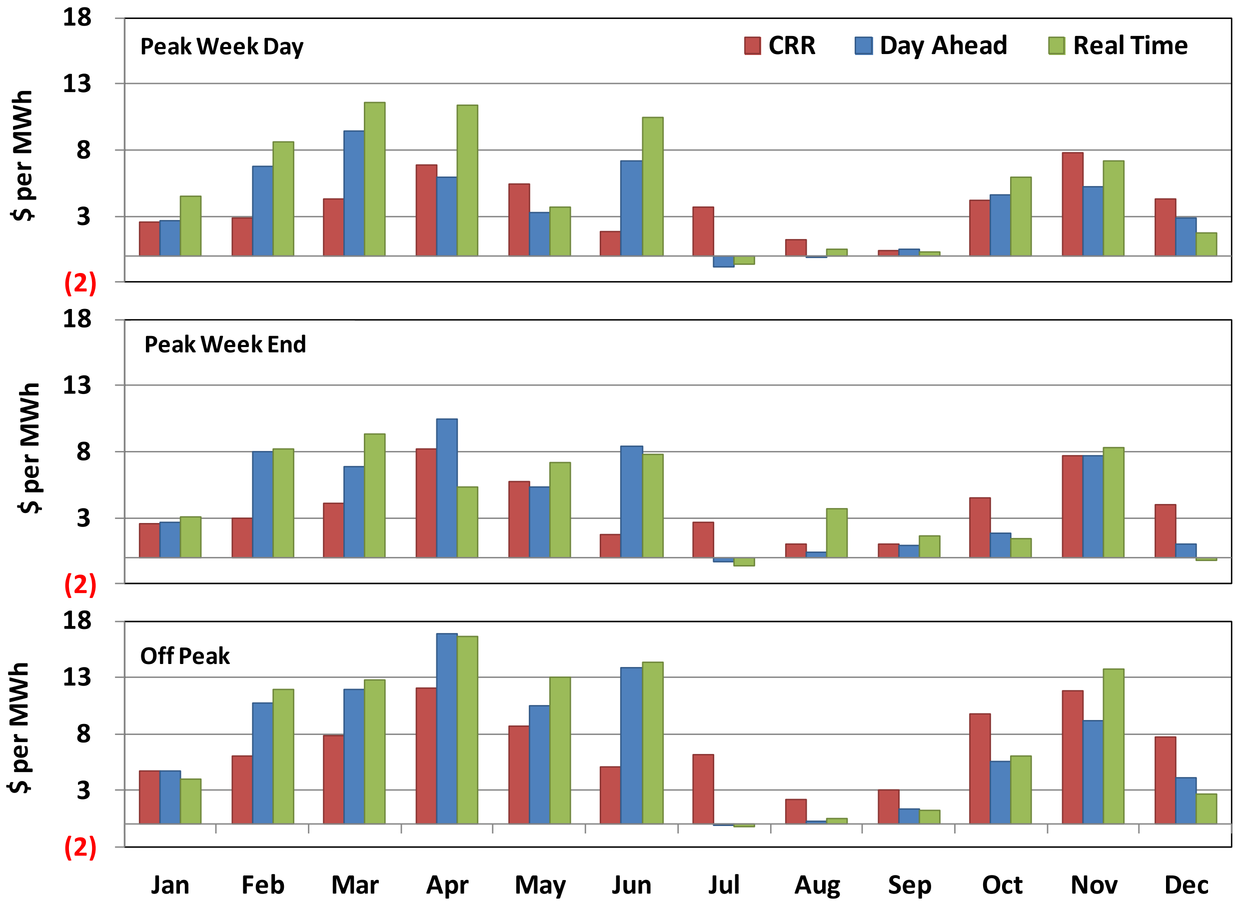
In Figure 43 the combined payments to PTP Obligation owners and CRR owners that have elected to receive real-time payments are compared to the total real-time congestion rent. For the year, real-time congestion rent was $529 million, payments for PTP Obligations were $463 million and payments for real-time CRRs were $99 million, resulting in a shortfall of approximately $33 million for the year.

This type of shortfall typically results from discrepancies between transmission topology assumptions used when clearing the day-ahead market and the actual transmission topology that exists during real-time. Specifically, if the day-ahead topology assumptions allows too many real-time congestion instruments (PTP Obligations and CRRs settled at real-time prices) with impacts on a certain path, and that path constrains at a lower limit in real-time, there will be insufficient rent generated to pay all of the congestion instruments.

From Figure 43 we can see that September and November were the months with the most noticeable deficiencies. During September there were multiple forced outages of major transmission facilities due to wild fires. These outages could not reasonably be anticipated and resulted in significant, short duration real-time congestion. In November there was a problem with the transmission modeling around the North DC Tie that allowed PTP Obligations to be sold between two points that should have been deemed "electrically close".

For our last look at congestion we examine the impacts of the West to North constraint in more detail. Figure 44 presents the price spreads between the West Hub and North load zone as valued at three separate points in time – at the monthly CRR auction, day-ahead and in real-time.

Figure 44: West Hub to North Load Zone Price Spreads



Since CRRs are sold for one of three defined time periods, weekday on peak, weekend on peak, and off peak, Figure 44 includes a separate comparison for each.

As expected, most real-time congestion, as evidenced by the largest price spread, occurred in the off peak period, for the months of February through June and November. The day-ahead price spreads were very similar for this period, but the prices paid for CRRs were generally less than the value received. Conversely, during the months of July through September, there was very little congestion. In July day-ahead and real-time prices were higher at the West Hub, which the results of the CRR auction did not anticipate.

IV. LOAD AND GENERATION

This section reviews and analyzes the load patterns during 2011 and the existing generating capacity available to satisfy the load and operating reserve requirements. We provide specific analysis of the large quantity of installed wind generation and conclude this section with a discussion of the daily generation commitment process.

A. ERCOT Loads in 2011

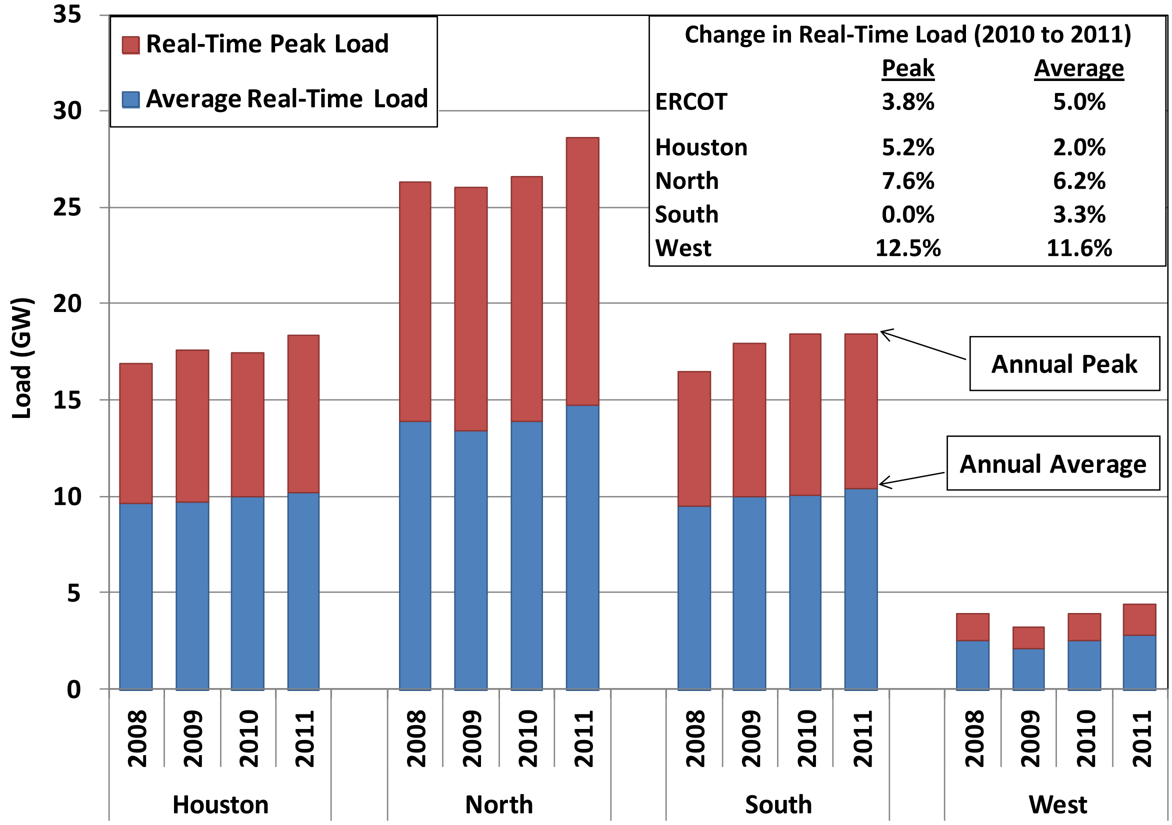
The changes in overall load levels from year to year can be shown by tracking the changes in average load levels. This metric will tend to capture changes in load over a large portion of the hours during the year. It is also important to separately evaluate the changes in the load during the highest-demand hours of the year. Significant changes in these peak demand levels have historically been very important and played a major role in assessing the need for new resources. They also affect the probability and frequency of shortage conditions (*i.e.*, conditions where firm load is served but minimum operating reserves are not maintained). The expectation of resource adequacy is based on the value of electric service to customers and the harm and inconvenience to customers that can result from interruptions to that service. Hence, both of these dimensions of load during 2011 are examined in this subsection and summarized in Figure 45.

This figure shows peak load and average load in each of the ERCOT zones from 2008 to 2011[10]. In each zone, as in most electrical systems, peak demand significantly exceeds average demand. The North Zone is the largest zone (with about 39 percent of the total ERCOT load); the South and Houston Zones are comparable (27 percent) while the West Zone is the smallest (7 percent of the total ERCOT load).

Figure 45 also shows the annual non-coincident peak load for each zone. This is the highest load that occurred in a particular zone for one hour during the year; however, the peak can occur in different hours for different zones. As a result, the sum of the non-coincident peaks for the zones was greater than the annual ERCOT peak load.

[10] For purposes of this analysis NOIE Load Zones have been included with the proximate geographic Load Zone.

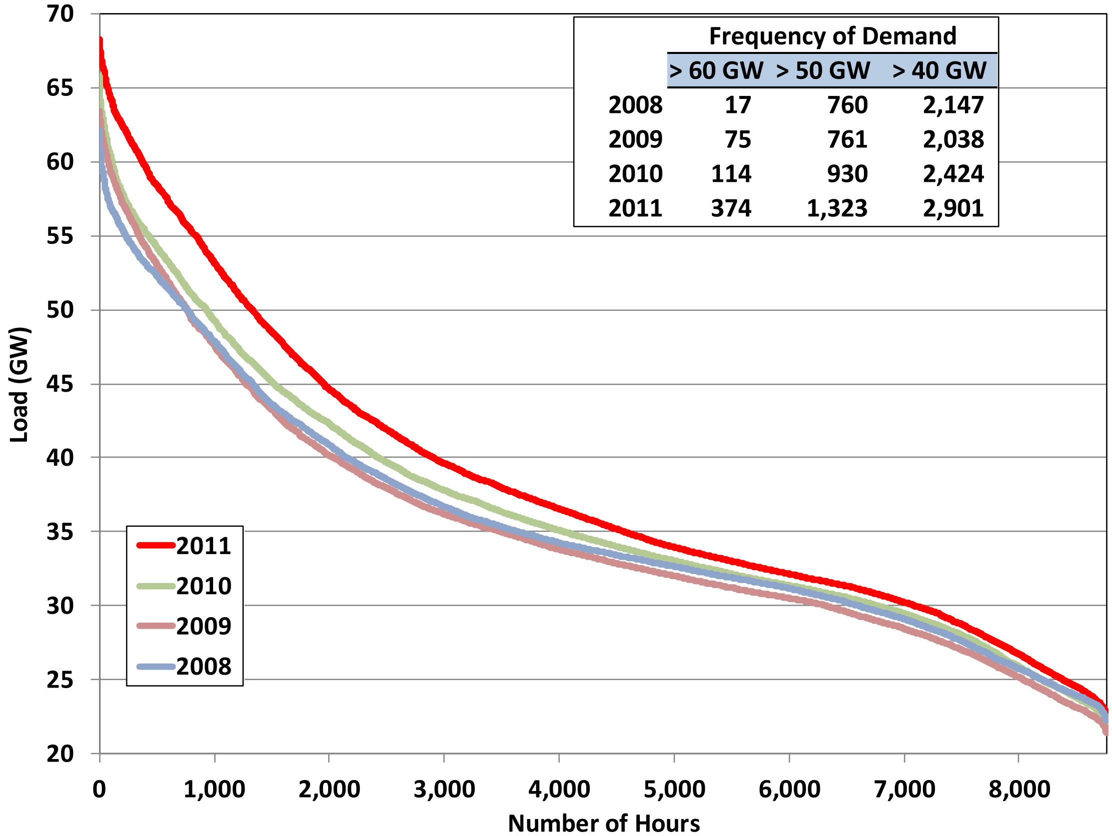
Figure 45: Annual Load Statistics by Zone



Total ERCOT load increased from 319 TWh in 2010 to 335 TWh in 2011, an increase of 5.0 percent or an average of approximately 1,800 MW every hour. Similarly, the ERCOT coincident peak hourly demand increased from 65,776 MW in 2010 to 68,379 MW, an increase of roughly 2,600 MW, or 4.0 percent.

To provide a more detailed analysis of load at the hourly level, Figure 46 compares load duration curves for each year from 2008 to 2011. A load duration curve shows the number of hours (shown on the horizontal axis) that load exceeds a particular level (shown on the vertical axis). ERCOT has a fairly smooth load duration curve, typical of most electricity markets, with low to moderate electricity demand in most hours, and peak demand usually occurring during the late afternoon and early evening hours of days with exceptionally high temperatures.

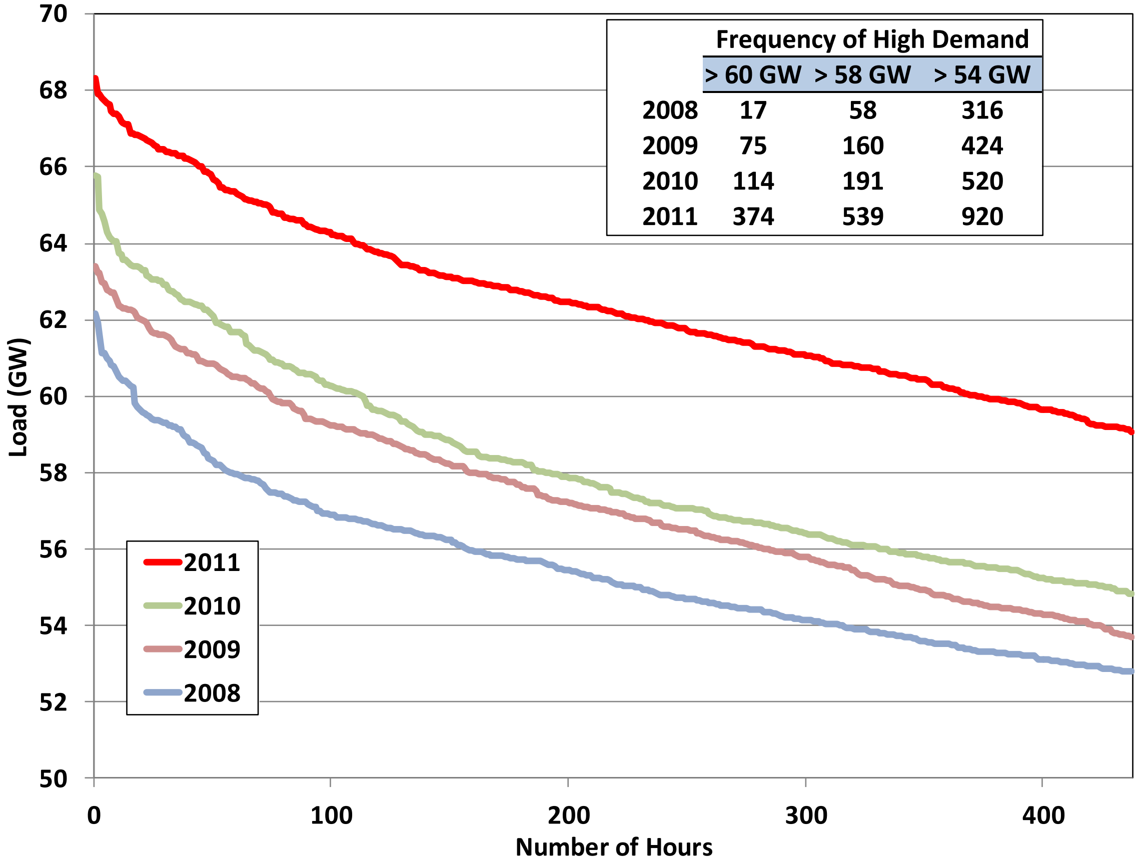
Figure 46: Load Duration Curve – All hours



As shown in Figure 46, the load duration curve for 2011 is significantly higher than in 2010 across all hours of the year. This is consistent with the aforementioned 5.0 percent load increase from 2010 to 2011.

To better illustrate the differences in the highest-demand periods between years, Figure 47 shows the load duration curve for the five percent of hours with the highest loads. This figure also shows that the peak load in each year is significantly greater than the load at the 95th percentile of hourly load. From 2008 to 2011, the peak load value averaged 18 percent greater than the load at the 95th percentile. These load characteristics imply that a substantial amount of capacity – approximately 10 GW – is needed to supply energy in less than 5 percent of the hours.

Figure 47: Load Duration Curve – Top five percent of hours

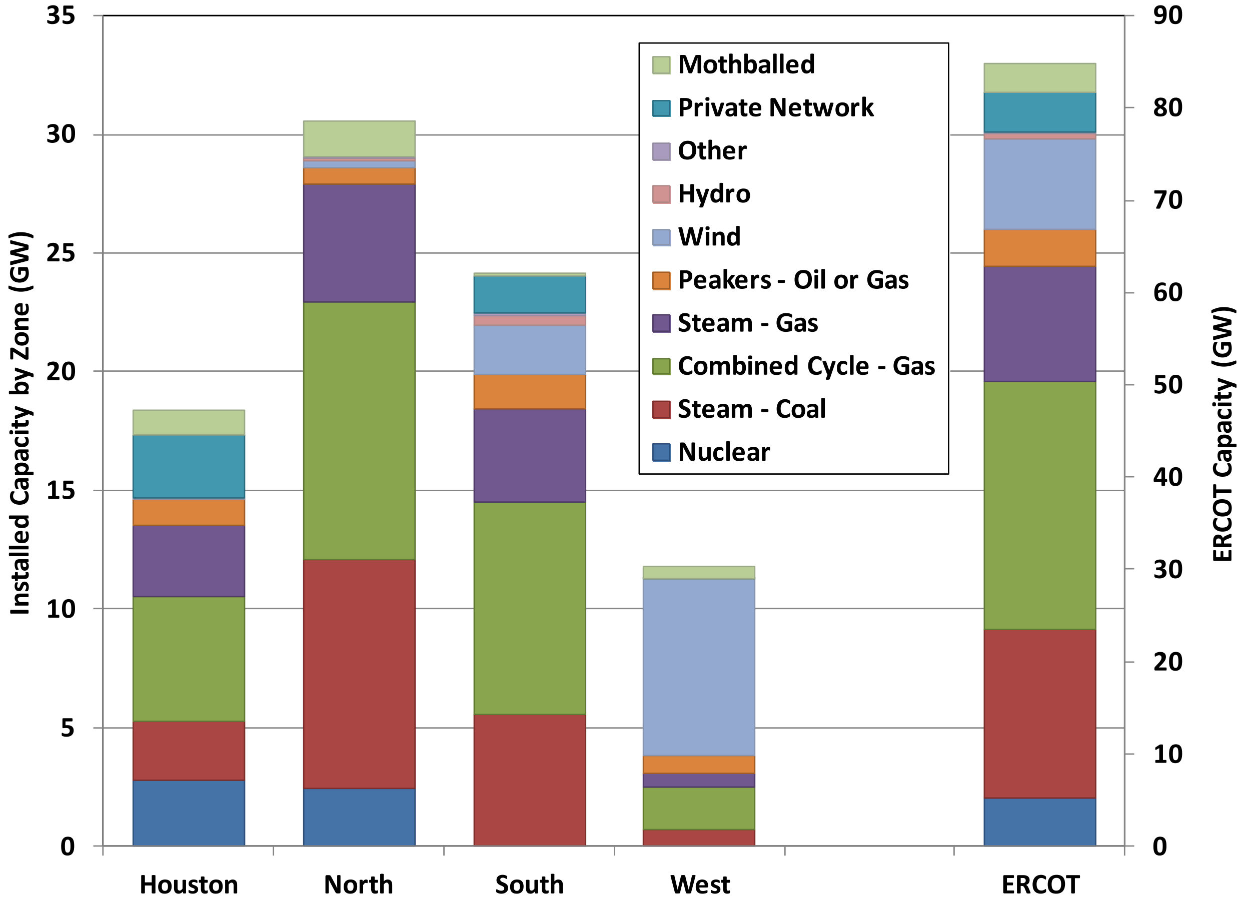


B. Generation Capacity in ERCOT

In this section we evaluate the generation mix in ERCOT. The distribution of capacity among the ERCOT zones is similar to the distribution of demand with the exception of the large amount of wind capacity in the West Zone. The North Zone accounts for approximately 36 percent of capacity, the South Zone 28 percent, the Houston Zone 22 percent, and the West Zone 14 percent. The Houston Zone typically imports power, while the West Zone typically exports power. Excluding mothballed resources and including only 8.7 percent of wind capacity as capacity available to reliably meet peak demand, the North Zone accounts for approximately 40 percent of capacity, the South Zone 30 percent, the Houston Zone 24 percent, and the West Zone 6 percent. Figure 48 shows the installed generating capacity by type in each of the ERCOT zones[11].

[11] For purposes of this analysis, generation located in a NOIE Load Zone has been included with the proximate geographic Load Zone

Figure 48: Installed Capacity by Technology for each Zone

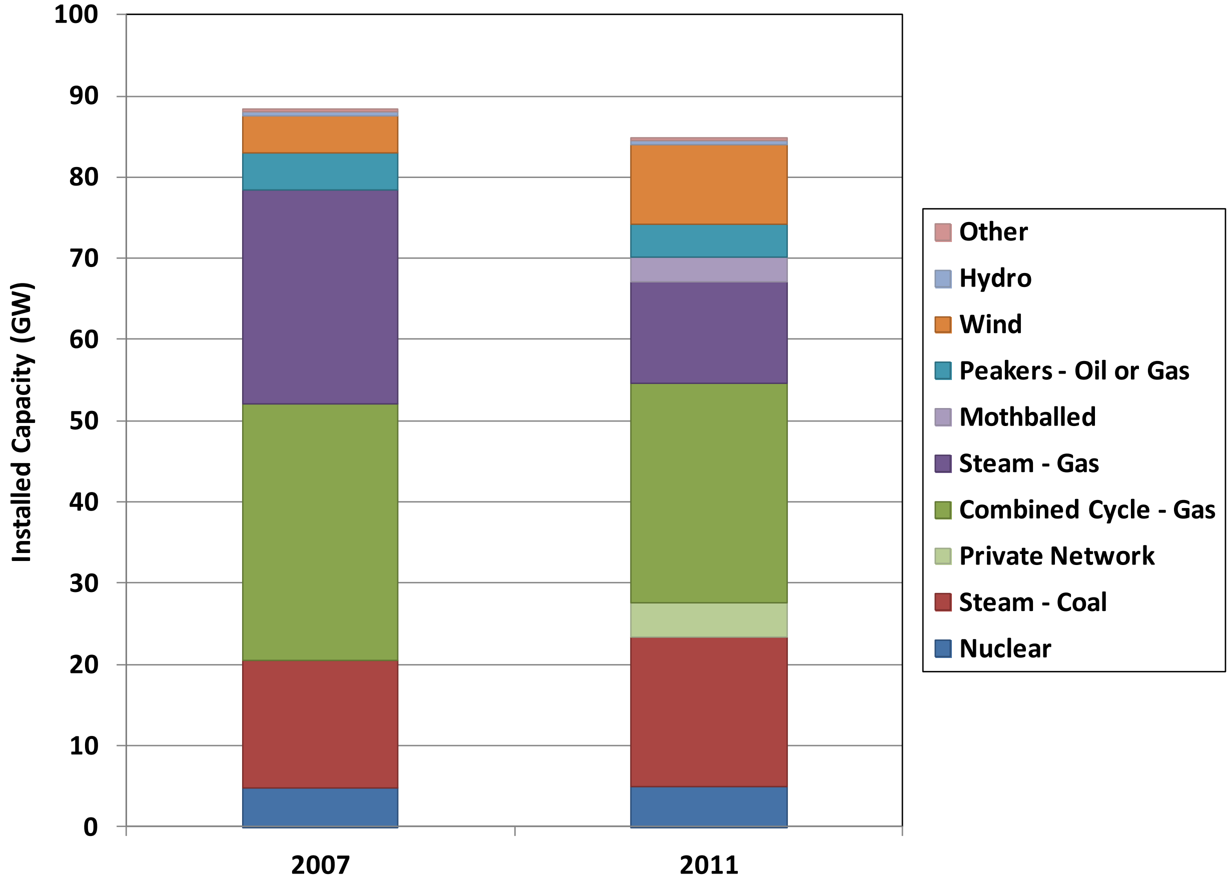


There were very few new units placed in service during 2011. Notable changes to ERCOT's installed generation during 2011 included two coal units being mothballed[12], additions of a new combined cycle unit and a wind unit. Even after these changes natural gas generation accounts for approximately 50 percent of the installed capacity in ERCOT.

By comparing the current mix of installed generation capacity to that in 2007, as shown in Figure 49, we can see the effects of longer term trends. Over these five years wind and coal generation are the only two categories with increased capacity. However, the sizable additions in these two categories have been more than offset by retirements of natural gas fueled steam units, resulting in less installed capacity in 2011 than there was in 2007.

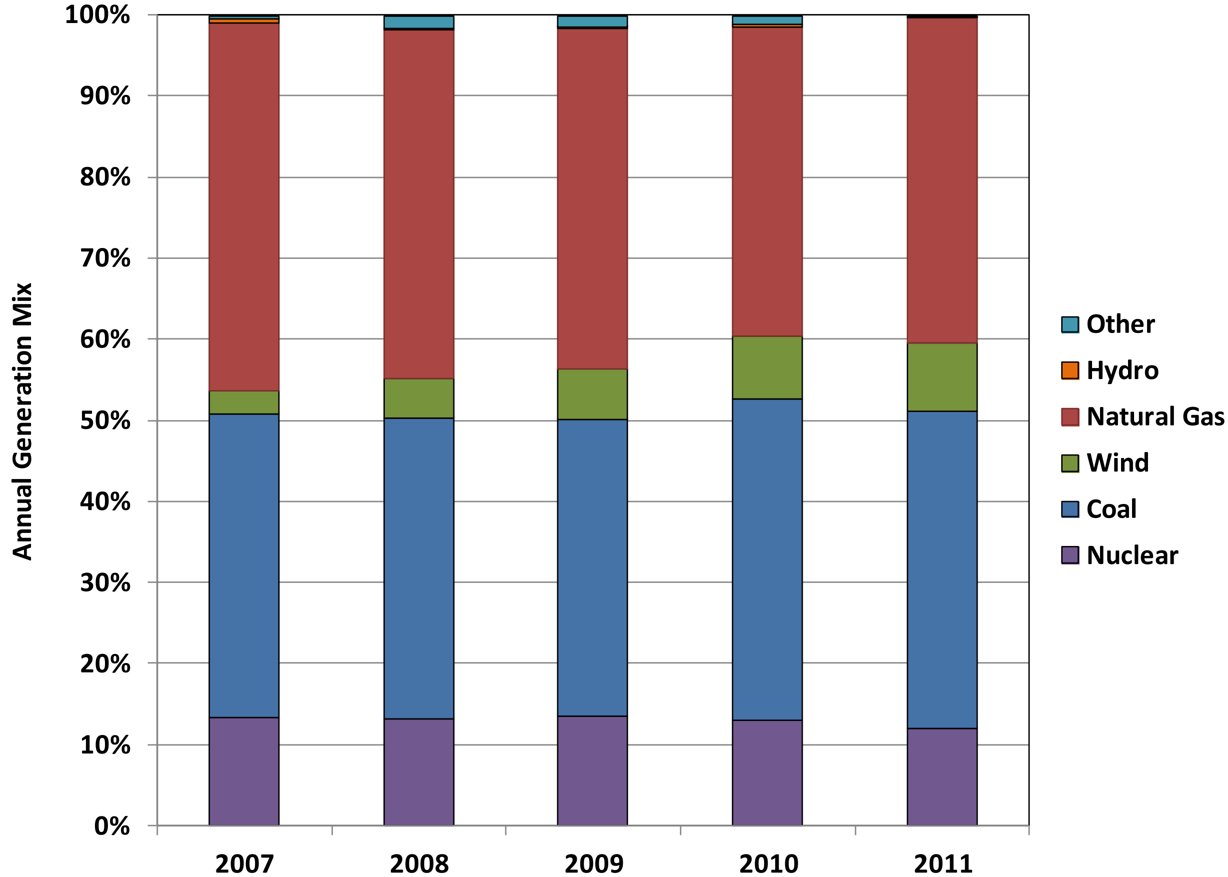
12 The mothball designation for these two units was subsequently rescinded in January 2012.

Figure 49: Installed Capacity by Type: 2007 to 2011



The shifting contribution of coal and wind generation is evident in Figure 50, which shows the percent of annual generation from each fuel type for the years 2007 through 2011. Over the five years shown, the percentage of generation produced by coal units increased slightly from 37 percent to 39 percent. Wind's generation share has increased every year, reaching 9 percent of the annual generation requirement in 2011, up from 3 percent in 2007. During the same period the percentage of generation provided by natural gas decreased from 45 percent to 40 percent.

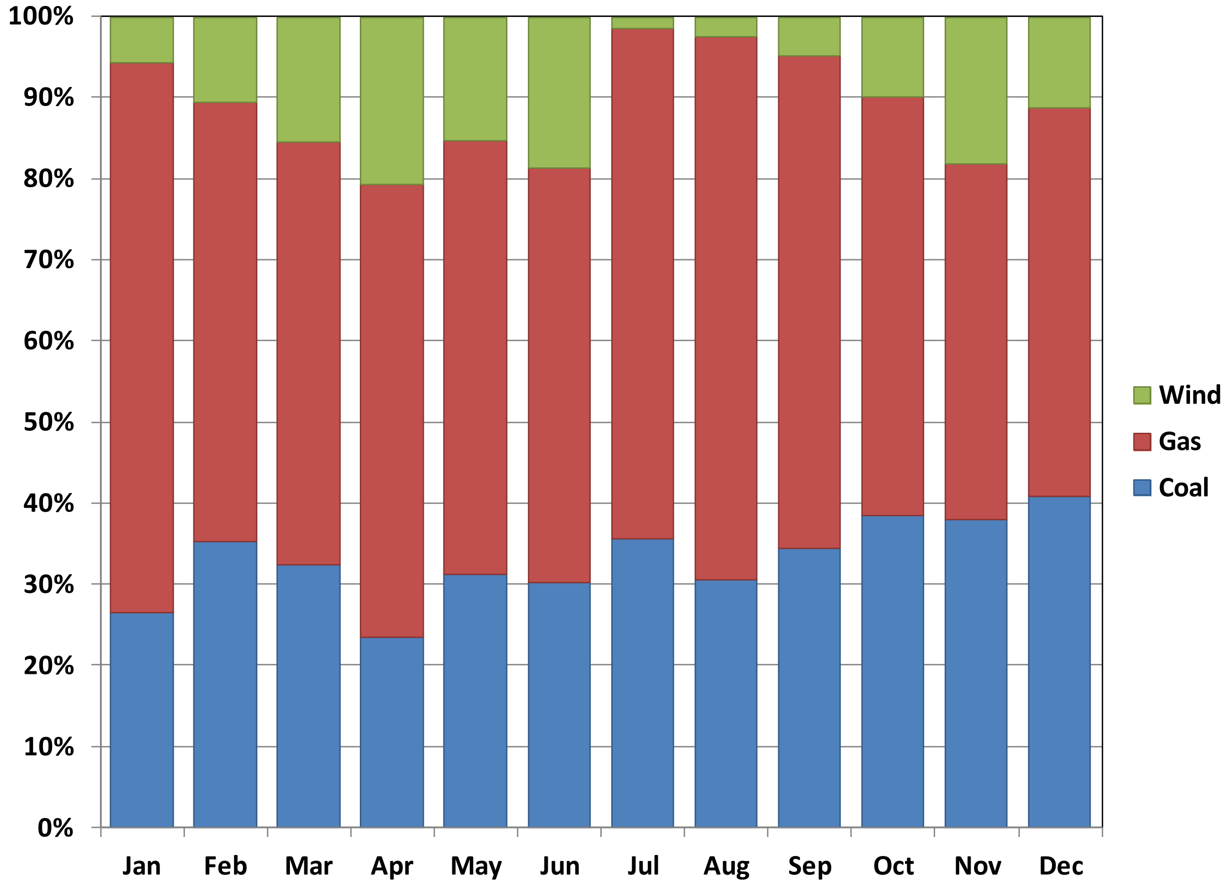
Figure 50: Annual Generation Mix



While ERCOT has coal/lignite and nuclear plants that operate primarily as base load units, its reliance on natural gas resources drives the high correlation between real-time energy prices and the price of natural gas fuel. There is approximately 25 GW of coal and nuclear generation in ERCOT. Generally, when ERCOT load is above this level, natural gas resources will be on the margin and set the real-time energy spot price. Although coal-fired and nuclear units combined to produce more than half of the energy in ERCOT, they have historically played a much less significant role in setting spot electricity prices. However, with the recent additions of new coal generation combined with continuing increases in wind capacity, with its low marginal production, the frequency at which coal and lignite are the marginal units in ERCOT was expected to increase in the future, particularly during the off-peak hours in the spring and fall, and even more as additional transmission capacity is added that will accommodate increased levels of wind production in the West Zone. This expectation is currently tempered by the impacts on existing coal units from impending additional environmental regulations and continuing low natural gas prices.

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



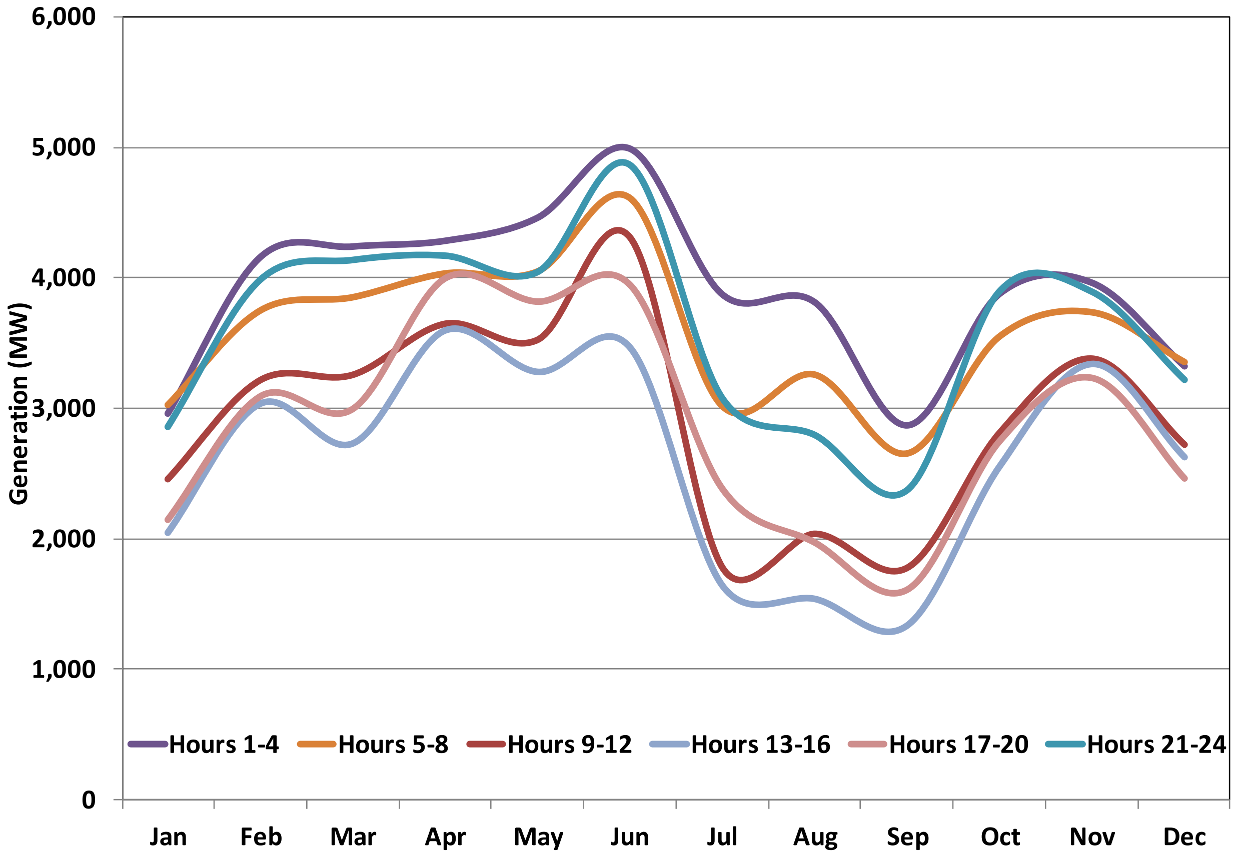
The methodology used in this analysis has been revised to reflect the details of the unit specific dispatch that are available under the nodal market design. For every five-minute interval we determine which units are marginal, that is they are being dispatched and their offer price is contributing to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

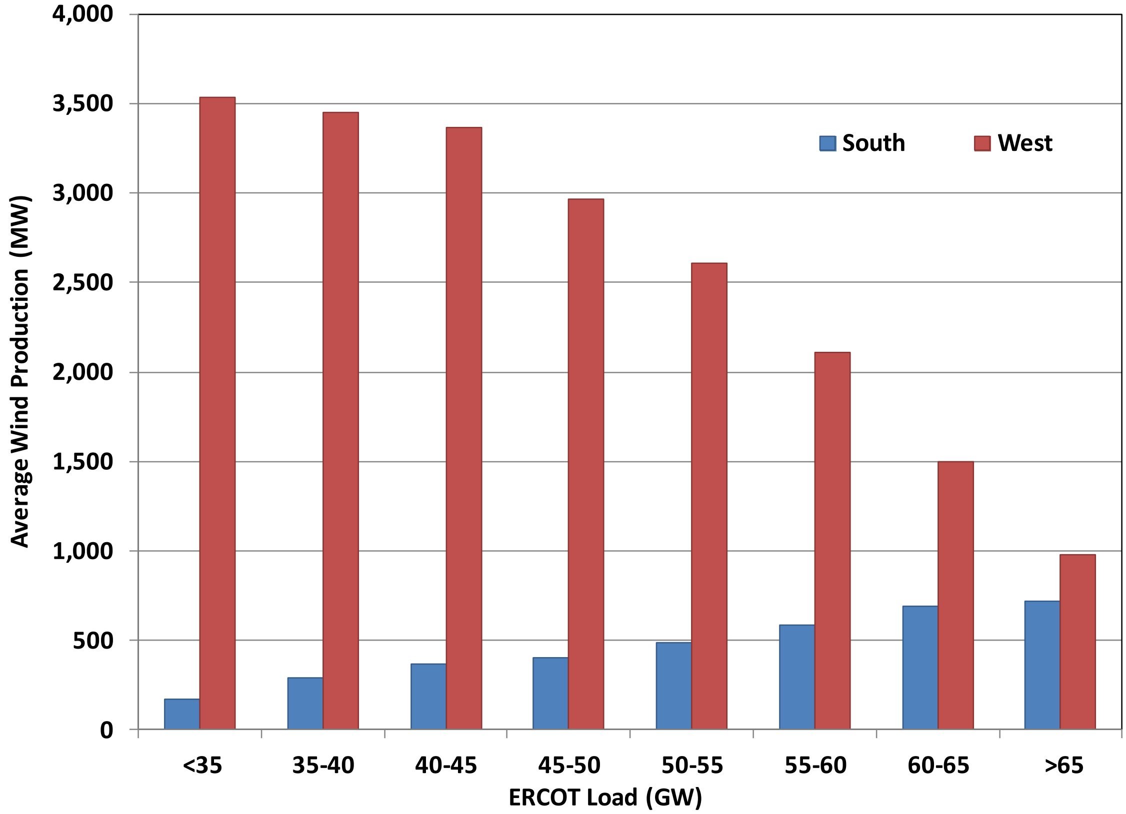


The average profile of wind production is negatively correlated with the load profile, with the highest wind production occurring during non-summer months, and predominately during off-peak hours. Figure 52 shows average wind production for each month in 2011, with the average production in each month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

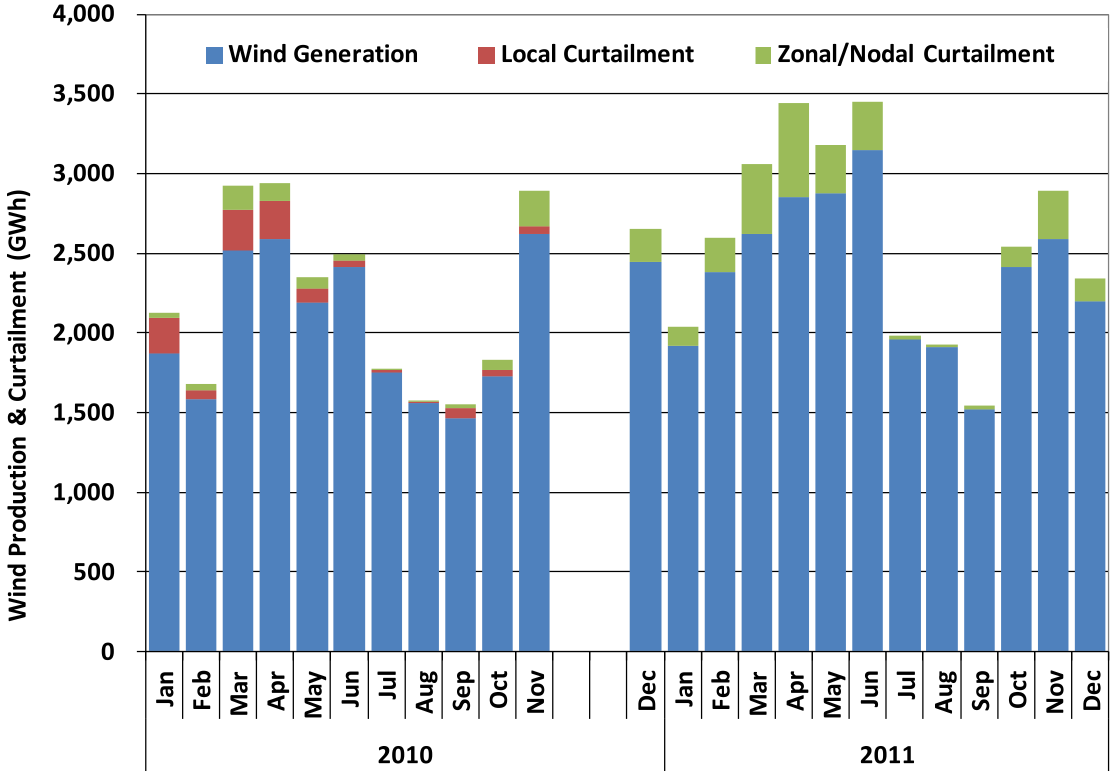
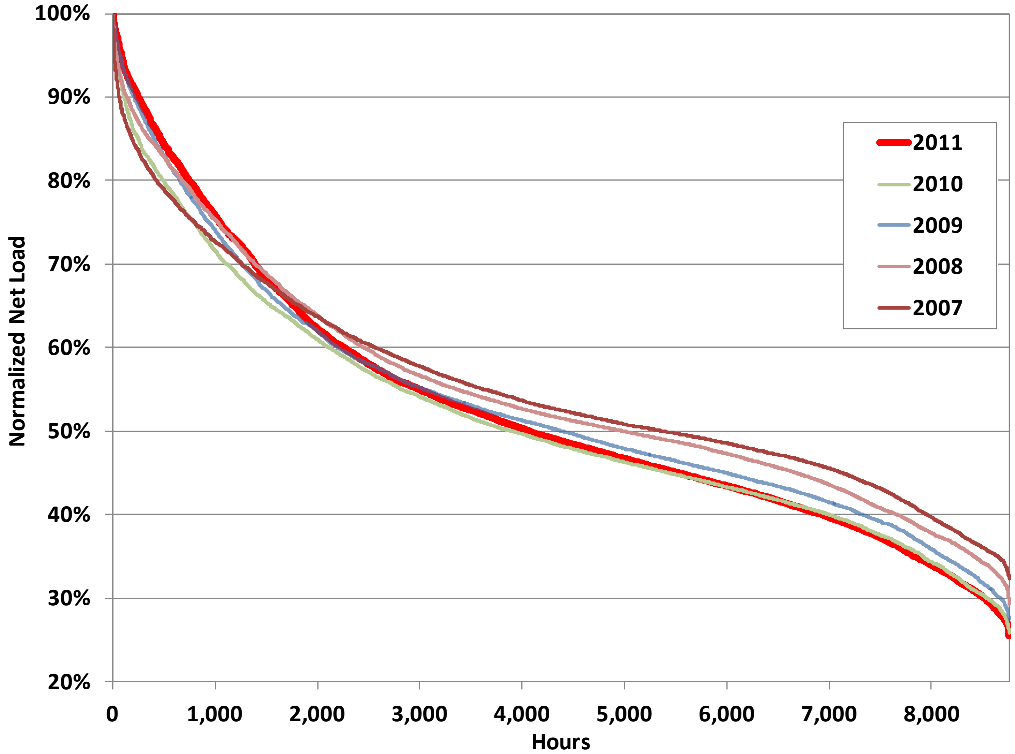


Figure 54 shows the wind production and local and zonal curtailment quantities for each month of 2010 and 2011. This figure reveals that the total quantity of curtailments for wind resources once again increased in 2011 when compared to 2010, even as actual production increased.

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

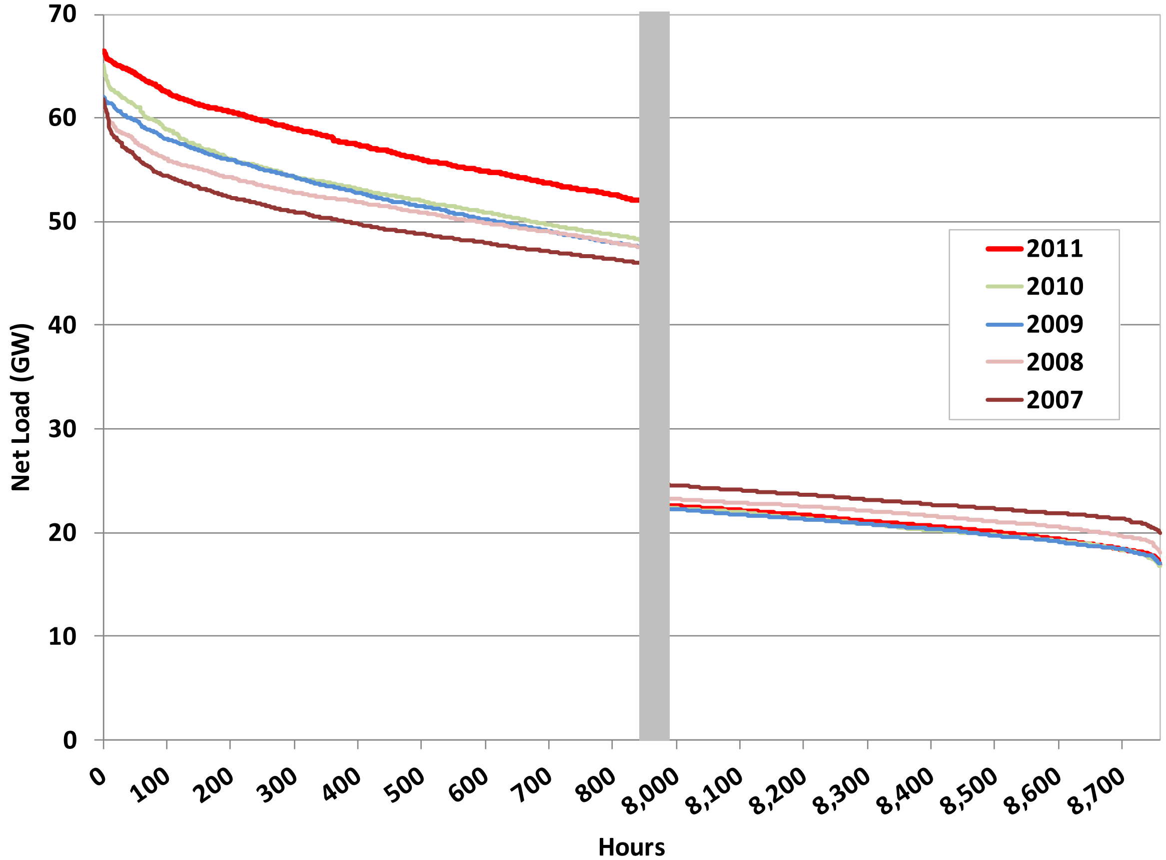


This figure shows the continued erosion of remaining energy available for non-wind units to serve during most hours of the year, with much less impact during the highest loads.

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.

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.

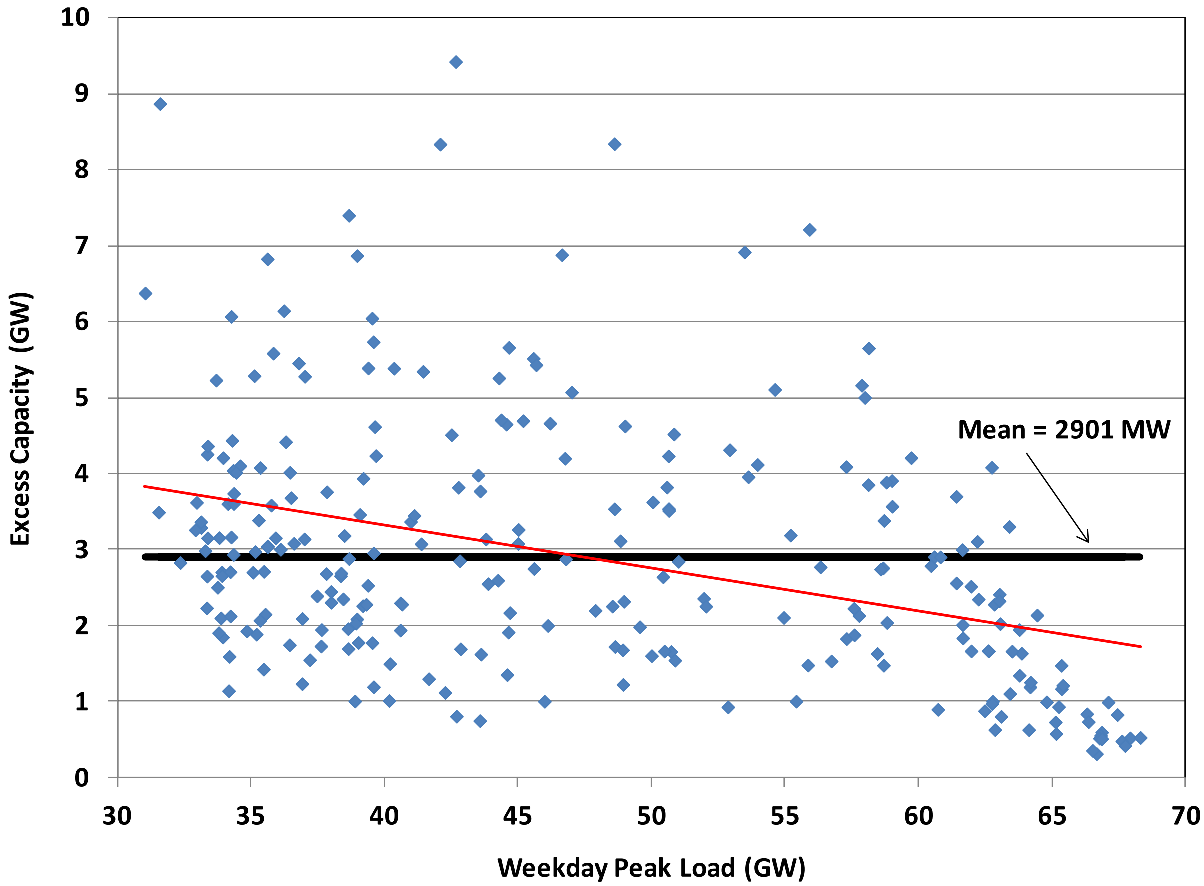
Thus, although the peak net load and reserve margin requirements are projected to continue to increase and create an increasing need for non-wind capacity to meet net load and reliability requirements, the non-wind fleet is expected to operate for fewer hours as wind penetration continues to increase. 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.

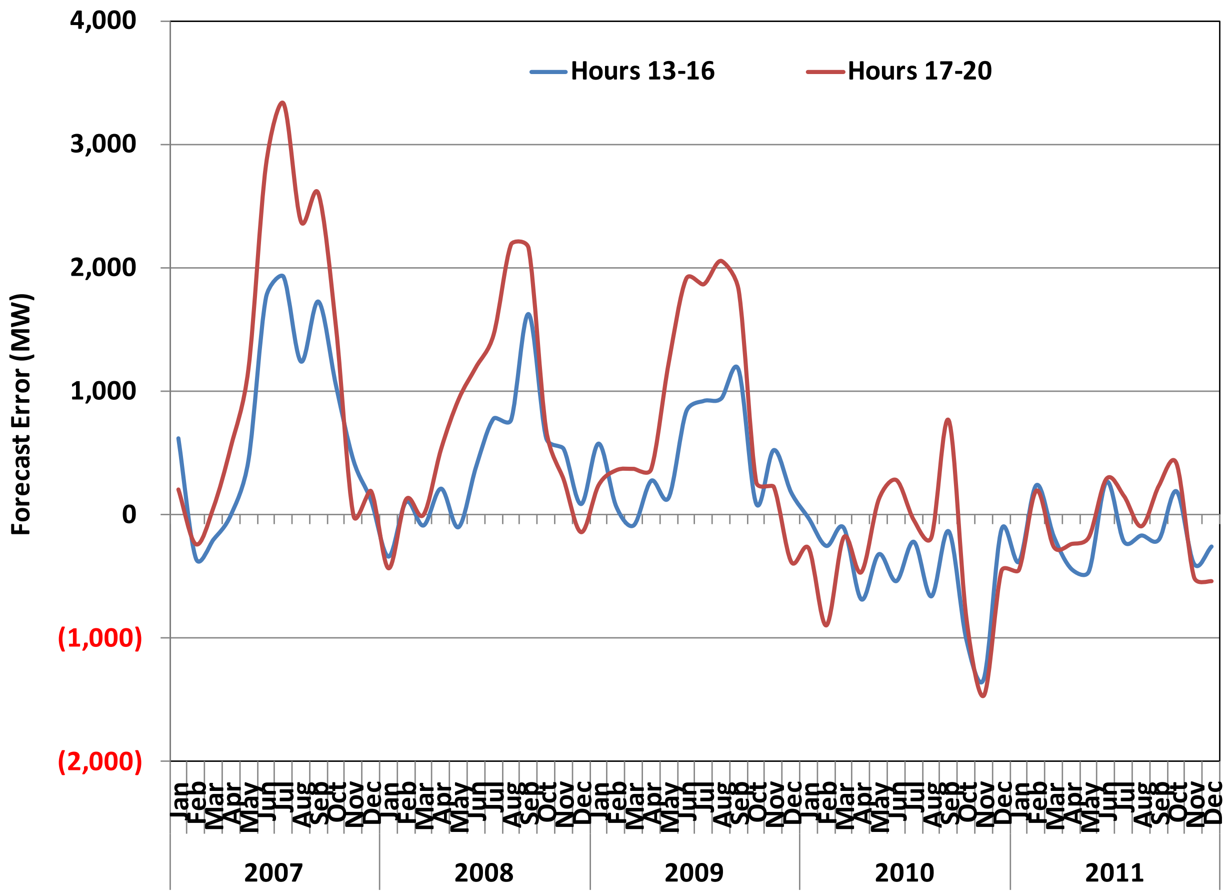
Figure 57: Excess On-Line and Quick Start Capacity



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

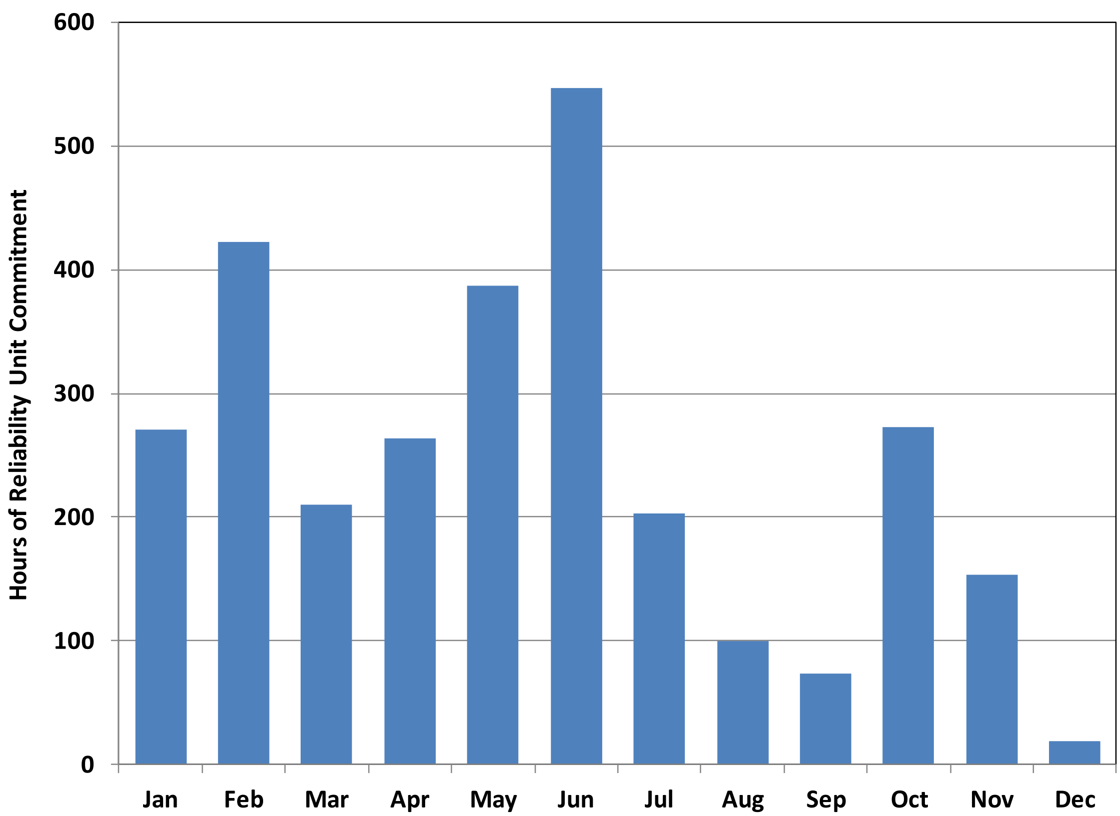


14 See 2010 ERCOT SOM report at 49-51 and 2009 ERCOT SOM report at 68-70.

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.

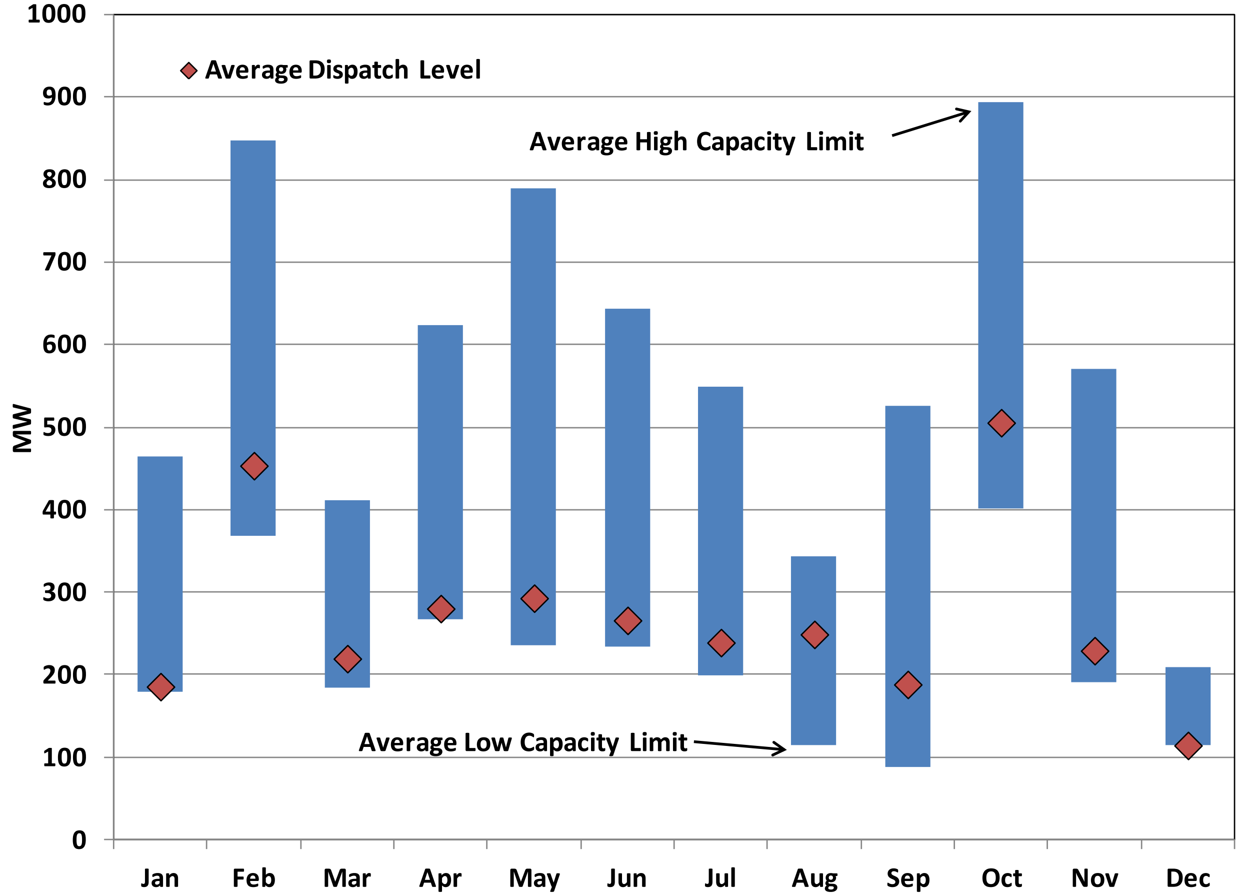
Once ERCOT assesses the unit commitments resulting from the Day-Ahead market, additional capacity commitments are made, if needed, using a reliability unit commitment process that executes both on a day-ahead and hour ahead basis. These additional unit commitments may be made for one of two reasons. 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.

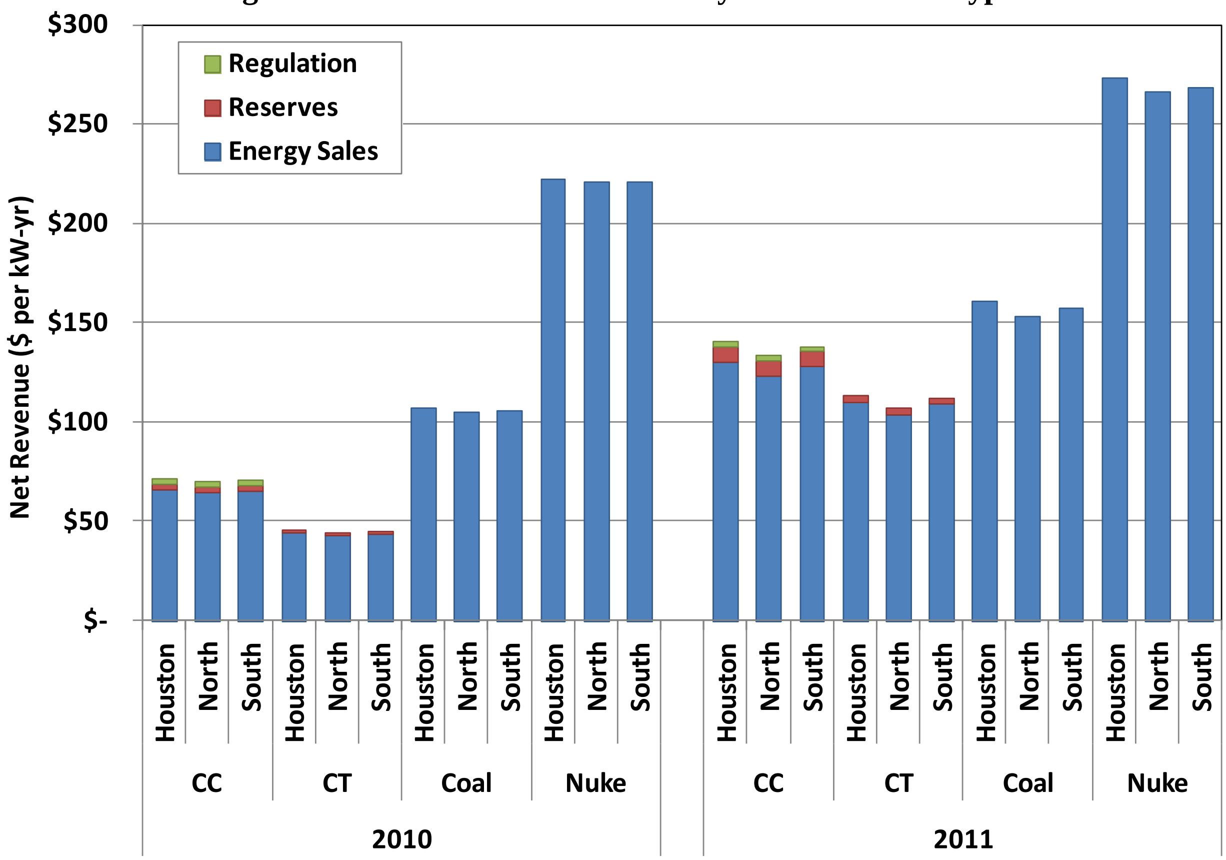
Figure 61 shows the results of the net revenue analysis for four types of hypothetical new units in 2010 and 2011. These are: (a) natural gas fueled combined-cycle, (b) natural gas fueled combustion turbine, (c) coal fueled generator, and (d) a nuclear unit. For the gas-fired technologies, net revenue 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



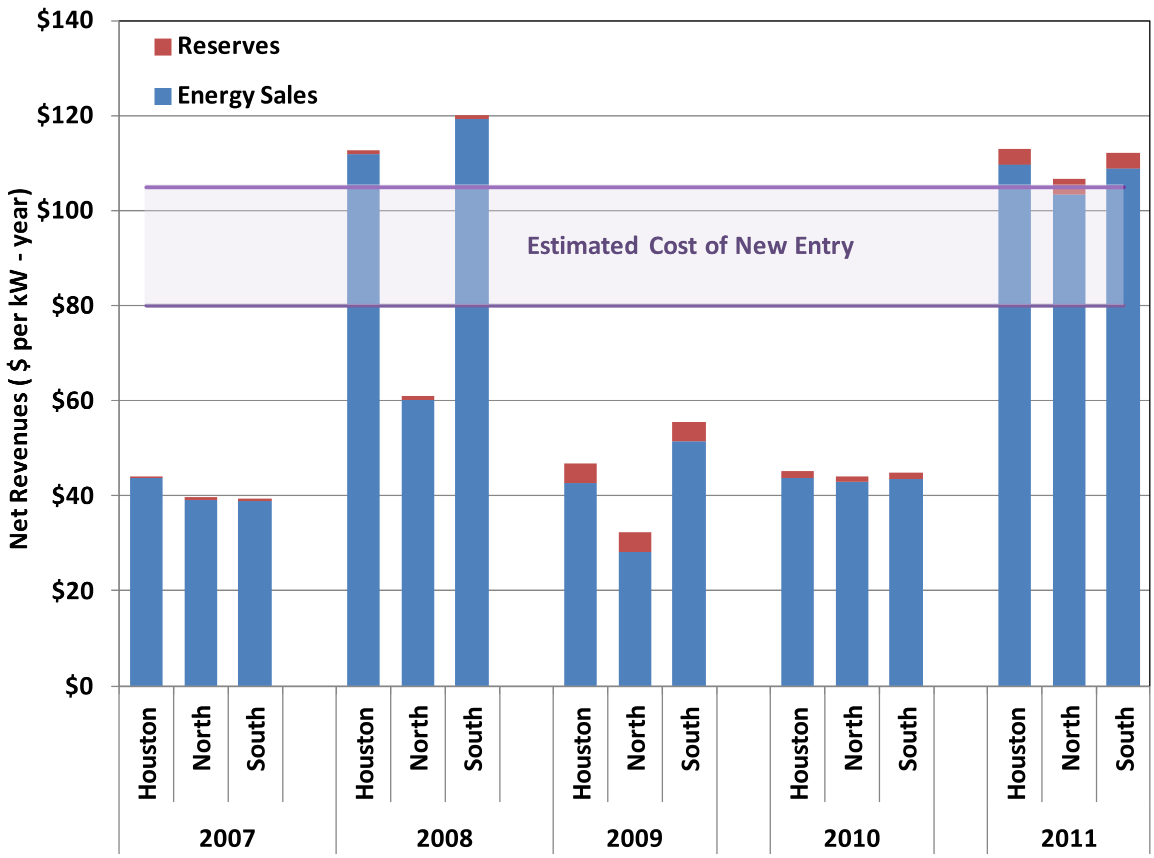
The estimated net revenue for both a new coal or a nuclear unit in ERCOT were well below the levels required to support new entry, despite the relatively frequent shortages in 2011.

- 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.

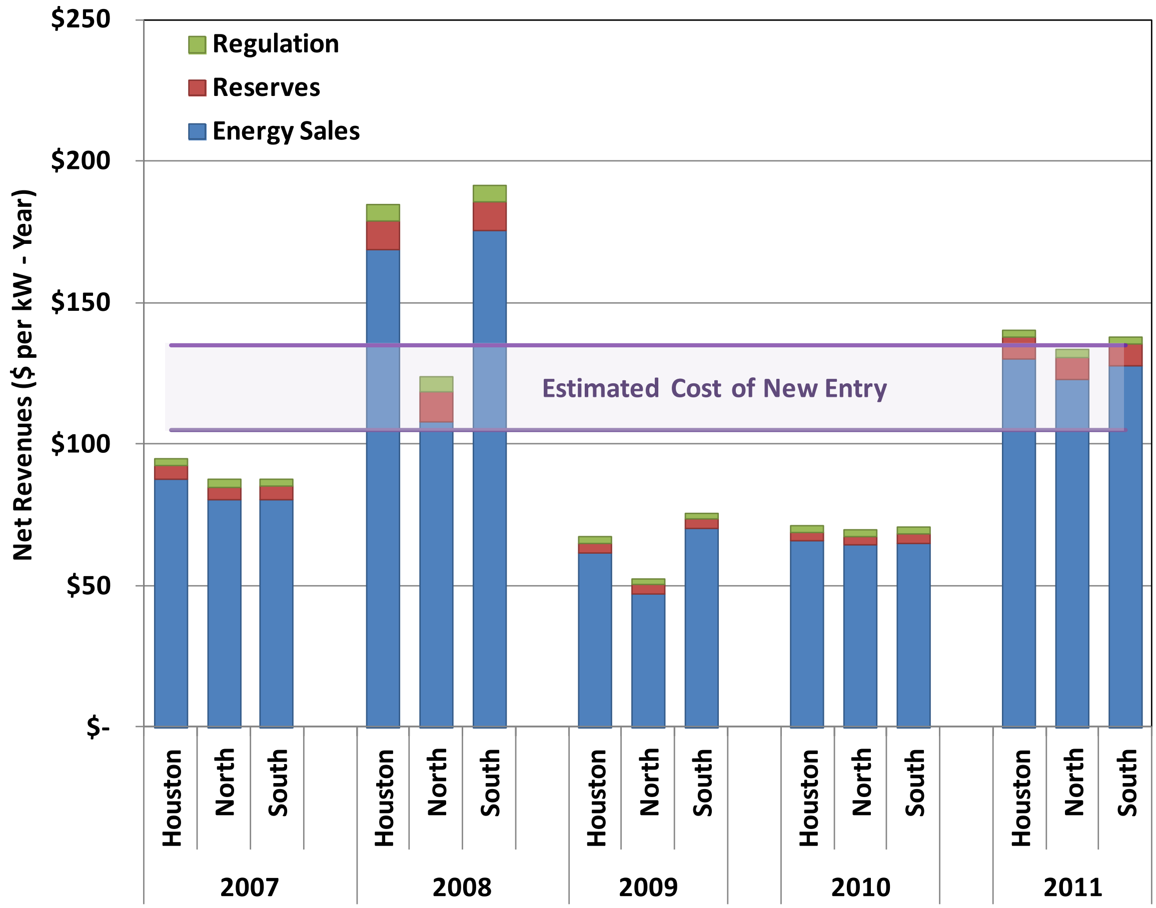
Based on our estimates of investment costs for new units, the net revenue required to satisfy the annual fixed costs (including capital carrying costs) of a new gas turbine unit ranges from $80 to $105 per kW-year. The estimated net revenue in 2011 for a new gas turbine ranged from $107 per kW-year in 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

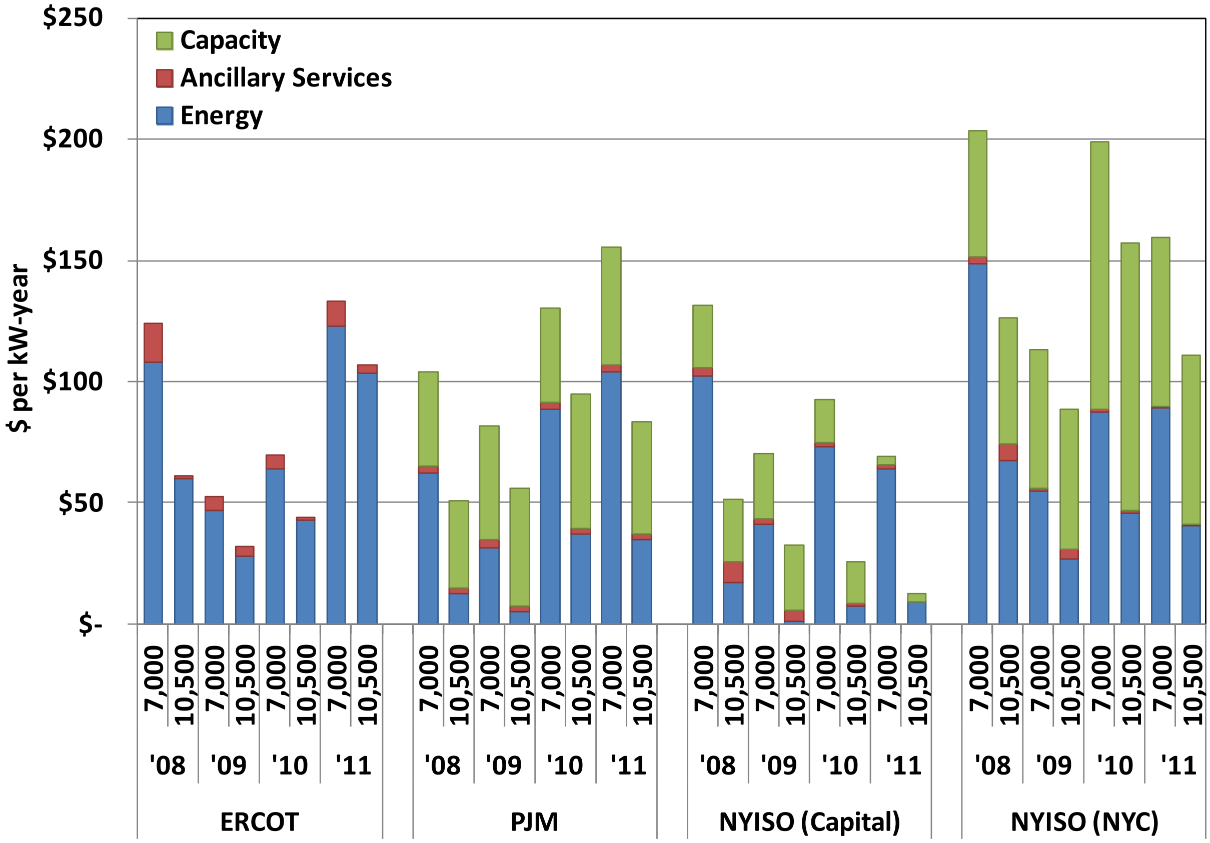


Figure 64 compares estimates of net revenue for the ERCOT North Zone, PJM, and two locations within the New York ISO. The figure includes estimates of net revenue from energy, reserves and regulation, and capacity. ERCOT does not have a capacity market, and thus, does not have any net revenue from capacity 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

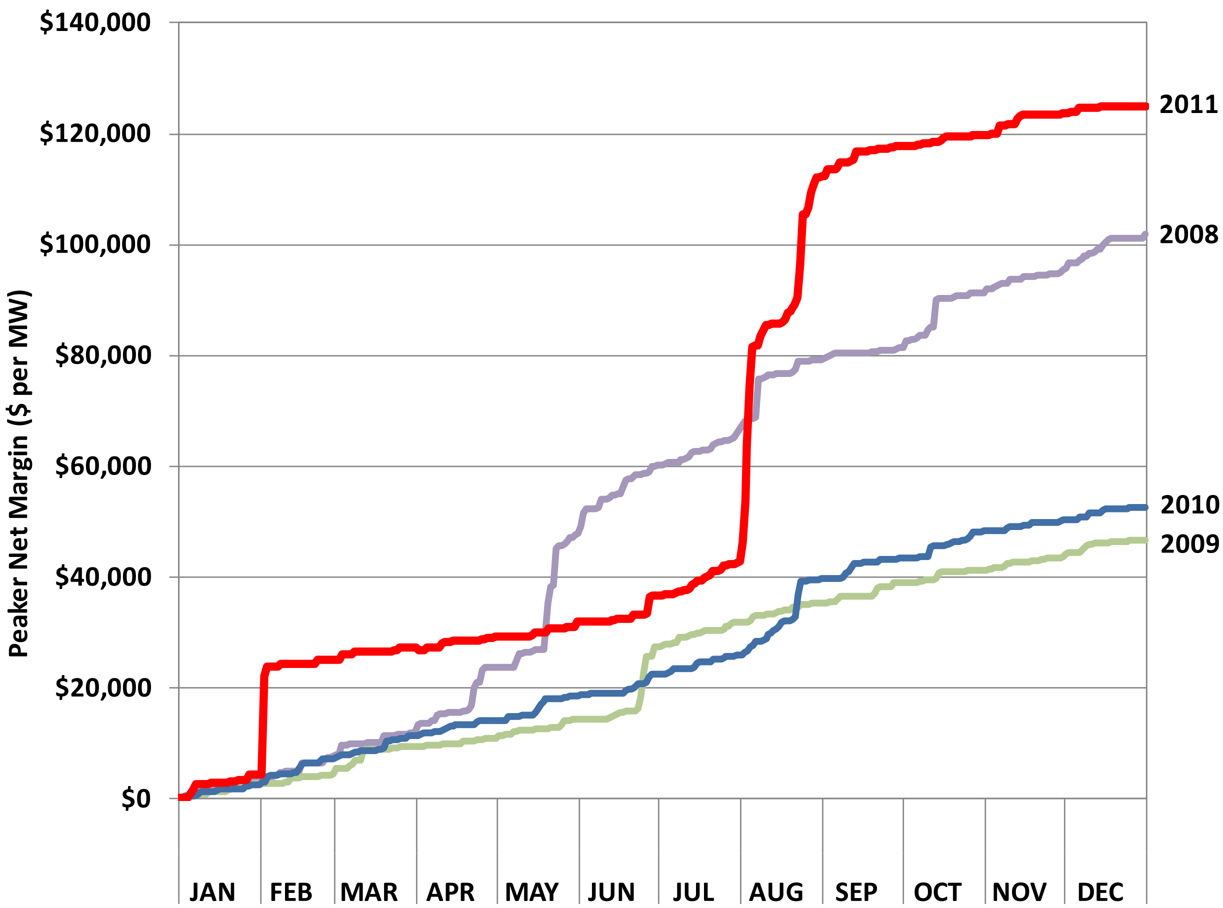
The Public Utility Commission of Texas ("PUCT") adopted rules in 2006 that define the parameters of an energy-only market. These rules include a Scarcity Pricing Mechanism ("SPM") that relaxed the existing system-wide offer cap by increasing it in multiple steps until it reached $3,000 per MWh shortly 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.

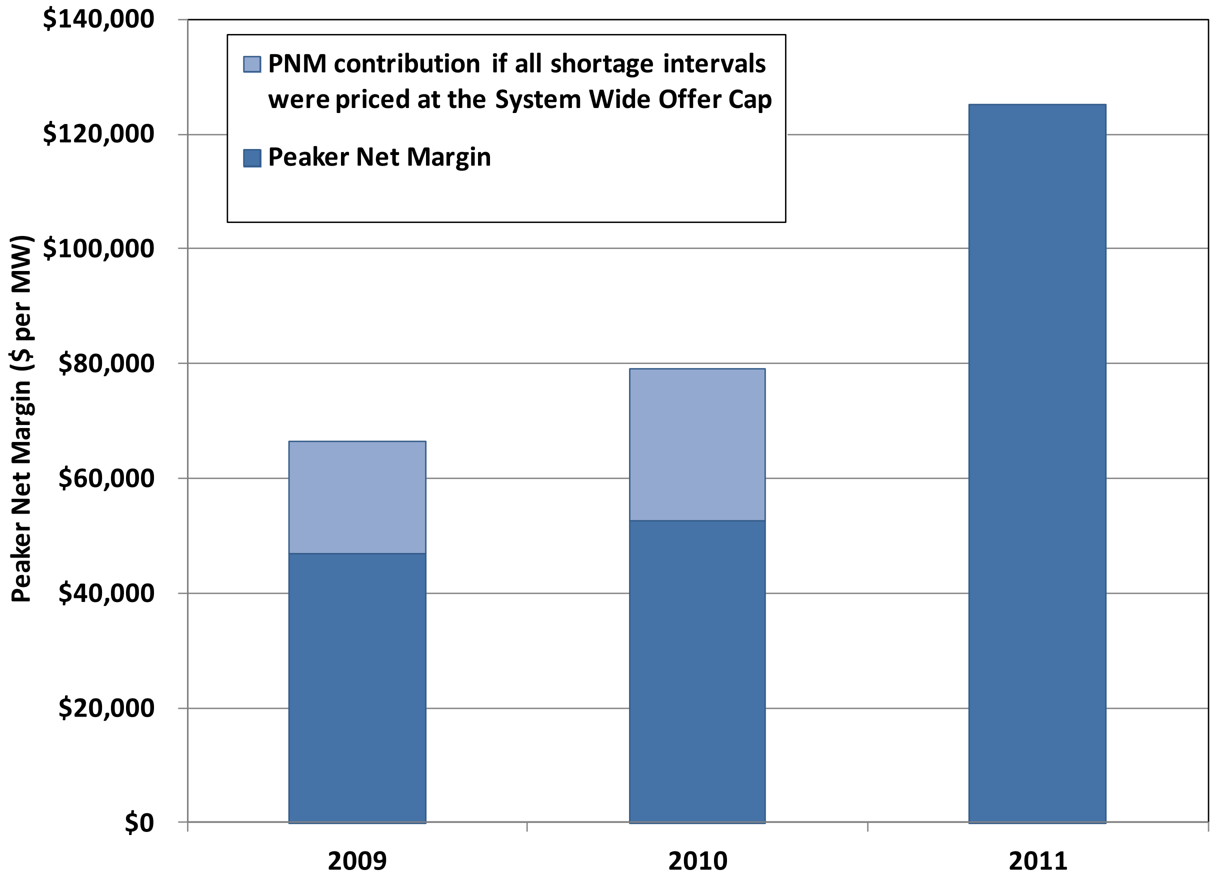
Under the PUCT rules governing the energy-only market, high-priced offers submitted by small market participants are deemed to not be an exercise of ERCOT-wide market power. Under the mechanics of the zonal market design, even during shortage conditions, prices were always set by generator offers. As 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

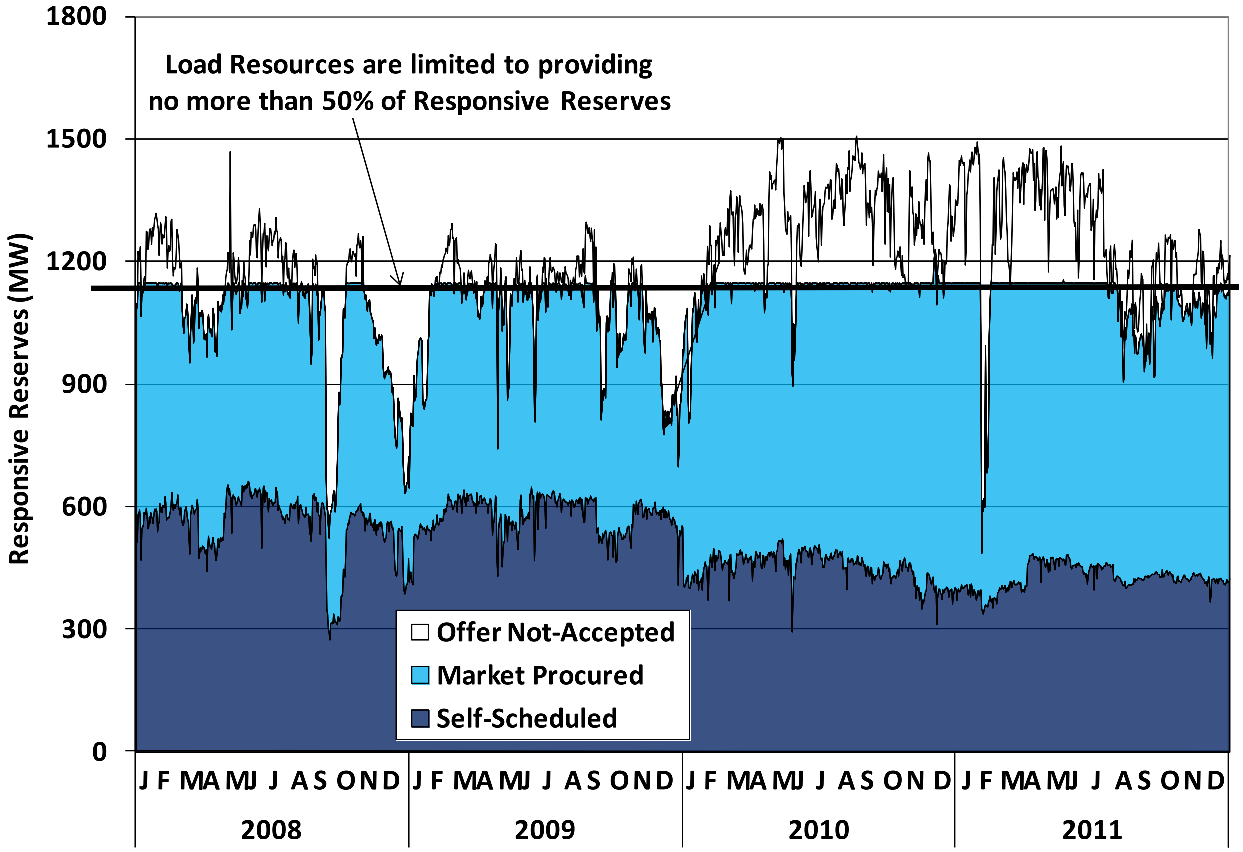
Demand response is a term that broadly refers to actions that can be taken by end users of electricity to reduce load in response to instructions from ERCOT or in response to certain market or system conditions. The ERCOT market allows participants with demand-response capability to provide energy and 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.

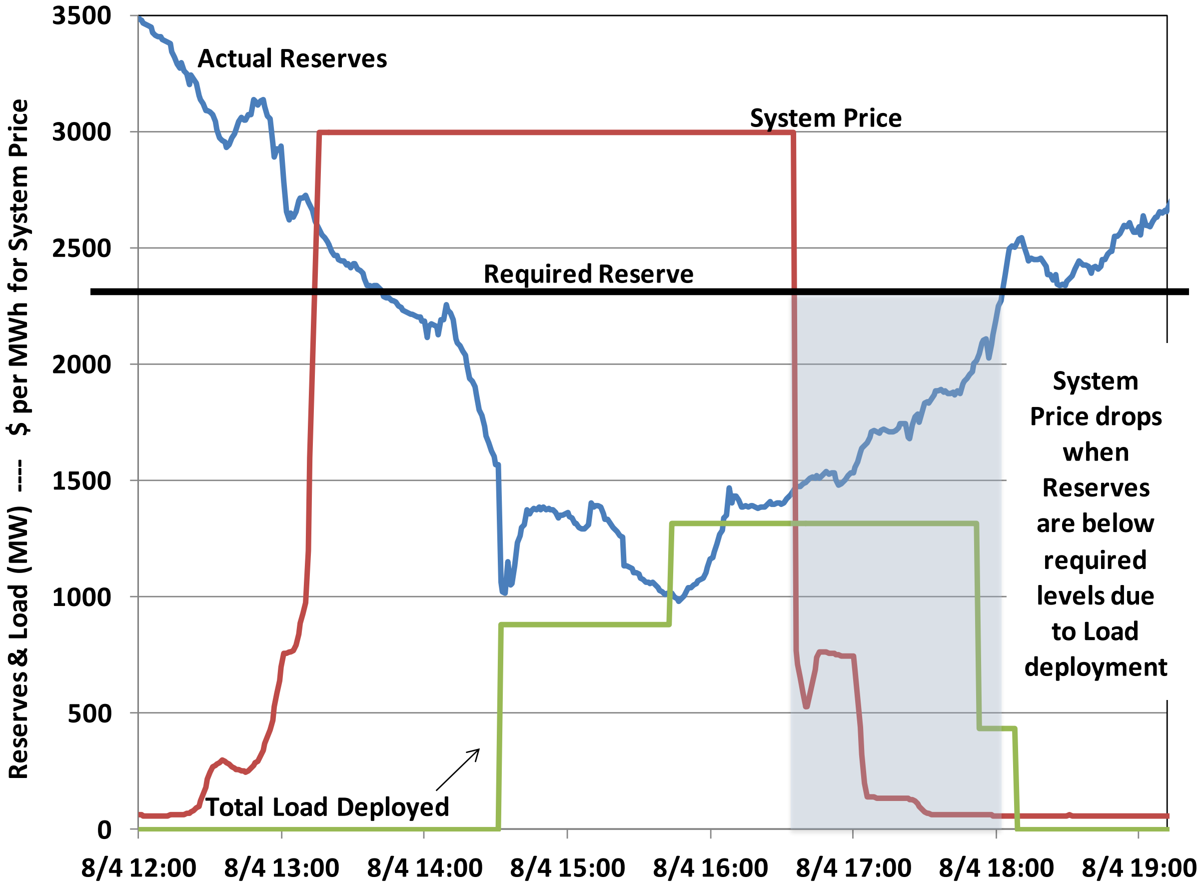
Figure 67: Daily Average of Responsive Reserves provided by Load Resources



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.

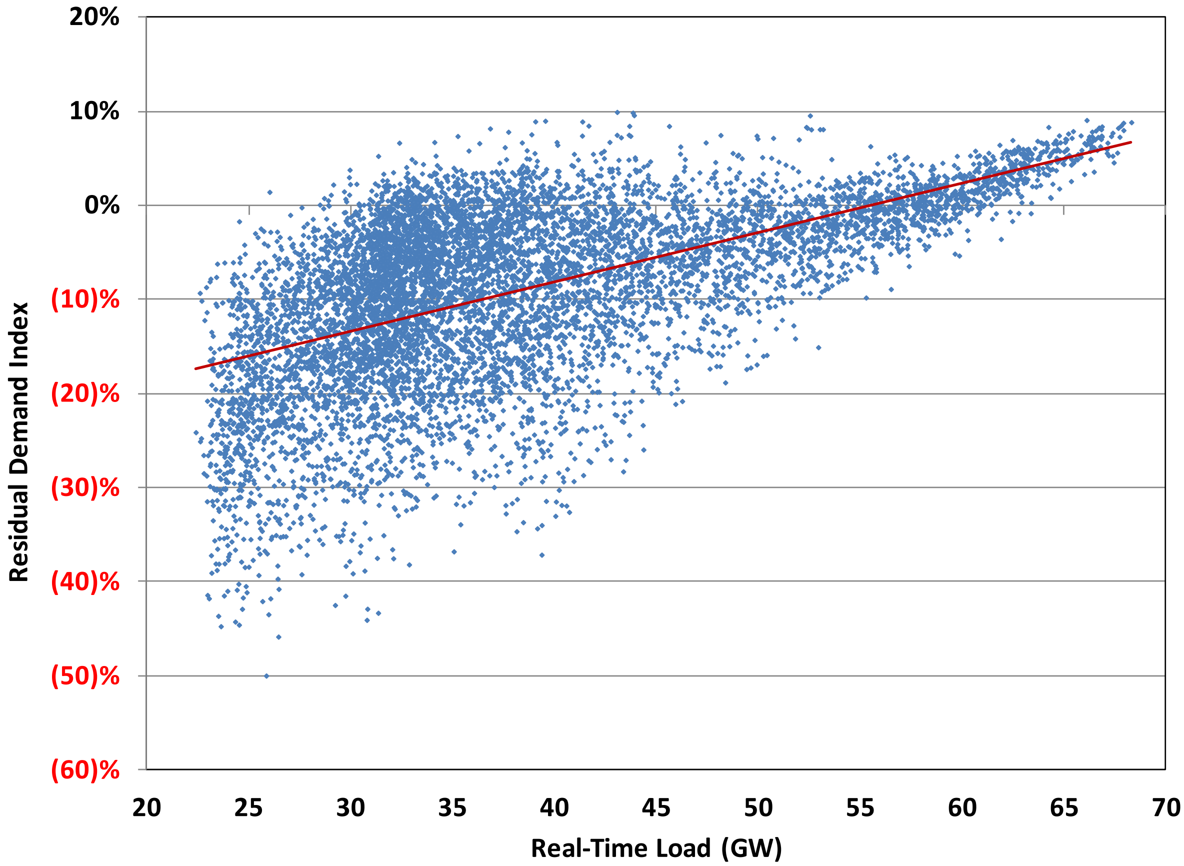
The RDI is a useful structural indicator of potential market power, although it is important to recognize its limitations. As a structural indicator, it does not illuminate actual supplier behavior to indicate whether a supplier may have exercised market power. The RDI also does not indicate whether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

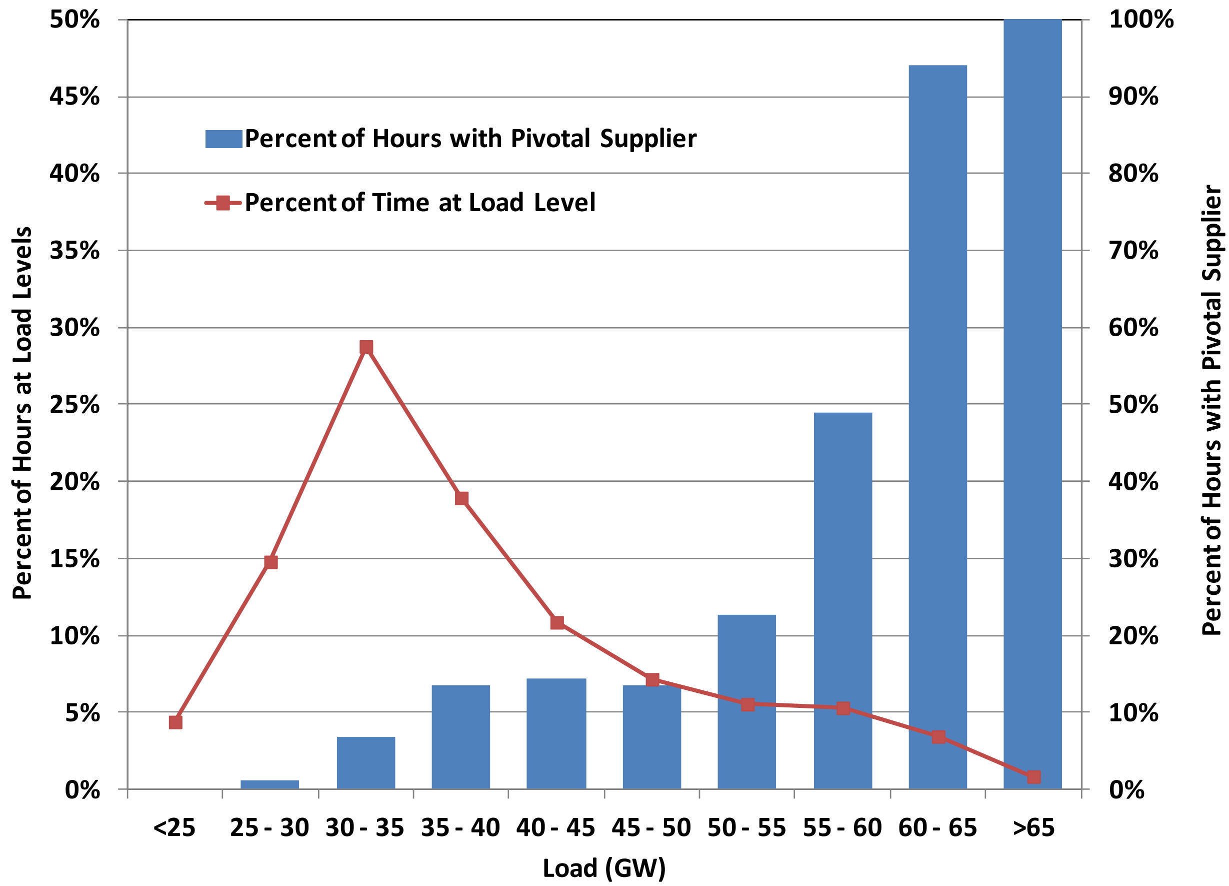
Figure 69 shows the RDI relative to load for all hours in 2011. The trend line indicates a strong positive relationship between load and the RDI. This analysis shown below is done at the QSE level because the largest suppliers that determine the RDI values own a large majority of the resources they are 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.

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

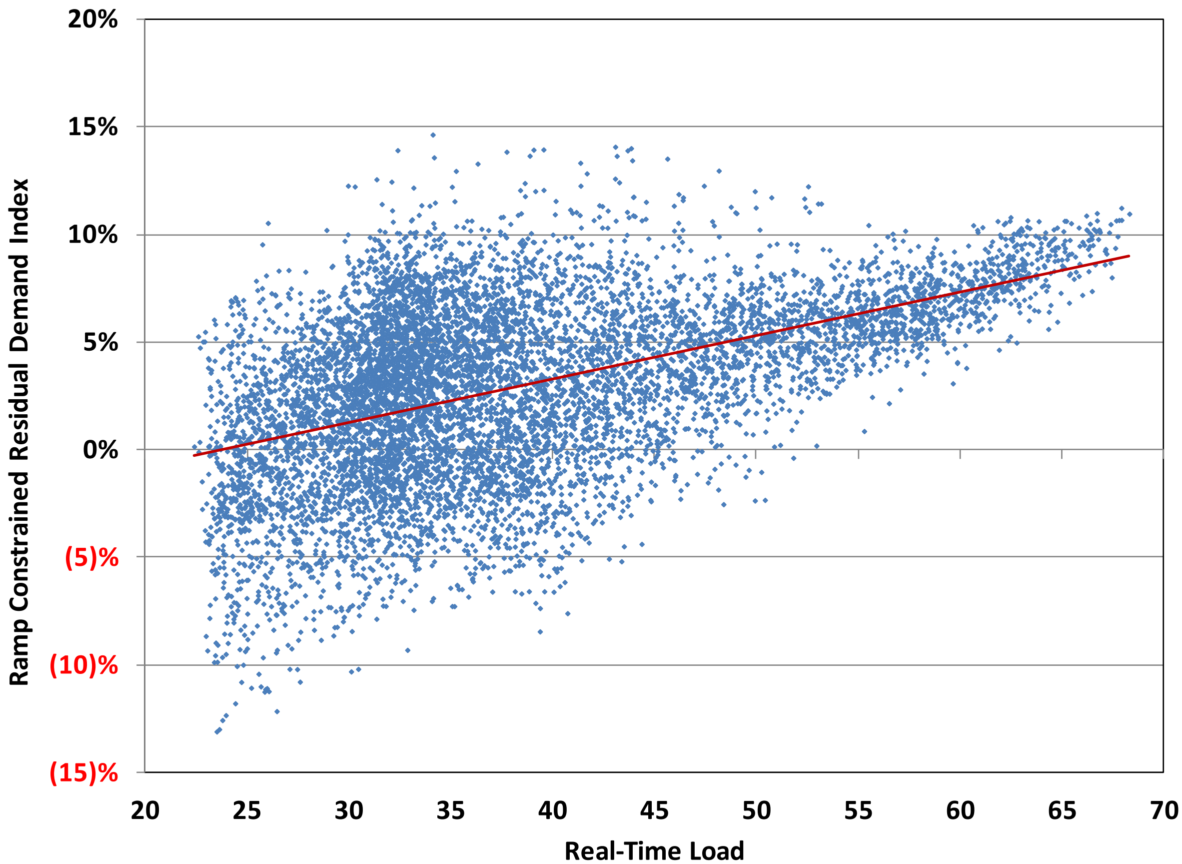
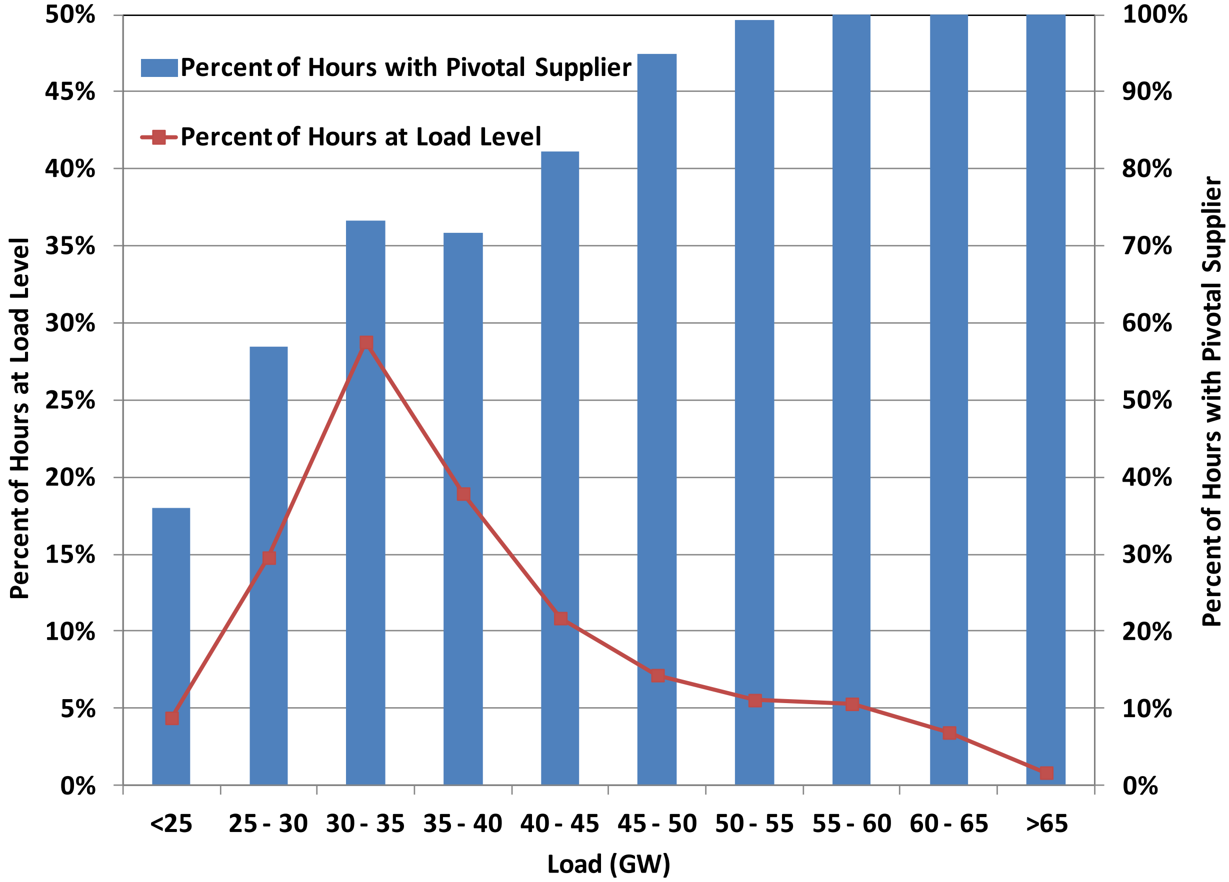


Figure 72 displays the percent of time at each load level there as a pivotal supplier when ramp rate constraints are considered. At loads greater than approximately 50 GW there is a pivotal supplier 100 percent of the time. Ramp rate constrained RDI indicates that there was a pivotal supplier in approximately 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.

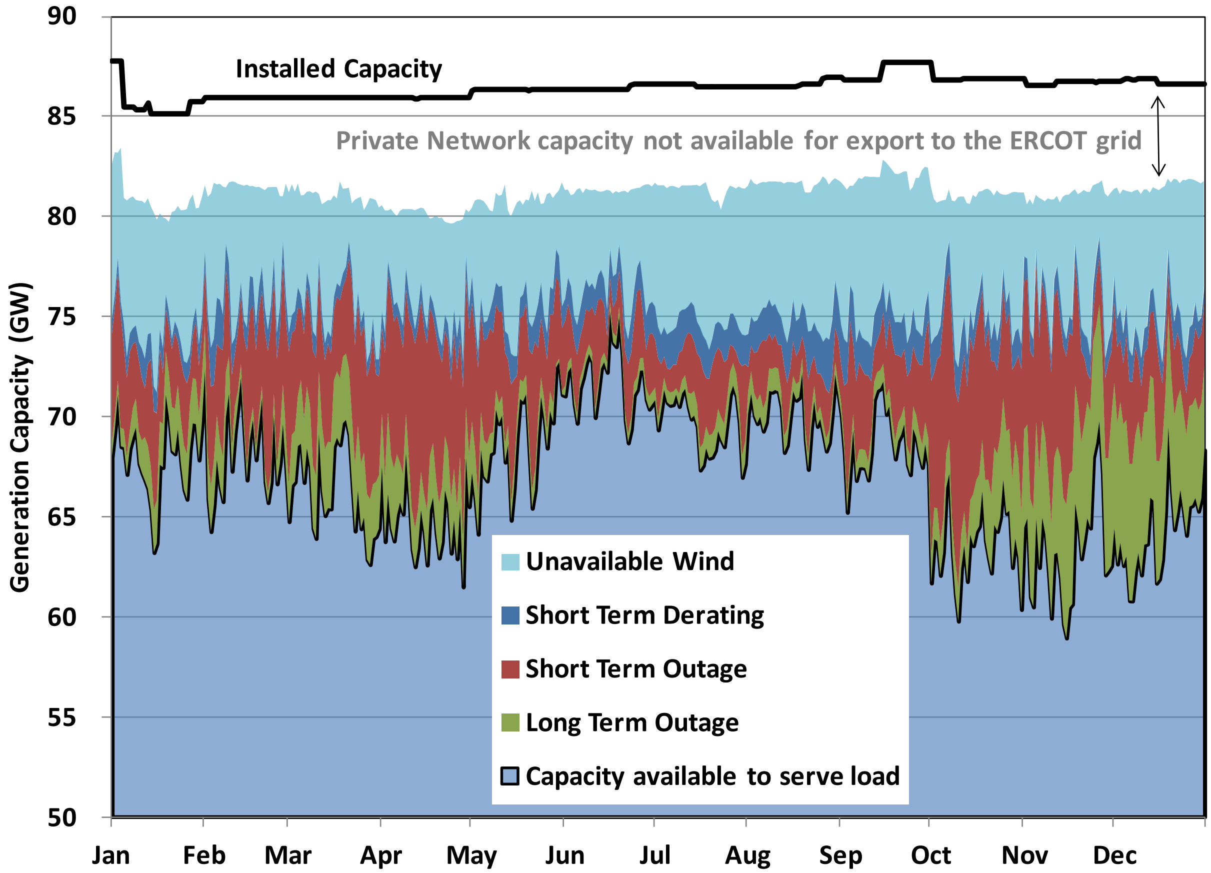
1. Generation Outages and Deratings

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

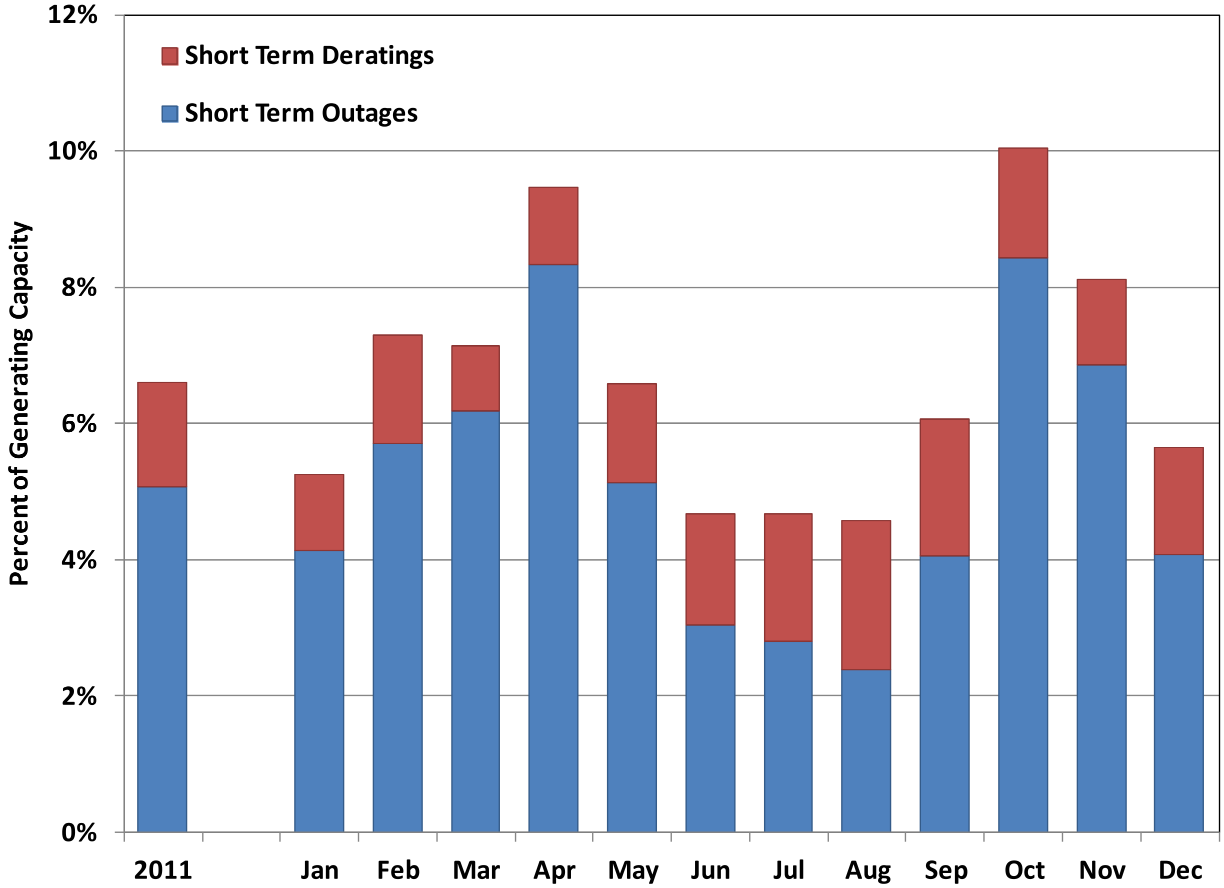


Figure 74 shows that total short-term deratings and outages were as large as 10 percent of installed capacity in the spring and fall, dropping to as low as 5 percent for the summer. Most of this fluctuation was likely due to anticipated planned outages.

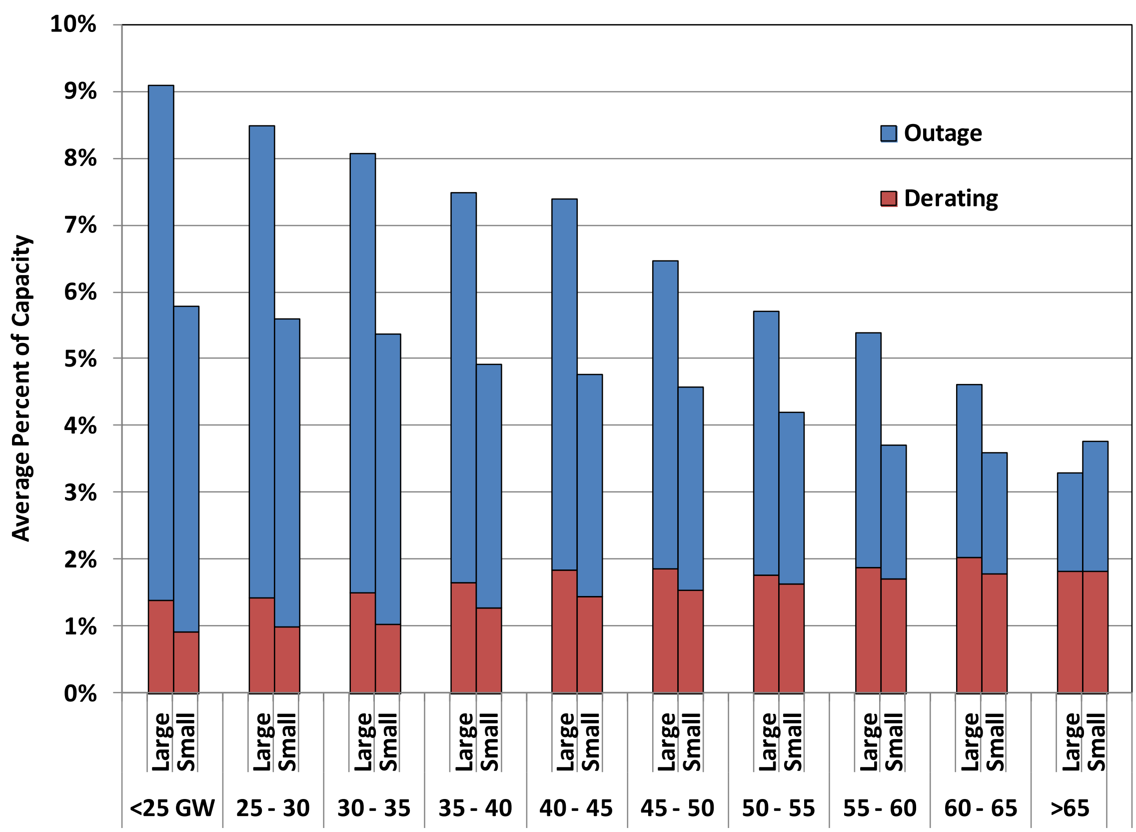
2. Evaluation of Potential Physical Withholding

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

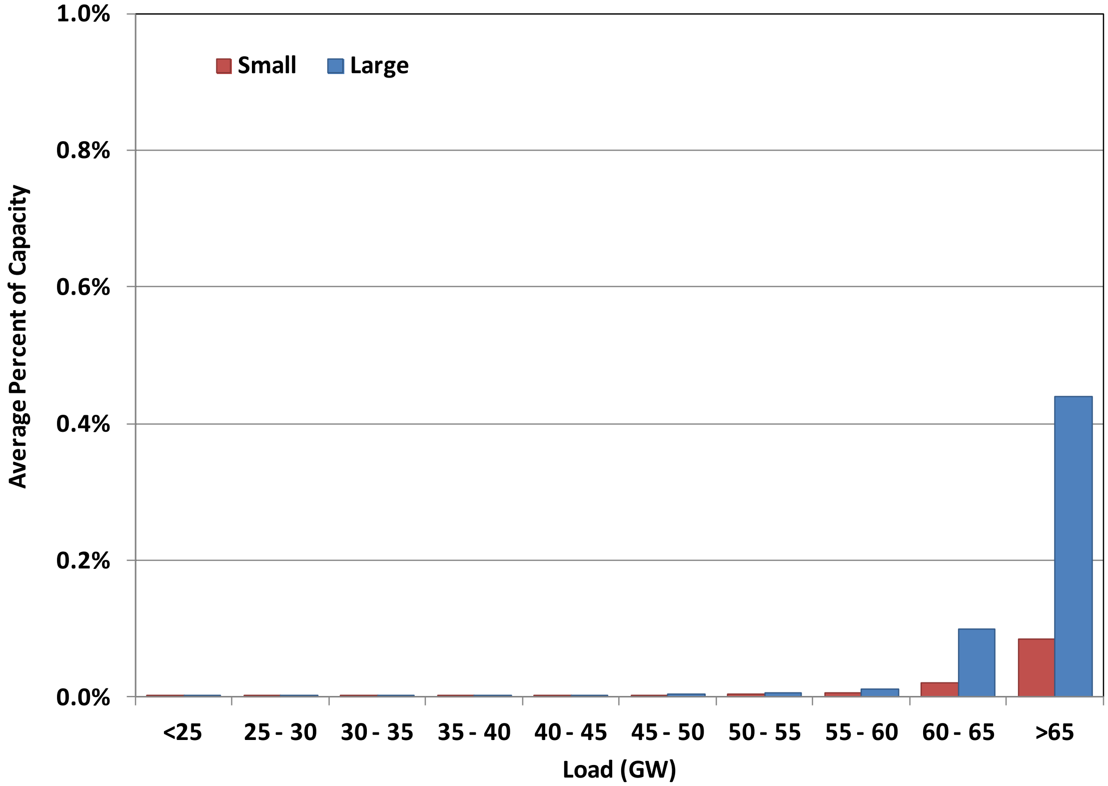
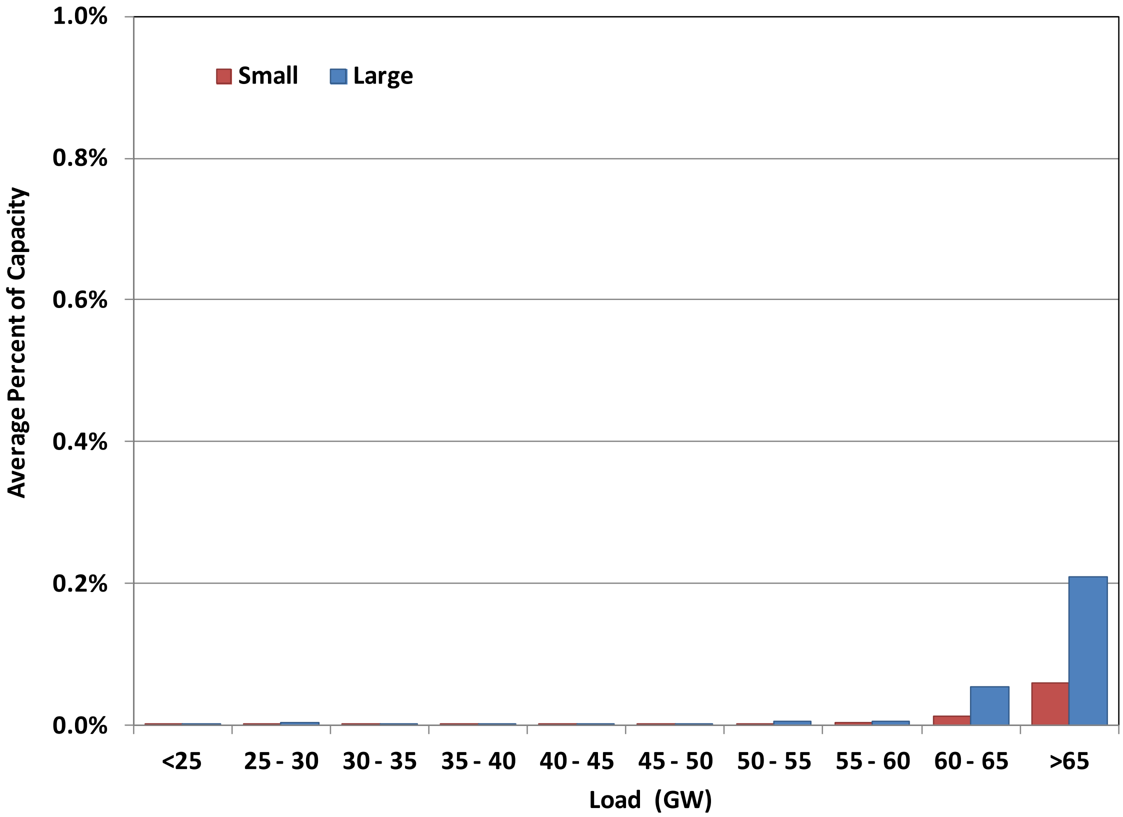


Figure 77 shows the ultimate output gap, measured by the difference between a unit's operating level and the output level had the unit been competitively offered to the market. In the second step of the dispatch, the after-mitigation offer curve is used to determine dispatch instructions and locational 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

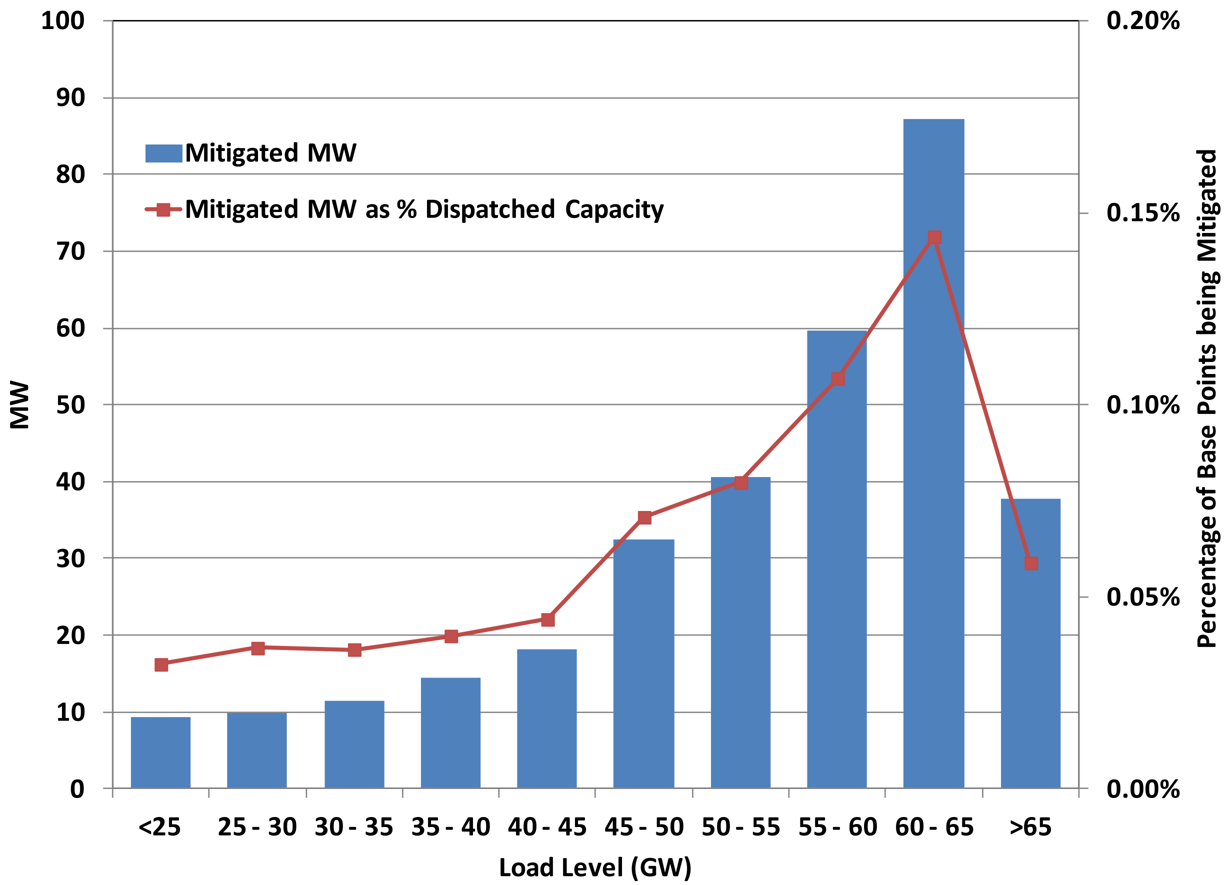


C. Mitigation

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.

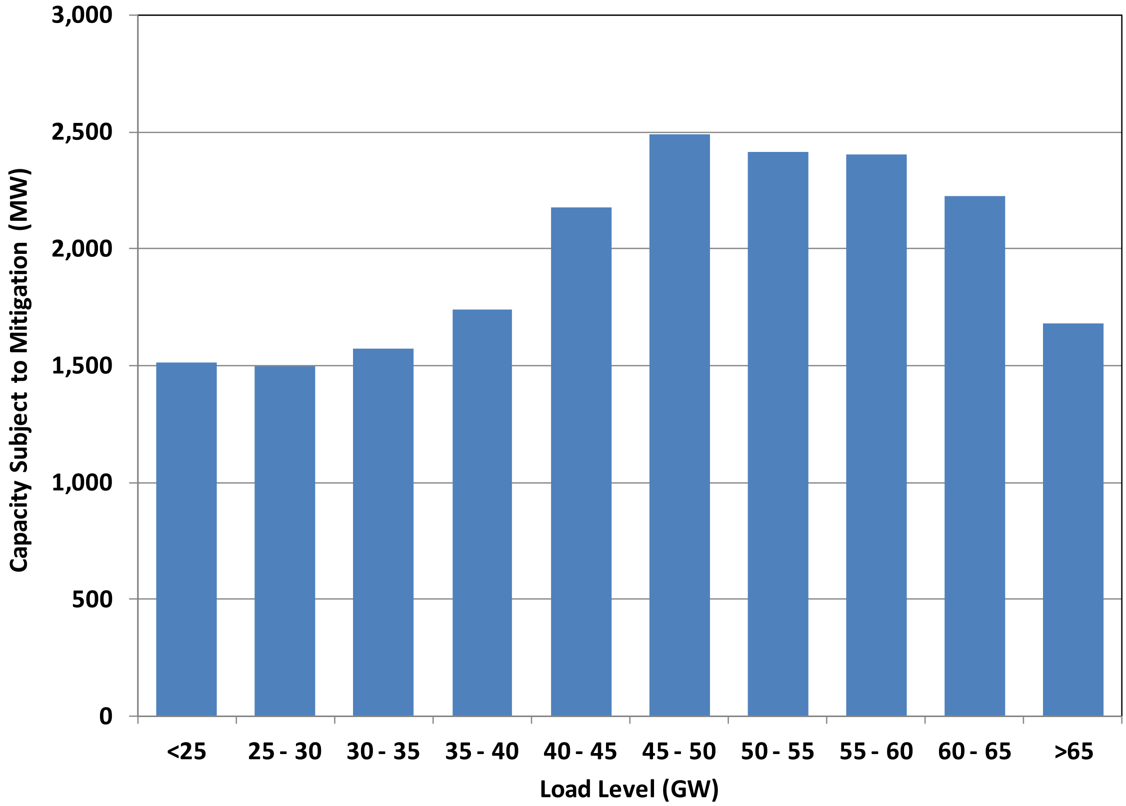
Our first analysis computes how much capacity, on average, is actually mitigated during each dispatch interval. The results, shown in Figure 78, are provided by load level. The quantities of capacity actually mitigated are relatively small, averaging 10 MW at low loads and increasing to almost 90 MW 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.