2016 State of the Market Report For the ERCOT Electricity Markets

POTOMAC ECONOMICS

Independent Market Monitor
for ERCOT

May 2017

TABLE OF CONTENTS

LIST OF FIGURES

<u>LIST OF TABLES</u>

EXECUTIVE SUMMARY

This report reviews and evaluates the outcomes of the ERCOT wholesale electricity markets in 2016 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 (PURA). It includes assessments of the incentives provided by the current market rules and analyses of the conduct of market participants. This report also assesses the effectiveness of the Scarcity Pricing Mechanism (SPM) pursuant to the provisions of 16 TEX. ADMIN. CODE § 25.505(g).

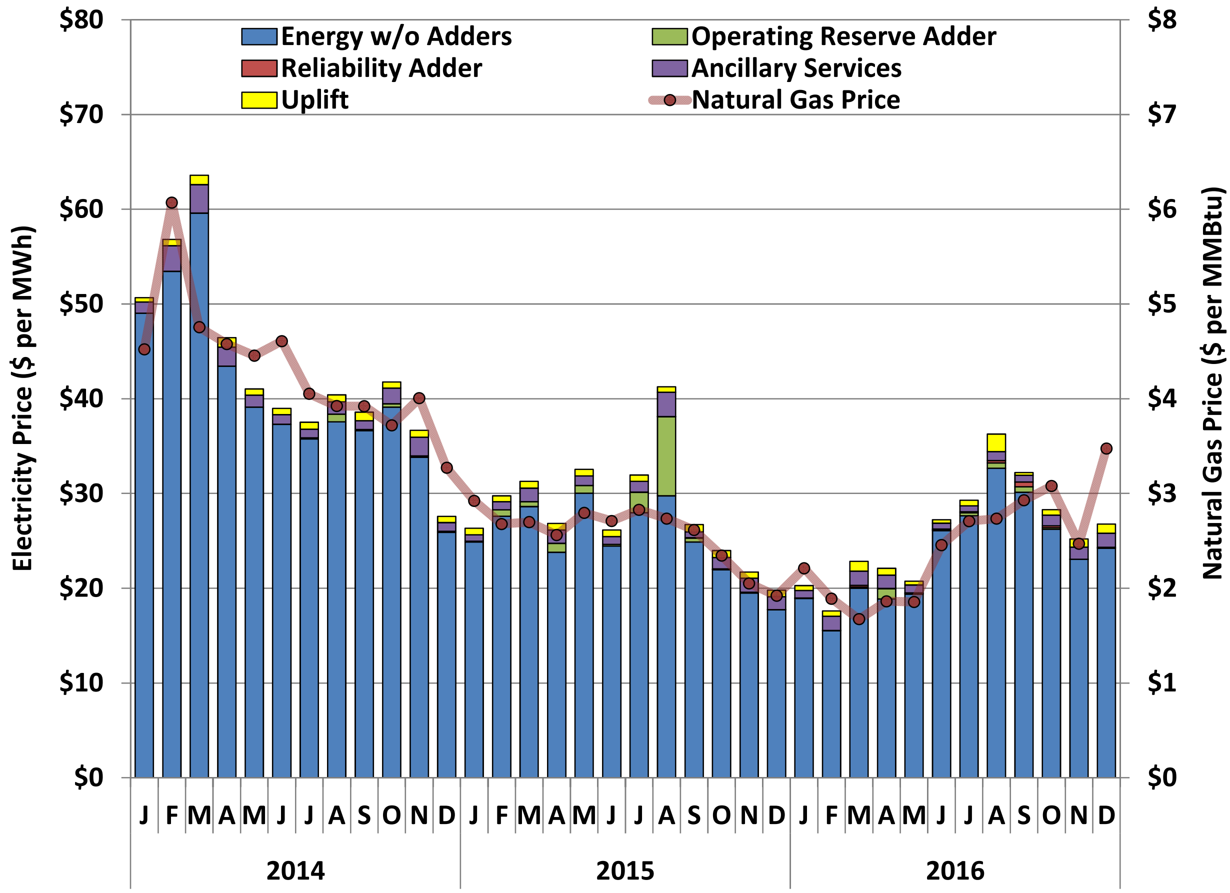
Overall, the ERCOT wholesale market performed competitively in 2016. Our key findings and results from 2016 include the following:

- Lower natural gas prices and surplus supply led to lower energy prices in 2016:
  - The ERCOT-wide load-weighted average real-time energy price was $24.62 per MWh in 2016, an 8 percent decrease from 2015.
  - The average price for natural gas was 4.7 percent lower in 2016 than in 2015, decreasing from $2.57 per MMBtu in 2015 to $2.45 per MMBtu in 2016.
- Real-time prices did not exceed $3,000 per MWh in 2016 and exceeded $1,000 per MWh for only 3.9 hours cumulatively for the year.
- ERCOT-wide real-time prices were negative for approximately 130 hours in 2016, a significant increase from the approximately 50 hours with negative prices in 2015.
- ERCOT set a new hourly demand record of 71,110 MW on August 11, 2016, an increase of 1.8 percent from the previous peak set in 2015. Average demand also rose in 2016, increasing 0.7 percent from 2015.
- The total congestion costs experienced in the ERCOT real-time market in 2016 were $497 million, an increase of 40 percent from 2015. Transmission outages were the primary causes for this increase.
- Net revenues provided by the market during 2016 were less than the estimated amount necessary to support new greenfield generation investment, which is not a surprise given that planning reserves are above the minimum target and shortages were rare in 2016. The Operating Reserve Demand Curve (ORDC), combined with a relatively high offer cap should increase net revenues when shortages become more frequent.

Review of Real-Time Market Outcomes

Although only a small share of the power produced in ERCOT is transacted in the spot market, real-time energy prices are very important because they set the expectations for prices in the day-ahead market and other forward markets where most transactions occur. 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 figure below summarizes changes in energy prices and other market costs by showing the all-in price of electricity, which is a measure of the total cost of serving load in ERCOT.

Average All-in Price for Electricity in ERCOT



The ERCOT-wide price in this figure is the load-weighted average of the real-time market prices from all load zones. Ancillary services costs and uplift costs are divided by real-time load to show them on a per MWh basis.[1] ERCOT developed two energy price adders that are designed to improve its real-time energy pricing when reserves become scarce or ERCOT takes out-of-

1 For this analysis uplift includes: Reliability Unit Commitment Settlement, Operating Reserve Demand Curve (ORDC) Settlement, Revenue Neutrality Total, Emergency Energy Charges, Base Point Deviation Payments, Emergency Response Service (ERS) Settlement, Black Start Service Settlement, and Block Load Transfer Settlement.

market actions for reliability. To distinguish the effects of the energy price adders, the Operating Reserve Demand Curve Adder (operating reserve adder) and the Reliability Deployment Price Adder (reliability adder) are shown separate from the energy price. The operating reserve adder was implemented in mid-2014 to account for the value of reserves based on the probability of reserves falling below the minimum contingency level and the value of lost load. The reliability adder was implemented in June 2015 as a mechanism to ensure that reliability deployments do not distort the energy prices.

The largest component of the all-in price is the energy cost. This figure above indicates that natural gas prices continued to be a primary driver of electricity prices. This correlation is expected in a well-functioning, competitive market because fuel costs represent the majority of most suppliers' marginal production costs. Since suppliers in a competitive market have an incentive to offer supply at marginal costs and natural gas is the most widely-used fuel in ERCOT, changes in natural gas prices should translate to comparable changes in offer prices. Hence, the reduction in natural gas prices of almost 5 percent contributed to an 8 percent reduction in ERCOT's average real-time energy prices. The all-in price in 2016 included small contributions from ERCOT's energy price adders – $0.27 per MWh from the operating reserve adder and $0.13 per MWh from the reliability adder.

Finally, the other classes of costs continue to be a small portion of the all-in electricity price – ancillary services costs were $1.03 per MWh, down from $1.23 per MWh in 2015 because of reductions in natural gas prices and lower ancillary service requirements. Uplift costs accounted for $0.74 per MWh of the all-in electricity price, similar to the uplift costs of $0.69 per MWh in 2015.

Real-Time Energy Prices

Energy prices vary across the ERCOT market because of congestion costs that are incurred as power is delivered over the network. The table below provides the annual load-weighted average price for each zone for the past six years.

Average Annual Real-Time Energy Market Prices by Zone

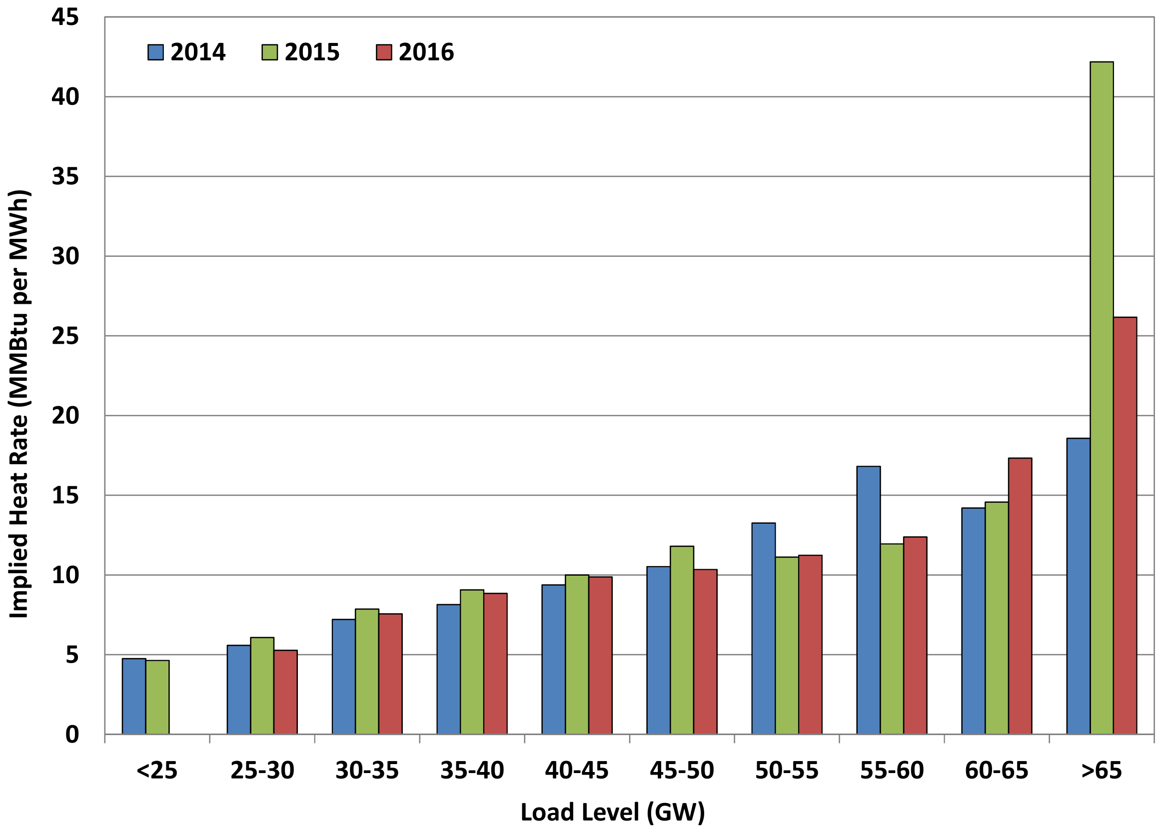
	2011	2012	2013	2014	2015	2016
ERCOT	**$53.23**	**$28.33**	**$33.71**	**$40.64**	**$26.77**	**$24.62**
Houston	$52.40	$27.04	$33.63	$39.60	$26.91	$26.33
North	$54.24	$27.57	$32.74	$40.05	$26.36	$23.84
South	$54.32	$27.86	$33.88	$41.52	$27.18	$24.78
West	$46.87	$38.24	$37.99	$43.58	$26.83	$22.05
Natural Gas ($/MMBtu)	$3.94	$2.71	$3.70	$4.32	$2.57	$2.45

The zonal prices in 2016 show greater disparities than 2015 because of congestion in the West and Houston. Prior to 2012, average prices in the West zone were lower than ERCOT-wide average prices. This changed in 2012 when demand in the West rose because of increased oil and gas production activity. The West zone average annual price remained higher than the ERCOT average until 2016 when increased congestion caused by high levels of wind output in the West caused the average prices in the West to be lower than the other zones. Additionally, transmission congestion related to power flows into Houston caused that zone to exhibit the highest average prices and reduced the average prices in the North zone.

Non-Fuel Energy Price Changes

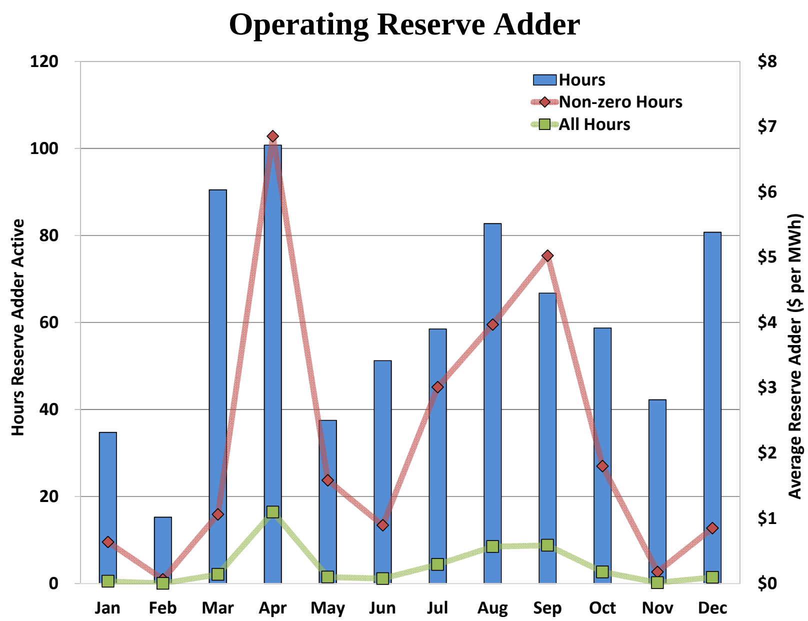
To summarize the changes in energy prices that were related to other factors, an "implied heat rate" is calculated by dividing the real-time energy price by the natural gas price. The following figure shows the average implied heat rate at various system load levels from 2014 through 2016. In a well-performing market, a clear positive relationship between these two variables is expected since resources with higher marginal costs are dispatched to serve higher loads.

Implied Heat Rate and Load Relationship



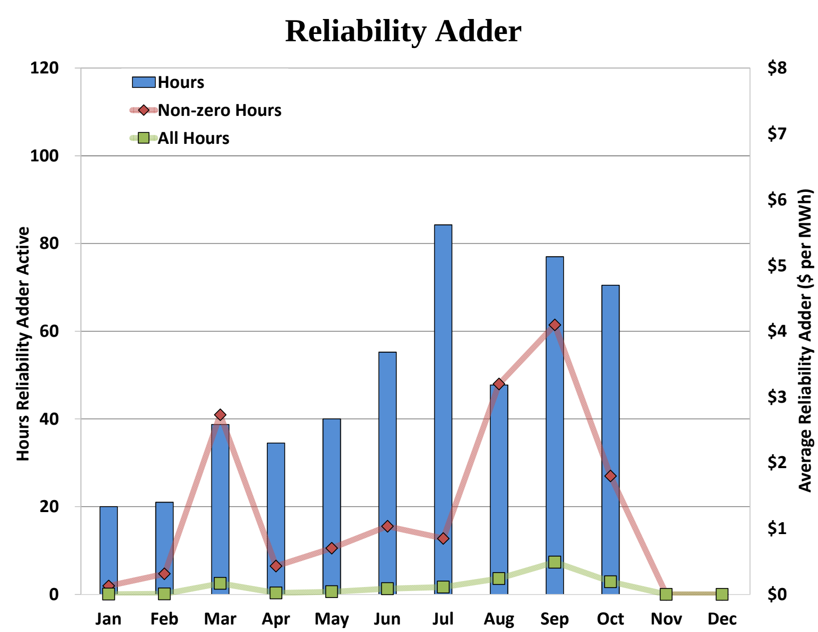
Energy Price Adders

As described above, the contributions of the energy price adders were relatively small in 2016. The first of the two adders is the operating reserve adder, which is based on the loss of load probability, considering online and offline reserve levels, multiplied by the deemed value of lost load. The following figure shows that the operating reserve adder had the largest impacts during April and September, rather than during the summer months as observed in 2015.



Overall, the operating reserve adder contributed $0.27 per MWh or 1 percent to the annual average real-time energy price.

The next figure shows the impacts of the reliability adder. The reliability adder reflects the incremental costs of reliability actions taken by ERCOT, including Reliability Unit Commitments (RUC) and deployed load capacity. When averaged across the active hours, the largest price impacts of the reliability adder occurred in August and September. The reliability adder is zero in most hours. There were no reliability adders in November and December. The reliability adder was non-zero for only 407 hours or 5 percent of the hours in 2016. The contribution from the reliability adder to the annual average real-time energy price was $0.13 per MWh.

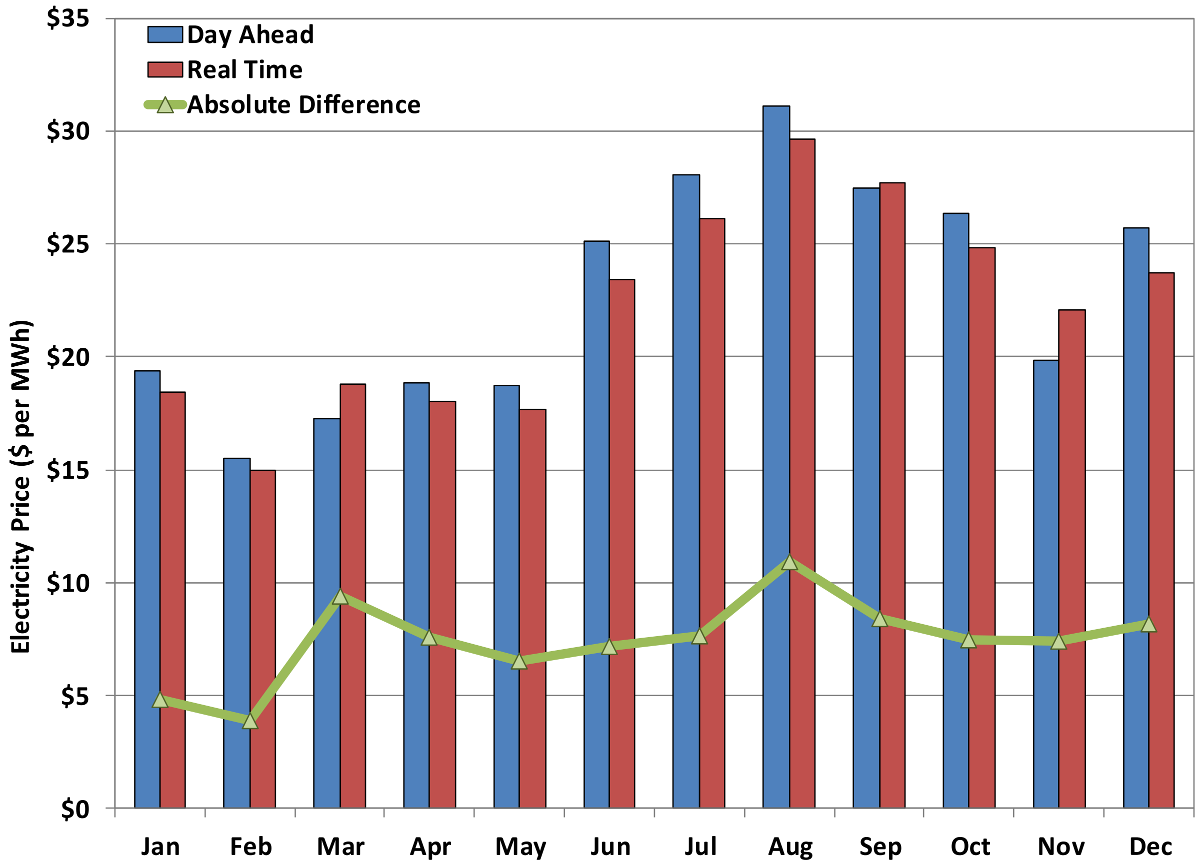


Day-Ahead Market Performance

ERCOT's day-ahead market allows participants to make financially binding forward purchases and sales of power for delivery in real-time. Although all bids and offers are evaluated for the ability to reliably flow on the transmission network, there are no operational obligations resulting from the day-ahead market. These transactions are made for a variety of reasons, including satisfying the participant's own demand, managing risk by hedging the participant's exposure to real-time prices or congestion, or arbitraging the real-time prices. 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 these reasons, the performance of the day-ahead market is essential.

Day-ahead market performance is primarily evaluated by its convergence with 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 on a risk-adjusted basis by making day-ahead purchases or sales to arbitrage the price differences. The next figure shows the price convergence between the day-ahead and real-time markets in 2016.

Convergence Between Day-Ahead and Real-Time Energy Prices



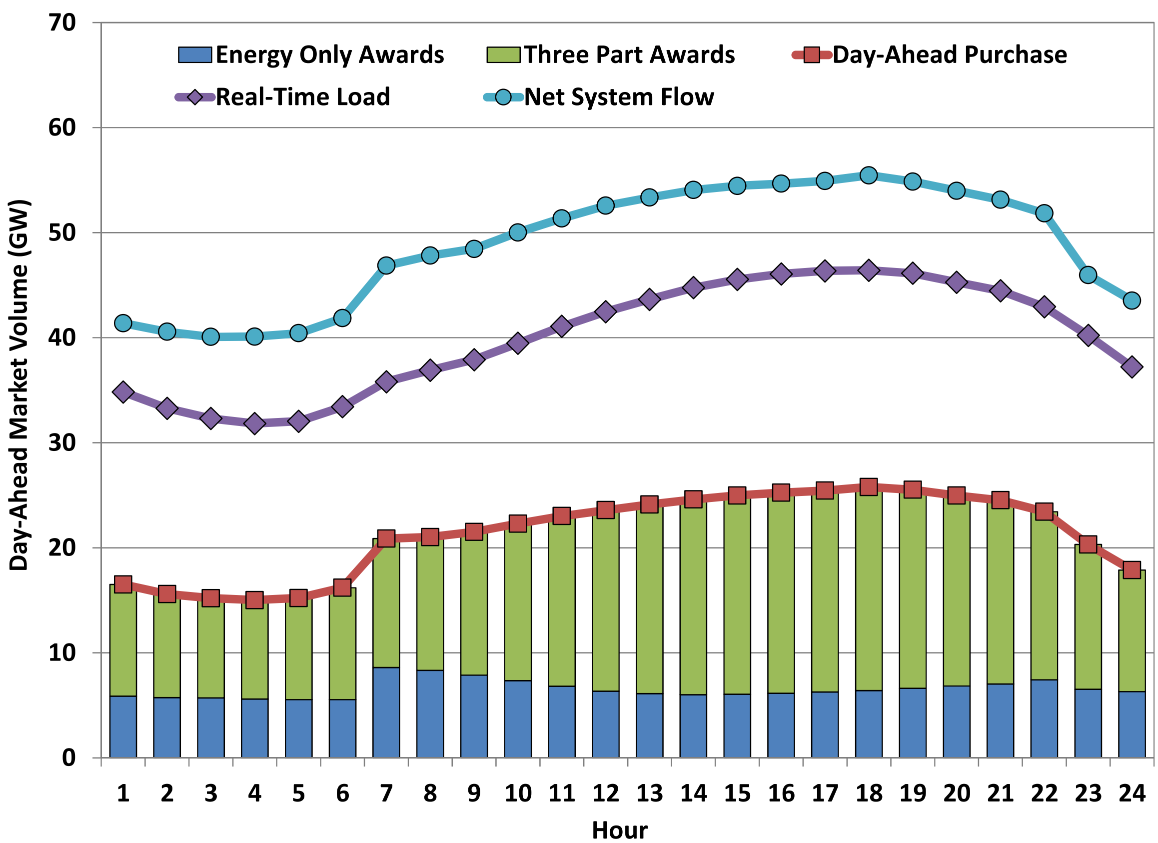
Price convergence was good in 2016 – day-ahead prices averaged $23 per MWh in 2016 compared to an average real-time price of $22 per MWh.[2] The overall day-ahead premium decreased slightly in 2016 from 2015. The average absolute difference between day-ahead and real-time prices was $7.44 per MWh in 2016, down slightly from $8.08 per MWh in 2015.

This day-ahead 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 market and higher for generators selling day ahead. The higher risk for generators is associated with the potential of incurring a forced outage and having to buy back energy at real-time prices. This explains why the highest premiums occurred during the summer months in 2016 with the highest relative demand and highest prices.

Day-Ahead Market Scheduling

The next figure summarizes the volume of day-ahead market activity by month, which includes both the purchases and sales of energy, as well as the scheduling of PTP obligations that represent the system flows between two locations.

Volume of Day-Ahead Market Activity by Month



2 These values are simple averages as previously presented.

The figure shows that the volume of day-ahead purchases provided through a combination of generator-specific and virtual energy offers was approximately 53 percent of real-time load in 2016, which was a slight increase compared to 51 percent in 2015.

PTP obligations are financial transactions purchased in the day-ahead market. Although PTP obligations do not themselves involve the direct supply of energy, PTP obligations allow a participant to buy the network flow from one location to another.[3] When coupled with a self-scheduled generating resource, the PTP allows a participant to service its load while avoiding the associated real-time congestion costs between the locations. Other PTPs are scheduled by financial participants seeking to arbitrage locational congestion differences between the day-ahead and real-time markets.

To provide a volume comparison, all of these “transfers” are aggregated with other day-ahead energy purchases and sales, netting location-specific injections against withdrawals to arrive at a “net system flow.” The net system flow in 2016 was more than 5 percent lower than in 2015. However, it exceeded real-time load by approximately 22 percent. This does not necessarily suggest that the real-time load is fully hedged by day-ahead purchases and PTP obligations since some of the PTP obligations are scheduled by financial participants that do not serve load. Nonetheless, it is likely that a much higher share of the real-time load is hedged in the day-ahead market than the 53 percent scheduling level discussed above.

Ancillary Service Prices

Total requirements for ancillary services declined in 2016, resulting in lower prices and lower total costs for ancillary services. Under the nodal market, ancillary services and energy are co-optimized in the day-ahead market. This means that market participants do not have to include expectations of forgone energy sales in ancillary service capacity offers. Because ancillary service clearing prices explicitly account for the opportunity costs of selling energy in the day-ahead market, ancillary service prices should generally be correlated with day-ahead energy prices. This correlation was not obvious in 2016 as other factors contributed to changes in ancillary service prices.

The next table compares the average annual price for each ancillary service in 2016 with 2015. The changes in total requirements for ancillary services in 2016 led to concomitant changes in ancillary service prices. The average price for responsive reserve remained about the same, as did the total requirements for the service. Reductions in the average price for non-spinning reserves and regulation up is consistent with the reduced requirements for each of those products.

The prices for all of the ancillary service products remain modest in part due to the lack of shortages in 2016. When ERCOT experiences a shortage of operating reserves, real-time prices

[3] PTP Obligations are equivalent to scheduling virtual supply at one location and virtual load at another.

will rise to reflect the expected value of lost load embedded in the ORDC mechanism. The expectation of higher real-time prices will tend to drive up the day-ahead price for ancillary services. Hence, the lack of shortages contributed to the low average ancillary service prices shown in the table.

Average Annual Ancillary Service Prices by Service

	2015 ($ per MWh)	2016 ($ per MWh)
Responsive Reserve	**10.87**	**11.10**
Non-Spinning Reserve	**6.92**	**3.91**
Regulation Up	**10.59**	**8.20**
Regulation Down	**6.01**	**6.47**

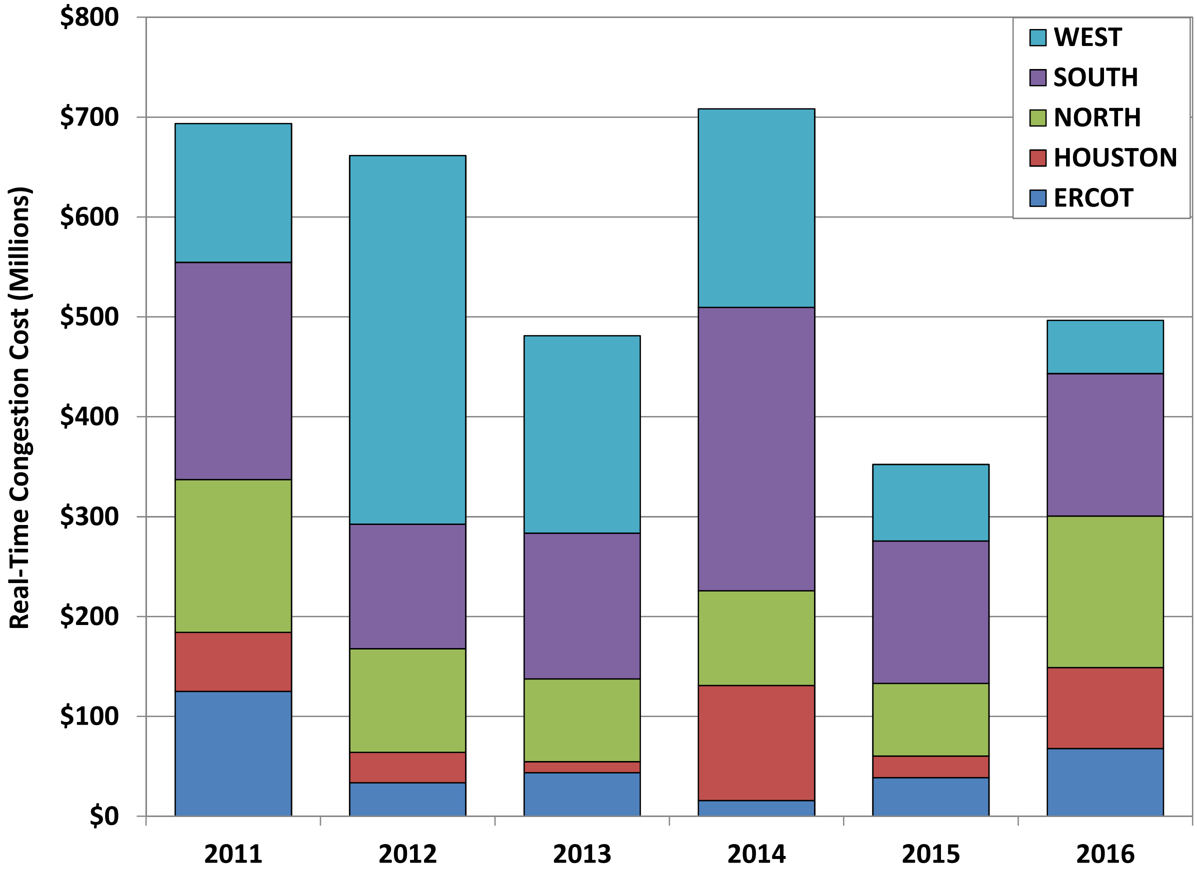
Transmission and Congestion

Congestion arises when the transmission network does not have sufficient capacity to dispatch the least expensive generators to satisfy demand. When congestion occurs, clearing prices vary by location to reflect the cost of meeting load at each location. These nodal prices reflect that higher-cost generation is required at locations where transmission constraints prevent the free flow of power from the lowest-cost resources.

The total congestion costs experienced in the ERCOT real-time market were $497 million in 2016, a 40 percent increase from 2015. This is a substantial increase, especially given the reduction in natural gas prices that would typically reduce transmission congestion. The increase in congestion occurred as constraints were binding in 8 percent more intervals in 2016. These increases were largely driven by higher congestion levels within the Houston and the North zones, and between these two zones. In fact, cross-zonal congestion in 2016 was the most costly since 2011 due to the increased frequency and cost associated with Houston import constraints. Most of the increased congestion was attributable to a variety of transmission outages, some of which were taken to perform system upgrades. The completion of these upgrades is expected to reduce associated congestion.

The next figure displays the amount of real-time congestion costs associated with each geographic zone. Costs associated with constraints that cross zonal boundaries, for example, North to Houston, are shown in the ERCOT category.

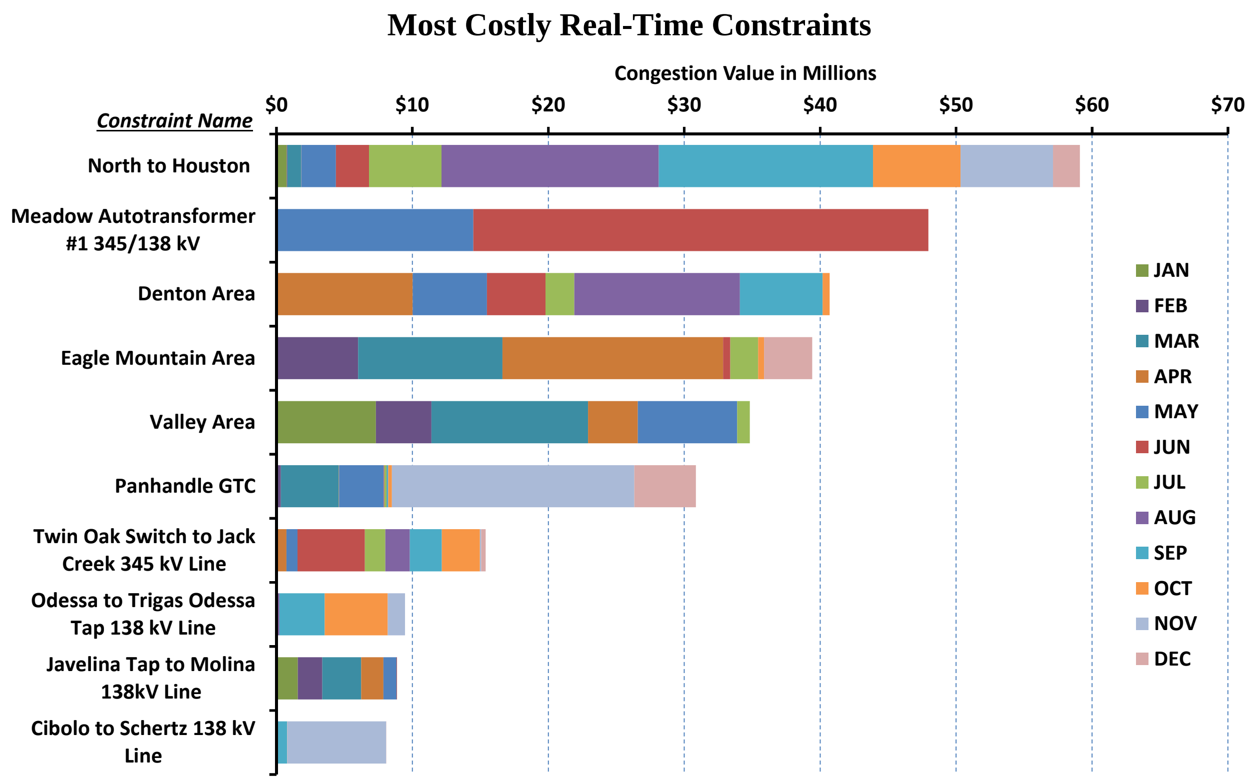
Real-Time Congestion Costs



The figure shows that the North and Houston zones experienced an increase in price impacts between and within the two zones in 2016. Congestion costs for the West and South zones were very similar to 2015.

To better understand the main drivers of congestion in 2016, the next analysis describes the congested areas with the highest financial impact. For this discussion, a congested area is determined by consolidating multiple real-time transmission constraints that are determined to be similar due to their geographic proximity and constraint direction.

The figure below displays the ten most costly real-time constraints as measured by congestion value. The North to Houston constraint, comprised of a generic transmission constraint (GTC) and multiple thermal constraints, was the most congested location in 2016 at $59 million. This area was also the most costly in 2015 at $38 million.



Demand and Supply

Load in 2016

Total ERCOT load over the calendar year increased 1.1 percent (approximately 450 MW on average) to total 351.5 TWh in 2016. As 2016 was a leap year, the relative increase in the total load is higher than the increase in average load. With the exception of the North zone, all zones showed an increase in average real-time load in 2016. Houston saw the largest average load increase at 2.9 percent. Changes in average loads were largely explained by summer weather. Cooling degree days increased 4 percent on average from 2015 to 2016 in Houston and decreased 3 percent in Dallas.

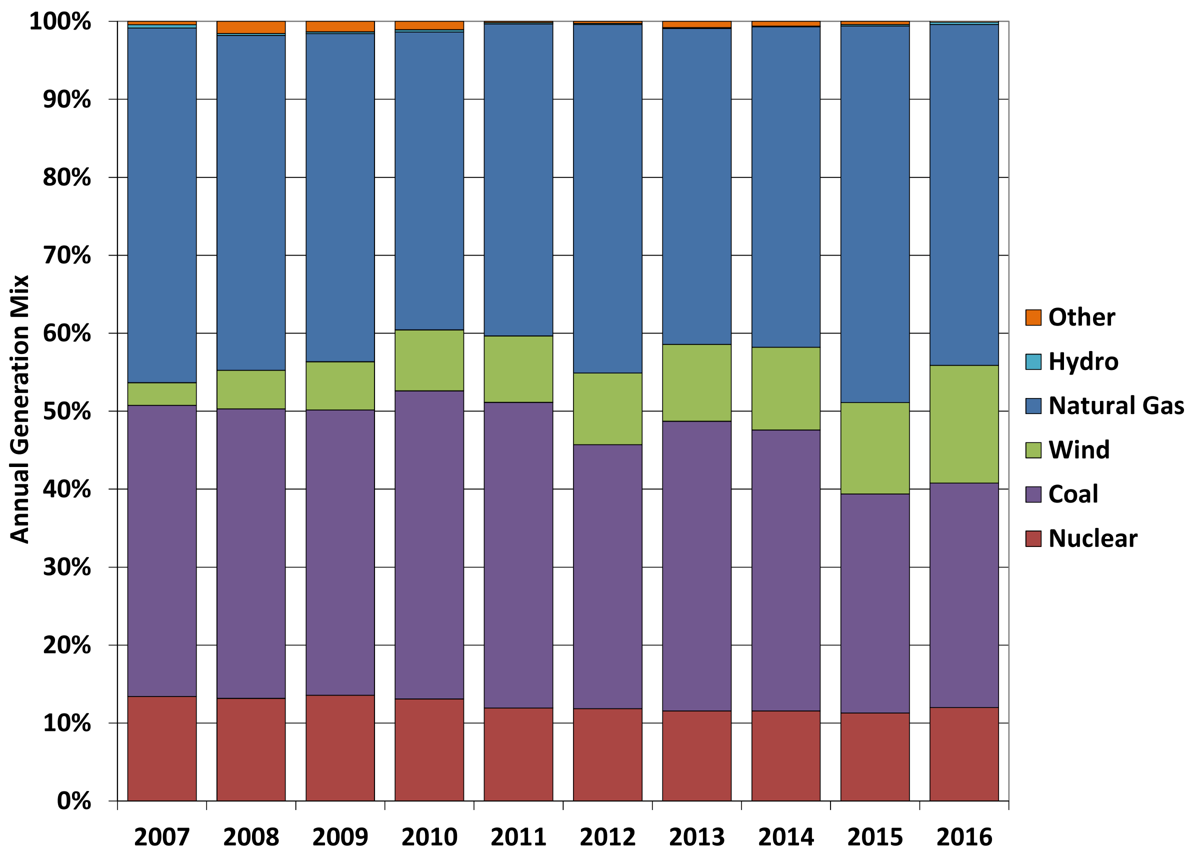
Summer conditions in 2016 also led to a new ERCOT-wide coincident peak hourly demand record of 71,110 MW on August 11, 2016. This was a 1.8 percent increase over the prior year's peak demand record of 69,877 MW. In fact, demand exceeded 70,000 MW five different times in 2016. The zones experienced varying changes in peak load. Although the West zone had shown a prior trend of increasing peak loads due to oil and gas production activity, that trend reversed in 2016 with a decrease in West zone peak load corresponding with a decline in oil and gas activity. Houston also showed a decrease in peak load. The South zone had the greatest increase in peak load at 4.6 percent.

Generating Resources

Approximately 5.5 GW of new generation resources came online in 2016, providing roughly 2 GW of net effective capacity. The overwhelming majority of new capacity was from wind generation. The 4.1 GW of newly installed wind capacity provides approximately 645 MW of peak capacity. The remaining 1.4 GW of new capacity consisted of 370 MW of solar resources, 10 MW of storage resources, and approximately 1 GW of new natural gas combined-cycle units.

Considering these additions and unit retirements in 2016, natural gas generation decreased slightly from 48 percent of total ERCOT installed capacity in 2015 to 45 percent in 2016. The share of total installed capacity for coal generation also decreased slightly from 20 percent in 2015 to 17 percent in 2016. The shifting contribution of coal and wind generation is evident in the figure below showing the percent of annual generation from each fuel type for the years 2007 through 2016.

Annual Generation Mix

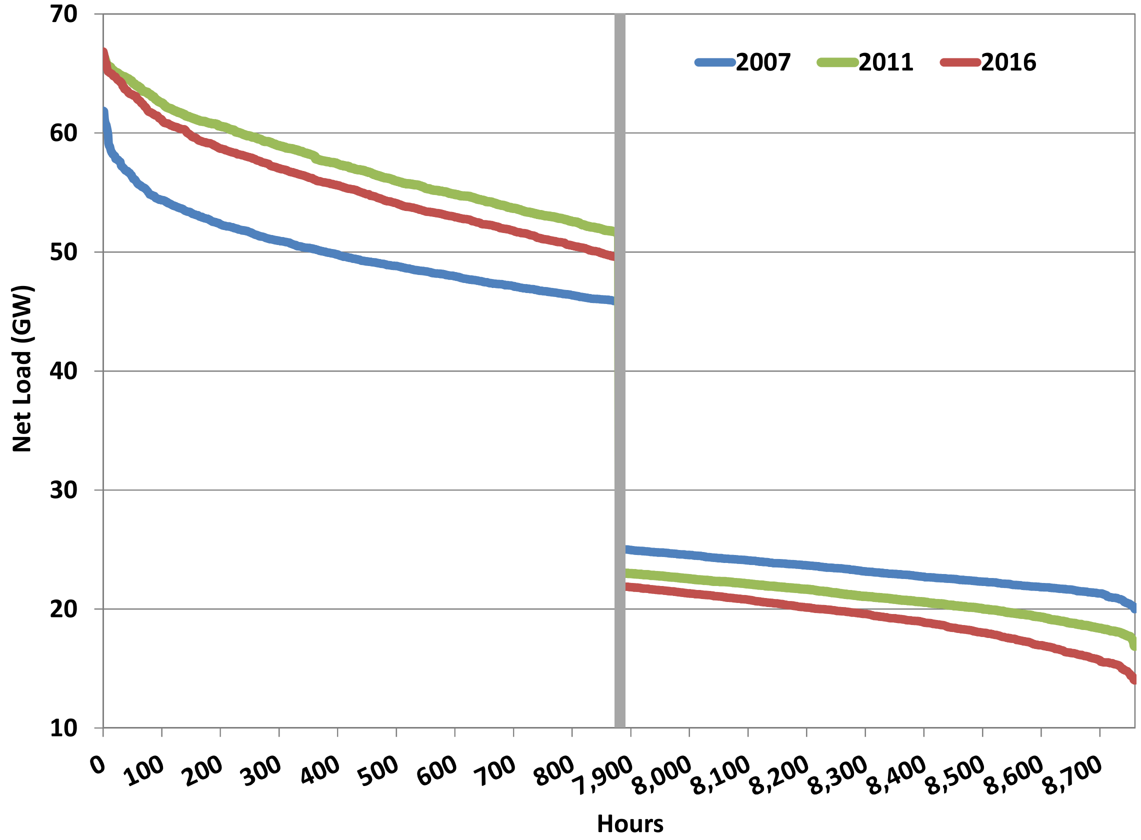


The generation share from wind has increased every year, reaching 15 percent of the annual generation requirement in 2016, up from 3 percent in 2007 and 12 percent in 2015. While the percent of generation from coal had declined significantly between 2014 and 2015, its share increased slightly to 29 percent in 2016. Natural gas declined from its high point in 2015 at 48 percent to 44 percent in 2016.

Wind Output

ERCOT continued to set new records for peak wind output in 2016. On December 25, wind output exceeded 16 GW, setting the record for maximum output and providing nearly 47 percent of the total load. Increasing levels of wind resources in ERCOT have 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 net load in the highest and lowest hours.

Top and Bottom Ten Percent of Net Load



Even with the increased development activity in the coastal area of the South zone, 73 percent of the wind resources in the ERCOT region are located in West Texas. The wind profiles in this area are such that most of the wind production occurs during off-peak hours or other times of low system demand. This profile results in only modest reductions of the net load relative to the actual load during the highest demand hours, but much larger reductions in the net load in the other hours of the year. Wind generation erodes the total load available to be served by base load coal units, while doing very little to reduce the amount of capacity necessary to reliably serve peak load.

In the hours with the highest net load (left side of the figure above), the difference between peak net load and the 95th percentile of net load has averaged 12.3 GW the past three years. This means that 12.3 GW of non-wind capacity is needed to serve load less than 440 hours per year.

In the hours with the lowest net load (right side of the figure), the minimum net load has dropped from approximately 20 GW in 2007 to below 15.4 GW in 2016, even with the sizable growth in annual load that has occurred. This continues to put operational pressure on the 24 GW of nuclear and coal 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 satisfy ERCOT's reliability requirements, the non-wind fleet can expect to operate for fewer hours as wind penetration increases. This outlook further reinforces the importance of efficient energy pricing during peak demand conditions and other times of system stress, particularly in the context of the ERCOT energy-only market design.

Reliability Commitments

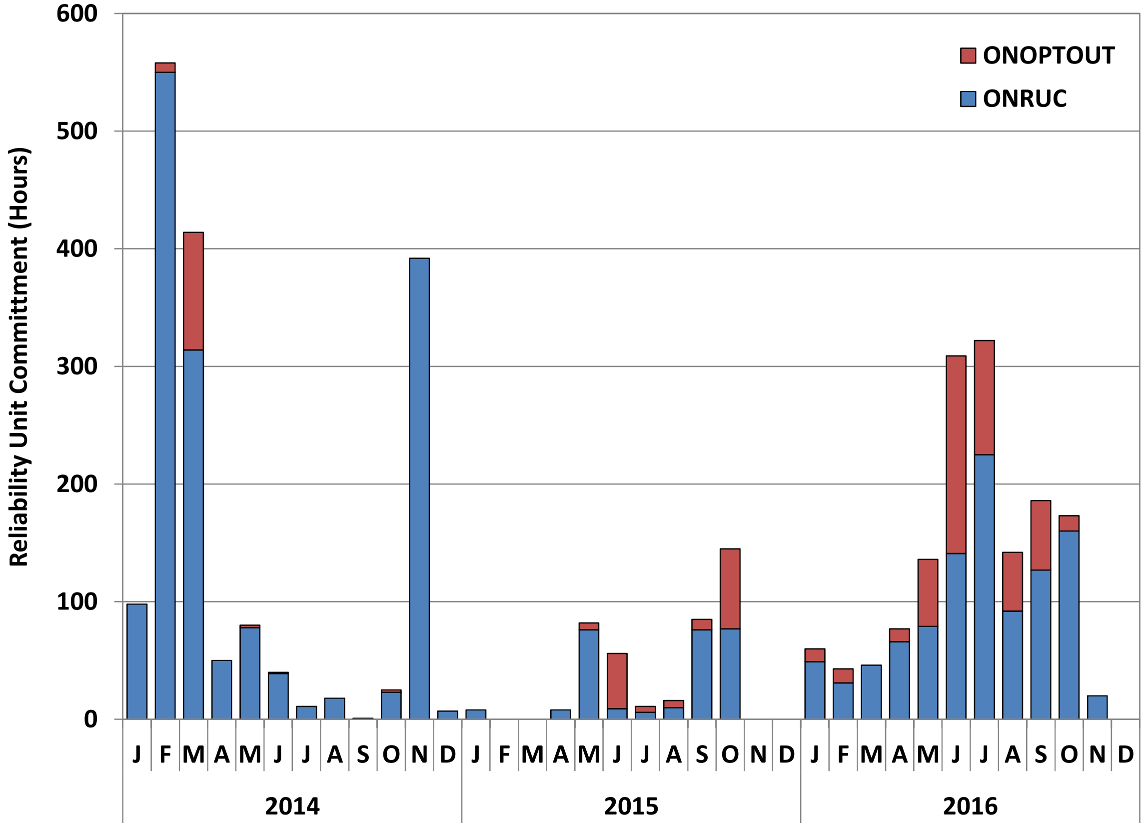
One of the important characteristics of any electricity market is the extent to which it results in the efficient commitment of generating resources. Under-commitment can cause apparent shortages in the real-time market and inefficiently high energy prices; while over-commitment can result in excessive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

The ERCOT market does not include a mandatory centralized unit commitment process. The decision to start-up or shut-down a generator is made by the market participant. ERCOT's day-ahead market outcomes help to inform these decisions, but ERCOT's day-ahead market is only financially binding. That is, when a generator's offer to sell is selected (cleared) in the day-ahead market there is no corresponding requirement to actually start that unit. The generator will be financially responsible for providing the amount of capacity and energy cleared in the day-ahead market whether or not the unit operates.

ERCOT continually assesses the adequacy of market participants' resource commitment decisions using a reliability unit commitment (RUC) process that executes both on a day-ahead and hour-ahead basis. Additional resources may be determined to be needed for two reasons – to satisfy the total forecasted demand, or to make a specific generator available resolve a transmission constraint. The constraint may be either a thermal limit or a voltage concern. The next figure shows how frequently these reliability unit commitments have occurred over the past three years, measured in unit-hours.

When a participant receives a RUC instruction, it may "opt-out" of the instruction by voluntarily starting its unit and receiving the real-time market revenue. If the supplier does not opt-out, it will receive a make-whole payment to cover its cost, but will relinquish the market revenues in excess of its cost through a "clawback" provision.

Frequency of Reliability Unit Commitments



RUC commitments in 2016 were more frequent than in recent years. Although the total unit-hours were similar to the unit-hours in 2014, they were much more consistent in 2016. Almost 12 percent of hours in 2016 had at least one unit receiving a reliability unit commitment instruction. The reliability commitments in 2016 were made primarily to manage transmission constraints (98 percent of unit-hours), most of which addressed persistent congestion in the Houston area and in the Rio Grande Valley.

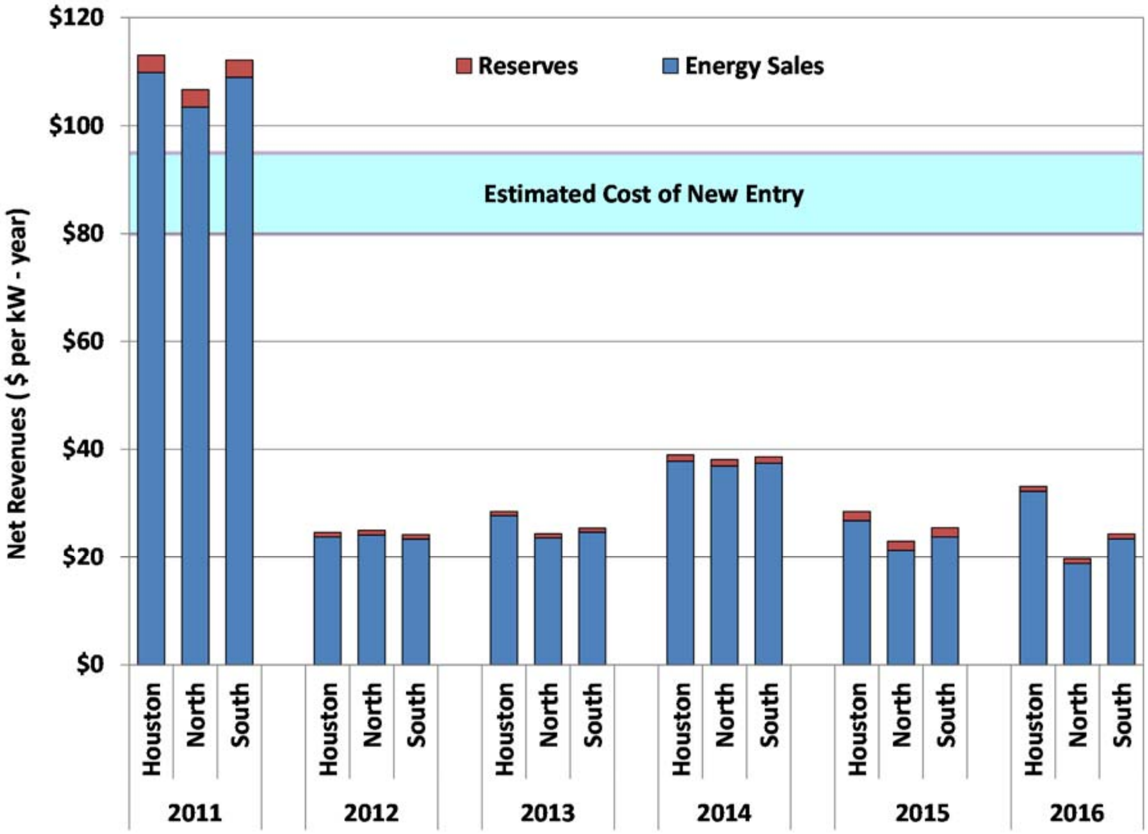
Suppliers opted-out of 32 percent of the RUC instructions in total. Although the quantities increased substantially in 2016, the RUC commitments did not increase costs to ERCOT loads because the make-whole payments were slightly smaller in aggregate than the clawback revenues.

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 system demands and reliability needs. These economic signals are best measured with the net revenue metric, which is calculated by determining the total revenue that could have been earned by a generating unit less its 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. In ERCOT's energy-only market, the net revenues from the real-time energy and ancillary services markets alone provide the economic signals that inform suppliers' decisions to invest in new generation or retire existing generation. To the extent that revenues are available through the day-ahead market or other forward bilateral contract markets, these revenues are ultimately derived from the expected real-time energy and ancillary service prices.

The next figure provides an historical perspective of the net revenues available to support investment in a new natural gas combustion turbine, selected to represent the marginal new supply that may enter when new resources are needed. The figure also shows the estimated "cost of new entry," which represents the revenues needed to break even on the investment.

Combustion Turbine Net Revenues



Based on 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 $95 per kW-year. These estimates reflect Texas-specific construction costs. The net revenue in 2016 for a new gas turbine was calculated to be approximately $23 to 29 per kW-year, depending on the zone location, which are well below the estimated cost of new gas turbine generation.

These results are consistent with the current surplus capacity, which contributed to infrequent shortages in 2015 and 2016. In an energy-only market, shortages play a key role in delivering the net revenues an investor would need to recover its investment. Such shortages will tend to be clustered in years with unusually high load and/or poor generator availability. Hence, these results alone do not raise substantial concerns regarding design or operation of ERCOT's ORDC mechanism for pricing shortages.

Given the very low energy prices during 2016 in non-shortage hours, the economic viability of existing coal and nuclear units was evaluated. Non-shortage prices, which have been substantially affected by the prevailing natural gas prices, determine the vast majority of net revenues received by these base load units. The generation-weighted average price for the four nuclear units in ERCOT - approximately 5 GW of capacity - was only $21.46 per MWh in 2016, down from $24.56 per MWh in 2015. According to the Nuclear Energy Institute (NEI), total operating costs for all nuclear units across the U.S. averaged $27.17 per MWh in 2016.[4] Assuming that operating costs in ERCOT are similar to the U.S. average, it is likely that these units were not profitable in 2016, based on the fuel and operating and maintenance costs alone. To the extent nuclear units in ERCOT had any associated capital costs, it is likely those costs were not recovered. Compared to other regions with larger amounts of nuclear generation, the four nuclear units in ERCOT are relatively new and owned by four entities with sizable load obligations. Although not profitable on a stand-alone basis, the nuclear units have substantial option value for the owners because they ensure that the cost of serving their load will not rise substantially if natural gas prices increase. Nonetheless, the economic pressure on these units does potentially raise a resource adequacy issue that will need to be monitored.

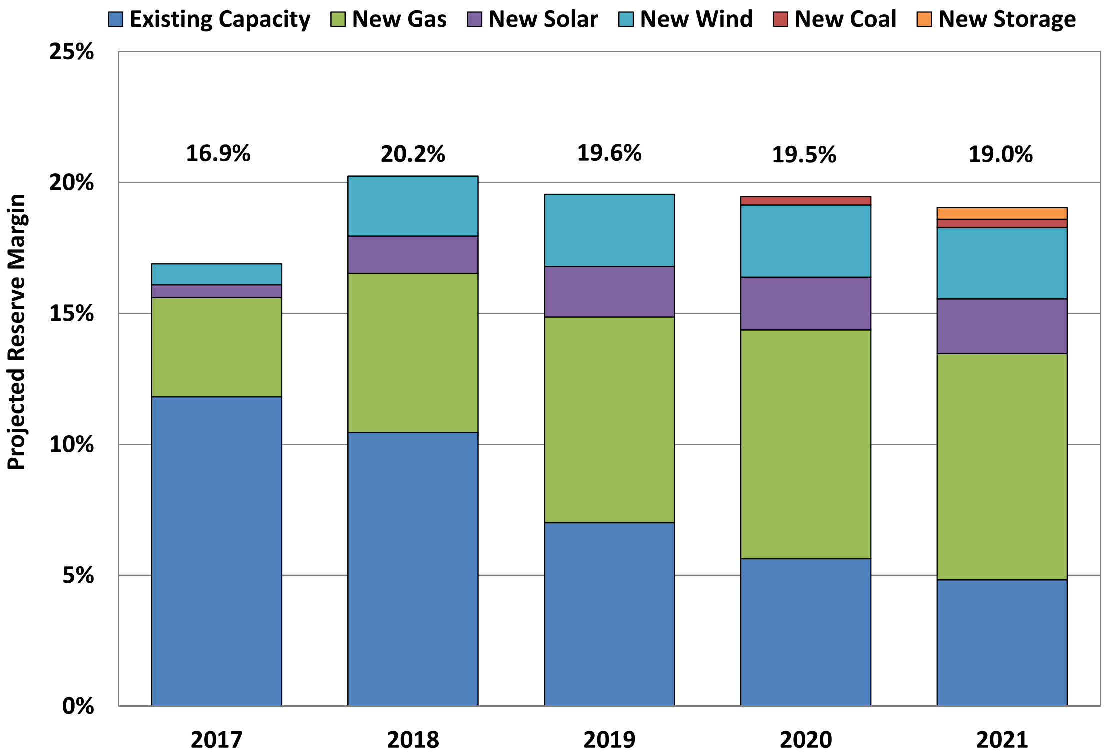
The generation-weighted price of all coal and lignite units in ERCOT during 2016 was $23.98 per MWh. Although specific unit costs may vary, index prices for Powder River Basin coal delivered to ERCOT were approximately $2.50 per MMBtu in 2016, a decrease from approximately $2.60 per MMBtu in 2015. For the past two years, delivered coal costs in ERCOT have been about $0.03 to $0.05 per MMBtu higher than natural gas prices at the Houston Ship Channel. Given that the coal units generally have higher heat rates and more expensive non-fuel operations and maintenance costs, they have been losing market share to natural gas. As with nuclear units, it appears that coal units were likely not profitable in ERCOT during 2016. With the bulk of the coal fleet in ERCOT being more than 30 years old, the retirement or suspended operation of some of these units could cause ERCOT's capacity margin to fall to unreliable levels more quickly than anticipated. While both nuclear and coal are feeling

4 NEI Whitepaper, "Nuclear Costs in Context", April 2017, available at https://www.nei.org/www.nei.org/files/fe/fed92b11-8ea6-40df-bb0c-29018864a668.pdf.

the pressure of an increased reliance on lower-priced natural gas units, coal units appear to be at greater risk of retirement than the nuclear units in ERCOT due to their relative age and inefficiency.

The next figure shows ERCOT's current projection of planning reserve margins and indicates that the region will have a 16.9 percent reserve margin heading into the summer of 2017. While these projections are slightly lower than those developed last year, the current outlook is very different than in 2013, when planning reserve margins were expected to be below the then-existing target level of 13.75 percent for the foreseeable future.[5]

Projected Planning Reserve Margins



Source: ERCOT Capacity, Demand and Reserves Report - December 2016

This current projection of planning reserve margins combined with relatively infrequent shortage pricing may raise doubts regarding the likelihood of announced generation coming on line as planned. Given the projections of continued low prices, investors of some of the new generation included in the Report on the Capacity, Demand, and Reserves in the ERCOT Region (CDR) may choose to delay or even cancel their project. Additionally, the profitability analysis of

5 The target planning reserve margin of 13.75 percent was approved by the ERCOT Board of Directors in November 2010, based on a 1 in 10 loss of load expectation (LOLE). The PUCT recently directed ERCOT to evaluate planning reserve margins based on an assessment of the Economically Optimal Reserve Margin (EORM) and the Market Equilibrium Reserve Margin (MERM). *See* PUCT Project No. 42303, ERCOT Letter to Commissioners (Oct. 24, 2016).

existing baseload resources casts doubt on the assumption embedded in the CDR that all existing generation will continue to operate. Hence, it is likely that the planning reserve margins will be lower than forecasted in the figure above.

Analysis of Competitive Performance

The report evaluates market power from two perspectives, structural (does market power exist) and behavioral (have attempts been made to exercise it).

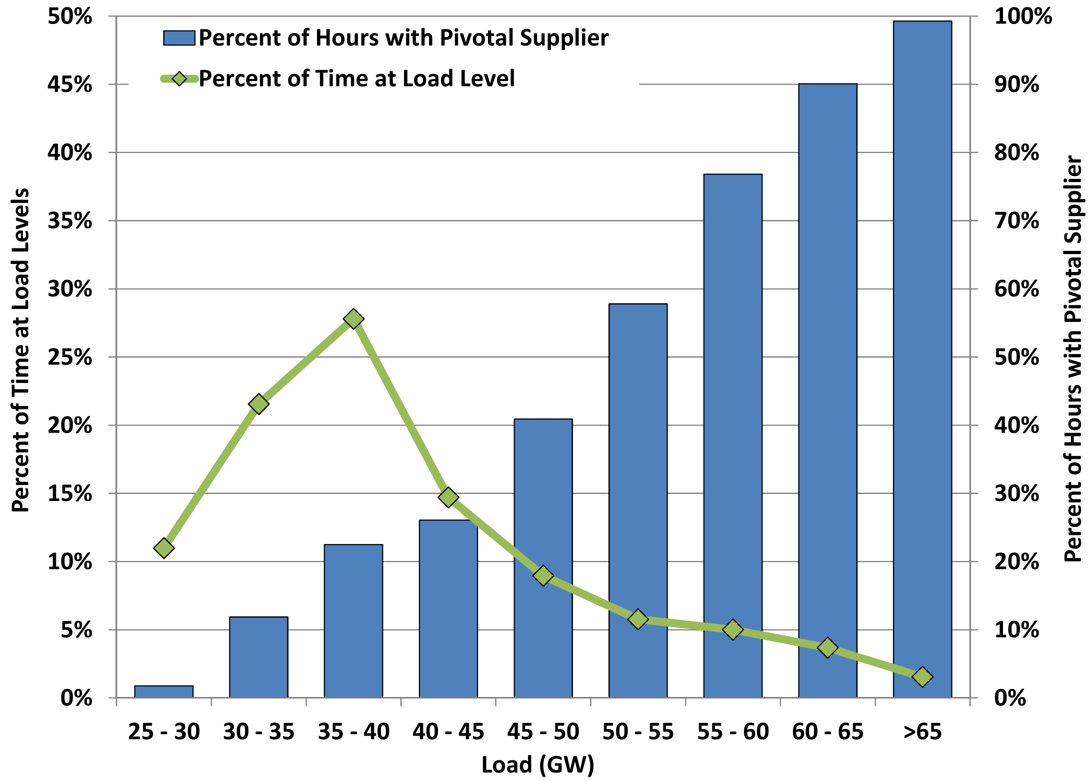
Structural Market Power

The market structure is analyzed by using the Residual Demand Index (RDI), a statistic that measures the percentage of load that could not 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 (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 if 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 could raise prices significantly by withholding resources.

The figure below summarizes the results of the RDI analysis by displaying the percentage of time at each load level there was a pivotal supplier. The figure also displays the percentage of time each load level occurs.

Pivotal Supplier Frequency by Load Level



This figure shows that at loads greater than 65 GW, there was a pivotal supplier 99 percent of the time. This is expected because at high load levels, larger suppliers are more likely to be pivotal because other suppliers' resources are more fully utilized serving the load. The frequency of relatively high loads increased in 2016. This led to an increase in the pivotal supplier frequency to 28.5 percent of all hours in 2016, up from 26 and 23 percent of all hours in 2015 and 2014, respectively. This indicates that market power continues to be a potential concern in ERCOT and underscores the need for effective mitigation measures to address it.

This analysis evaluates the structure of the entire ERCOT market. In general, local market power in narrower areas that can become isolated by transmission constraints raise more substantial competitive concerns. This local market power is addressed through: (a) structural tests that determine "non-competitive" constraints that can create local market power; and (b) the application of limits on offer prices in these areas.

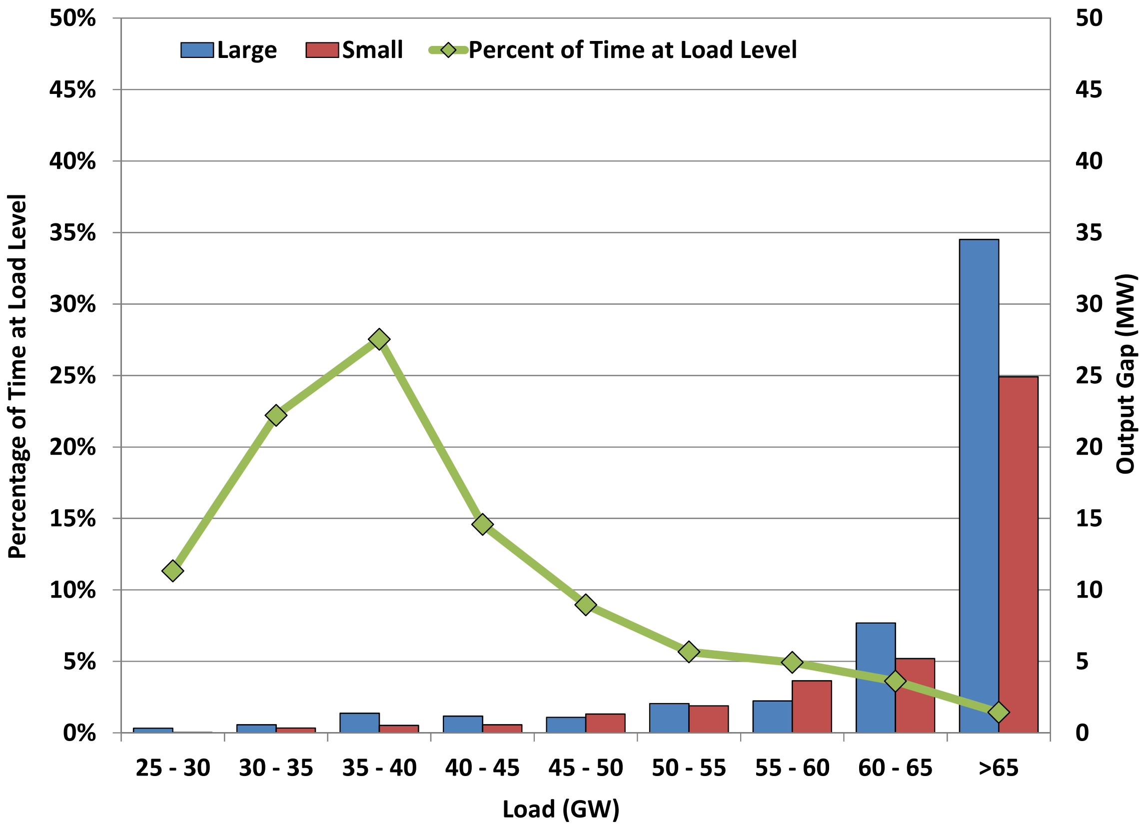
Evaluation of Conduct

In addition to the structural market power analyses above, actual participant conduct was evaluated to assess whether market participants have attempted to exercise market power through physical or economic withholding. An "output gap" metric is used to measure potential economic withholding, which occurs when a supplier raises its offer prices to reduce its output.

The output gap is the quantity of energy that is not being produced by online resources even though the output is economic to produce by a substantial margin given the real-time energy price. A margin of $30 per MWh is used for this analysis. To determine whether the output from a resource is economic to produce, the mitigated offer cap serves as a proxy for the marginal production cost of energy.

The next figure shows the output gap levels, separately showing the results aggregated for the five largest suppliers (those with greater than 5 percent of ERCOT installed capacity) and all other suppliers (i.e., the small category).[6]

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



These results show that potential economic withholding levels were extremely low for the largest suppliers and small suppliers alike in 2016. Output gaps for the largest suppliers are routinely monitored individually and were found to be consistently low across all load levels. These results, together with our evaluation of the market outcomes presented in this report, allow us to conclude that the ERCOT market performed competitively in 2016.

6 In the second step of the dispatch, the after-mitigation offer curve is used to determine dispatch instructions and locational prices. The output gap at Step 2 showed very small quantities of capacity that would be considered part of this output gap.

Recommendations

Overall, we find that the ERCOT market performed well in 2016. However, we have identified and recommended a number of potential improvements to the ERCOT markets. We make seven recommendations in this report, four of which we have previously recommended. These recommendations are categorized by their principle objective: a) to improve the operation of the ERCOT system and its resources; and b) to improve price formation in ERCOT's energy and ancillary services markets. We describe each recommendation below and the benefits that each would provide. For recommendations repeated from prior reports, we discuss the status of progress made to evaluate or implement the recommendation.

Improving Real-Time Operations and Resource Performance

One of the primary functions of the wholesale markets is to coordinate the operations of all resources to satisfy the system's needs at the lowest cost. The recommendations in this section are principally intended to improve the operation of the ERCOT markets, but in doing so will also improve ERCOT's prices and performance incentives. The first two recommendations in this section were considered over the past year, which we describe in the status section for each recommendation.

1. **Evaluate policies and programs that create incentives for loads to reduce consumption for reasons unrelated to real-time energy prices, including: (a) the Emergency Response Service (ERS) program and (b) the allocation of transmission costs.**

Any incentives that cause market participants to take actions that are inconsistent with the real-time prices will undermine the performance of the market and its prices. These concerns are heightened when these actions are taken under peak or emergency conditions because the ERCOT market relies on efficient pricing under such conditions to motivate efficient long-term resource decisions by participants. By curtailing load in response to incentives or programs that are not aligned with the real-time energy market, supply is uneconomically reduced and the real-time market is adversely affected. The following two aspects of the ERCOT market raise these concerns.

ERS Program. A load that wishes to actively participate in the ERCOT market can participate in ERS, provide ancillary services, or simply choose to curtail in response to high prices. Participating in ERS greatly limits a load's ability to provide ancillary services or curtail in response to high prices. Given the high budget allotted and the low risk of deployment, ERS is an attractive program for loads. Because the ERS program is so lucrative, we are concerned that it is limiting the motivation for loads to actively participate and contribute to price formation in the real-time energy market.

Transmission Cost Allocation. Transmission costs in ERCOT are allocated on the basis of load contribution in the highest 15-minute system demand during each of the four months from June through September. This allocation mechanism is routinely referred to as four coincident peak, or 4CP. Over the last three years, transmission costs have risen by more than 60 percent, significantly increasing an already substantial incentive to reduce load during probable peak intervals in the summer. ERCOT estimates that 835-1,491 MW of load were actively pursuing reduction during the 4CP intervals in 2016, an increase from the estimated response in 2015.[7]

Load curtailment to avoid transmission charges may be resulting in price distortion during peak demand periods since the response is targeting peak demand rather than responding to wholesale prices. This was readily apparent in 2016 as there were significant load curtailments corresponding to peak load days in June, July and September when real-time prices on those days were in the range of $25 to $40 per MWh.

Status: In docket number 45927, the PUCT considered changes to the ERS program. Ultimately, the PUCT decided to retain ERS in its current structure, but elected to permit an ERS resource selected as a must-run alternative to a reliability must run contract to modify or terminate its obligations under a pre-existing ERS contract.[8] While the PUCT is considering changes to transmission service rates in Docket No. 46393, changes to the 4CP allocation method are not part of that project.[9] At this time, no final changes have been adopted to transmission service rates.

2. Modify the real-time market software to better commit load and generation resources that can be online within 30 minutes.

The real-time market relies primarily on two classes of resources: online resources and offline resources that can start quickly. The real-time market efficiently dispatches online resources and sets nodal prices that reflect the marginal value of energy at every location, but ERCOT lacks real-time processes to facilitate efficient commitment and decommitment of peaking resources that can start quickly (i.e., within 30 minutes). This is a concern because suboptimal dispatch of these resources raises the overall costs of satisfying the system's needs, distorts the real-time energy prices, and affects reliability. For these reasons, other markets have implemented this type of look-ahead process to optimize short-term commitments of peaking resources. In contrast, ERCOT relies on de-centralized commitment where individual participants bear most of the costs of their own commitment decisions. Because participants lack the information ERCOT has on upcoming conditions and the plans of other participants, this decentralized process will necessarily be less efficient than a fully-optimized real-time process coordinated by ERCOT.

---

7 *See* ERCOT, *2016 Annual Report of Demand Response in the ERCOT Region* (Mar. 2017) at 8, available at http://www.ercot.com/services/programs/load.

8 PUCT Docket Number 45927, *Rulemaking Regarding Emergency Response Service*, Order Adopting Amendment to § 25.507 As Approved at the March 30, 2017 Open Meeting (Mar. 30, 2017).

9 PUCT Docket Number 46393, Rulemaking Proceeding to Repeal and Replace 16 Texas Administrative Code § 25.192, Relation to Transmission Service Rates.

Further, as ERCOT attracts more variable wind and solar resources, the value of having access to and optimally utilizing fast-starting controllable resources will grow. Hence, we continue to recommend that ERCOT develop this capability.

Status: We have been recommending this change since the start of ERCOT's nodal market. After taking interim steps to produce non-binding generation dispatch and price projections and then to improve the short term forecasting procedures, ERCOT evaluated the potential improvement from a multi-interval real-time market. This evaluation determined that because the costs to implement were greater than the projected benefits, moving forward with implementation was not supported at this time.[10] The finding of insufficient benefits is not surprising given the current low-price environment and the level of surplus capacity on the system. However, as planning reserve margins fall and installation of intermittent renewable resources increases, the benefits of enhancement will grow.

3. Implement real-time co-optimization of energy and ancillary services.

Substantial benefits can be achieved by implementing real-time co-optimization of energy and ancillary services. First, jointly optimizing all products in each interval allows ancillary service responsibilities to be continually adjusted in response to changing market conditions. The efficiencies of this continual adjustment would flow to all market participants and would be greater than what can be achieved by QSEs acting individually. The continual, optimal system-wide allocation of resources between providing energy and providing reserves will lower the cost of satisfying both requirements. Additionally, it will ensure that energy is produced in locations where it may be most valuable.

The second benefit from real-time co-optimization will be improved shortage pricing. The Operating Reserve Demand Curve (ORDC) provides a mechanism for setting real-time energy prices that reflect the expected value of lost load. However, jointly-optimizing the energy and reserve markets would allow this shortage pricing to be more accurate. In a co-optimized system, the real-time market will determine in each five minutes whether a shortage of either energy or reserves exists and set prices accordingly. Currently, capacity providing responsive or regulating reserves are not available to be converted into energy at any price. Under a co-optimized system, a demand curve would be established for every type of reserve (potentially including locational reserve products in the future). When it is economic to release these reserves to provide energy, the value of these reserve shortages will be reflected efficiently in the energy and reserve prices. This is especially important in ERCOT because pricing during shortage conditions is key for the success of ERCOT's energy-only market.

[10] *See* PUCT Docket No. 41837, *PUCT Review of Real-Time Co-Optimization in the ERCOT Region*, ERCOT Report on the Multi-Interval Real-Time Market Feasibility Study (Apr. 6, 2017).

Other economic benefits would be achieved by allowing all suppliers to participate equally in ERCOT's ancillary service markets. Currently, QSEs without large resource portfolios are effectively precluded from participating in ancillary service markets because of the replacement risk they face having to rely on a supplemental ancillary services market (SASM).

For all of these reasons, implementing real-time co-optimization of energy and ancillary services is our highest priority recommendation.

Status: The PUCT initiated a project to consider the feasibility of implementing real-time co-optimization in September 2013.[11] After some initial investigation including a draft whitepaper by ERCOT, the project was temporarily put on hold to consider whether a Multi-Interval Real-Time Market (MIRTM) should be pursued first or in conjunction with real-time co-optimization. In early 2017, the PUCT provided direction to ERCOT to restart the evaluation of implementing real-time co-optimization.[12]

Improving Price Formation in the ERCOT Market

4. Price future ancillary services based on the shadow price of procuring the service.

In a well-functioning real-time market, the market model will indicate the marginal cost of satisfying any requirement, which is the shadow price of the requirement. This shadow price is the most efficient clearing price for each of ERCOT's ancillary service requirements. Hence, we recommend that any new or updated ancillary services be priced on this basis.

Status: In the context of stakeholder discussions about Future Ancillary Services, we re-introduced our recommendation that the clearing price of a service be based on the shadow price of any constraint used in the procurement of that service. At this point, we are not recommending any changes to the current ancillary services procurement or pricing practices, although the current pricing of responsive reserves is inefficient. As changes are made to ancillary services, we believe it is appropriate to include this change to improve the pricing of these products and suppliers' incentives.

5. Ensure that the price of any energy deployed from a reliability must run (RMR) unit reflects the shortage conditions that exist by the fact that there is an RMR unit.

Currently RMR units are required to submit energy offer curves with prices equal to the system-wide offer cap. This requirement was implemented shortly after four units were brought back to service from mothball status during the extreme heat of the summer of 2011. The

11 *See* PUCT Docket No. 41837.

12 *Id.*, ERCOT Letter to Chairman and Commissioners (Apr. 27, 2017), responding to Commissioner direction at the April 13, 2017 Open Meeting directing ERCOT "to restart the evaluation of the potential implementation of the co-optimization of energy and operating reserves in the real-time market."

purpose was to ensure that the energy from these RMR units, needed for overall generation adequacy, was priced to reflect the value of lost load.

Other, future RMR units may be needed to resolve local transmission constraints, as was the case with Greens Bayou RMR. In that situation, the RMR unit energy offer price will likely be mitigated. Mitigating energy offers from an RMR unit may result in the unit being dispatched prior to other competitively-offered units, especially if output from the RMR unit is more helpful in unloading the relevant transmission constraint. In the absence of any other market changes designed to reflect the reliability needs that caused the RMR, we believe that pricing the energy from the RMR unit such that its costs to resolve the relevant constraint are higher than the costs of other available market-based resources will establish more efficient economic signals in the ERCOT market.

Status: This is a new recommendation.

6. Evaluate the need for a local reserve product.

In an energy-only market, all economic signals to support long-term investment and retirement decisions are provided by the energy and ancillary service markets. A substantial component of these economic signals is the prices and revenues generated in shortage conditions. ERCOT's ORDC establishes shortage pricing ERCOT-wide, but does not allow for shortage pricing in local areas. Therefore, ERCOT's current market design may support adequate resources in aggregate, but may not support adequate resource in some local areas.

It is common in other markets to plan and operate the system to be able to maintain reliability in a local area even after the two largest contingencies occur (transmission or generation outages). This is one of the most common reasons that a unit may be deemed needed for reliability and given an RMR contract, but such an action should be seen as a failure of the wholesale market to provide sufficient revenues to support the continued operation of the resource.

In ERCOT's energy-only market, the primary means to ensure that sufficient revenues are provided to satisfy both the market-wide and local resource adequacy needs is to strive for alignment between ERCOT's operating requirements and its planning requirements. In other words, if having sufficient resources to respond to the two largest contingencies is a reasonable planning requirement, it is also likely a reasonable operating requirement. Other RTO's include this requirement in their operating reserve markets by establishing a separate, localized 30-minute reserve product. The advantage of defining such an ancillary service product in ERCOT is that it would allow the real-time energy and reserve markets to price local reserve shortages and provide the revenues necessary to satisfy local capacity needs. In doing so, it should eliminate the need to sign out-of-market RMR contracts.

Hence, we recommend that ERCOT align its planning requirements and real-time operating requirements and begin evaluating the need for a local reserve product. Changes to the process for determining whether an RMR unit is needed, implemented in NPRR788, were important clarifications. However, if there is a local reliability concern that is best addressed by maintaining additional operating reserves in a specific area, we suggest that ERCOT develop and implement a new local reserve product.

Status: This is a new recommendation.

7. Consider including marginal losses in ERCOT locational marginal prices.

When electricity is produced in one location and consumed at another location, the electricity flows through the transmission system and some of it is lost. The transmission losses vary depending on the distance the electricity is traveling and the voltage of the lines it must flow over. Ideally, the real-time dispatch model should recognize the marginal losses that will result from dispatching units in different locations and set prices accordingly. Recognizing marginal losses will allow the real-time market to produce more from a higher-cost generator located electrically closer to the load, thus resulting in fewer losses. Optimizing this trade-off in the real-time dispatch lowers the overall costs of satisfying the system's needs.

The ERCOT market is unique in its treatment of transmission losses. Marginal losses are not included in ERCOT real-time energy prices and the costs of losses are collected from loads on an average basis. This approach may have been reasonable at the time ERCOT was implementing its initial real-time energy markets because generators were relatively close to load centers. However, as open access transmission expansion policies and other factors have led to a wider dispersion of the generation fleet, the failure to recognize marginal losses in the real-time dispatch and pricing has led to larger dispatch inefficiencies and price distortions. Therefore, we are now recommending that the ERCOT real-time market be upgraded to recognize marginal losses in its dispatch and prices.

Accompanying this change, a revenue allocation methodology will need to be developed because marginal loss pricing results in the collection of more payments for losses than the aggregate cost of losses. This occurs because the marginal losses are always larger than the average losses (i.e., losses increase as more power flows over the transmission system). Most other RTOs in the U.S. recognize marginal losses and may provide examples of allocation approaches that could be used in ERCOT.

Status: This is a new recommendation.

I. REVIEW OF REAL-TIME MARKET OUTCOMES

Although only a small share of the power produced in ERCOT is transacted in the spot market, real-time energy prices are very important because they set the expectations for prices in the day-ahead market and bilateral forward markets where most transactions occur. 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, low prices in the real-time energy market will translate to 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 2016.

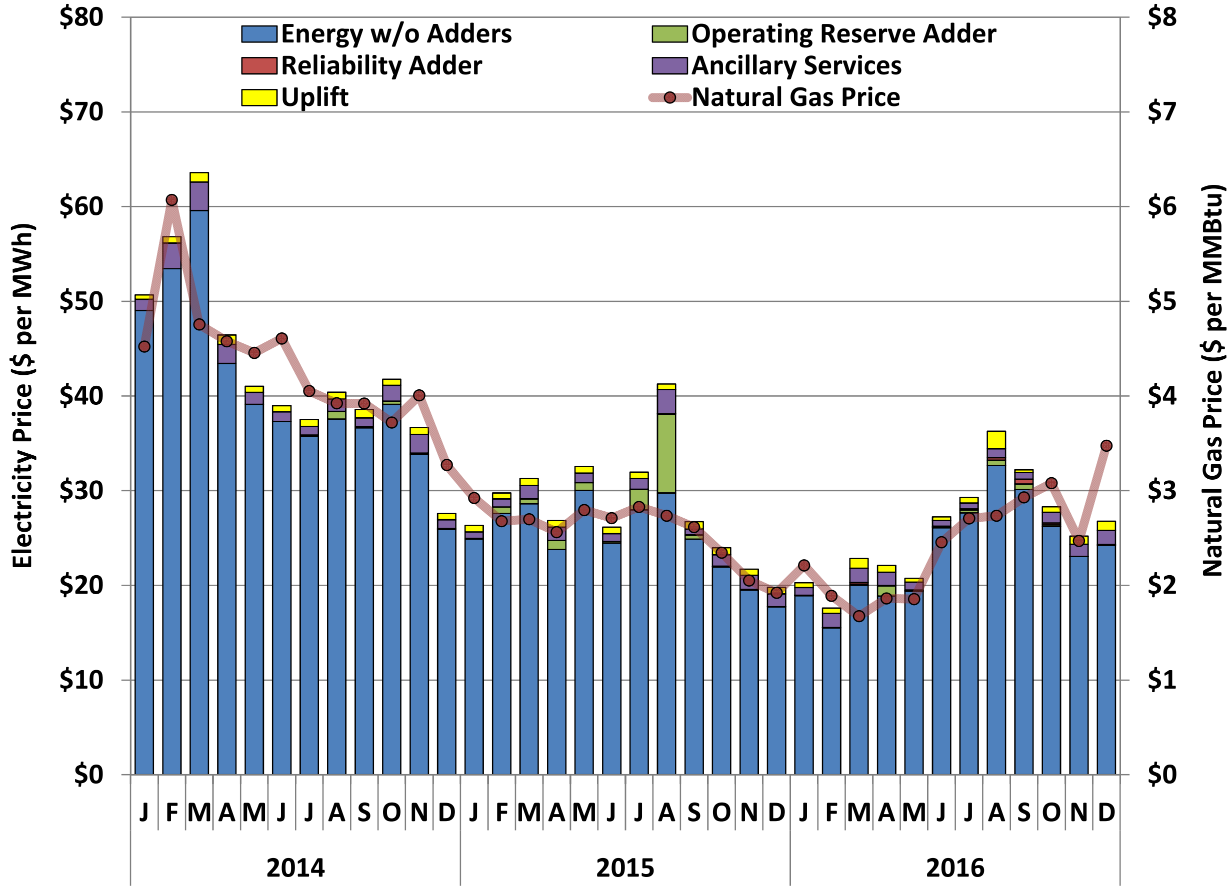
A. Real-Time Market Prices

The 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." An average "all-in" price of electricity has been calculated for ERCOT that is intended to reflect wholesale energy costs as well as these additional costs.

Figure 1 summarizes changes in energy prices and other market costs by showing the all-in price of electricity, which is a measure of the total cost of serving load in ERCOT for 2014 through 2016. The ERCOT-wide price in this figure is the load-weighted average of the real-time market prices from all zones. Ancillary services costs and uplift costs are divided by real-time load to show them on a per MWh basis.[13] ERCOT developed two energy price adders that are designed to improve its real-time energy pricing when conditions or ERCOT takes out-of-market actions for reliability. To distinguish the effects of the energy price adders, the Operating Reserve Demand Curve Adder (operating reserve adder) and the Reliability Deployment Price Adder (reliability adder) are shown separate from the energy price. The operating reserve adder was implemented in mid-2014 to account for the value of reserves based on the probability of reserves falling below the minimum contingency level and the value of lost load. The reliability adder was implemented in June 2015 as a mechanism to ensure that reliability deployments do not distort the energy prices. The reliability adder is calculated using a separate price run of SCED, removing any Reliability Unit Commitments (RUC) or deployed load capacity and recalculating prices. When the recalculated system lambda (average load price) is higher than the initial system lambda, the increment is the adder.

[13] For this analysis Uplift includes: Reliability Unit Commitment Settlement, Operating Reserve Demand Curve (ORDC) Settlement, Revenue Neutrality Total, Emergency Energy Charges, Base Point Deviation Payments, Emergency Response Service (ERS) Settlement, Black Start Service Settlement, and Block Load Transfer Settlement.

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

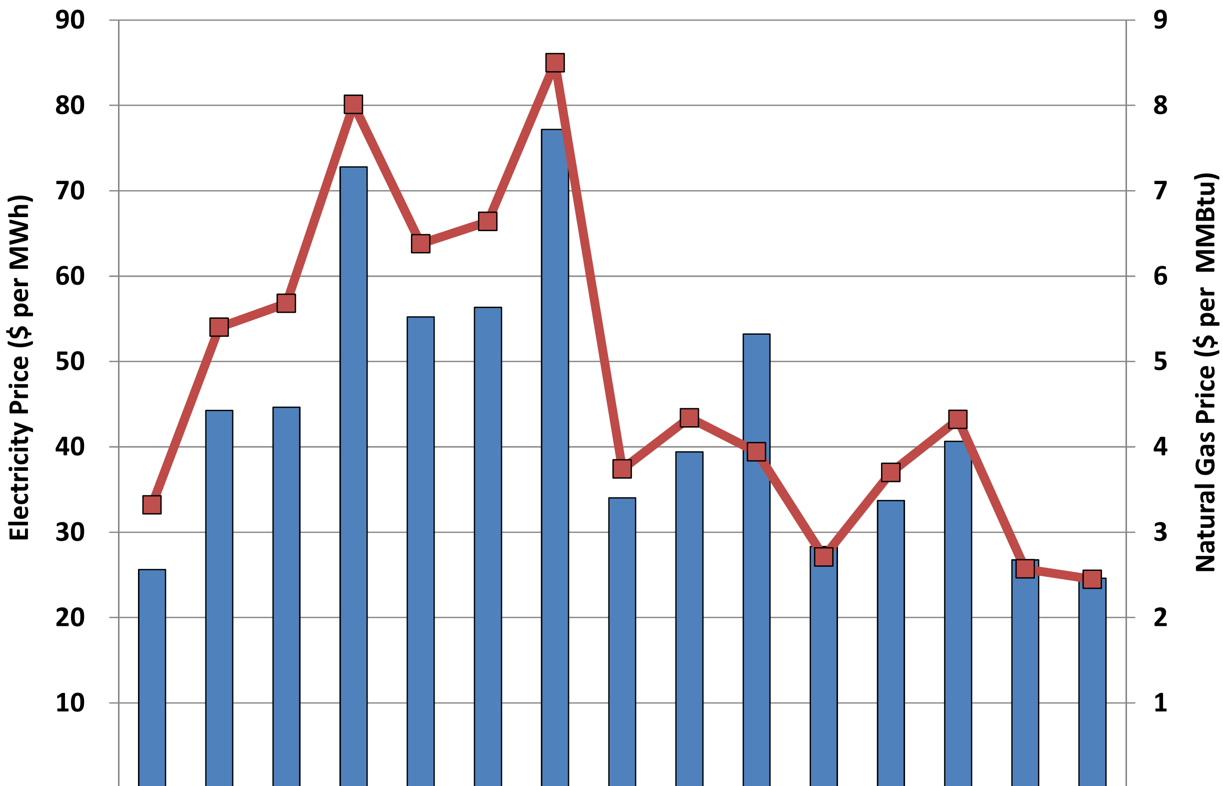


The largest component of the all-in price is the energy cost. The figure above indicates that natural gas prices continued to be a primary driver of energy prices. This correlation is expected in a well-functioning, competitive market because fuel costs represent the majority of most suppliers' marginal production costs. Since suppliers in a competitive market have an incentive to offer supply at marginal costs and natural gas is the most widely-used fuel in ERCOT, changes in natural gas prices should translate to comparable changes in offer prices. The average natural gas price in 2016 was $2.45 per MMBtu, down approximately 5 percent from the 2015 average price of $2.57 per MMBtu. ERCOT average real-time energy prices were also down 8 percent, declining from $26.77 in 2015 to $24.62 in 2016. The all-in price in 2016 included small contributions from ERCOT's energy price adders - $0.27 per MWh from the operating reserve adder and $0.13 per MWh from the reliability adder. The highest monthly average operating reserve adder occurred in April; while the highest monthly average reliability adder occurred in September.

Finally, the other classes of costs continue to be a small portion of the all-in electricity price – ancillary services costs were $1.03 per MWh, down from $1.23 per MWh in 2015 because of reductions in natural gas prices and lower ancillary service requirements. Uplift costs accounted for $0.74 per MWh of the all-in electricity price, similar to $0.69 per MWh in 2015.

Figure 2 below provides additional historic perspective on the ERCOT average real-time energy prices as compared to the average natural gas prices in each year from 2002 through 2016.

Figure 2: ERCOT Historic Real-Time Energy and Natural Gas Prices



	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
ERCOT	25.64	44.26	44.64	72.79	55.22	56.35	77.19	34.03	39.40	53.23	28.33	33.71	40.64	26.77	24.62
Gas	3.32	5.40	5.68	8.01	6.38	6.64	8.50	3.74	4.34	3.94	2.71	3.70	4.32	2.57	2.45

Like Figure 1, Figure 2 shows the close correlation between the average real-time energy price in ERCOT and the average natural gas price. Such relationship is consistent with expectations in ERCOT where natural-gas generators predominate and tend to set the marginal price. A noticeable exception occurred in 2011, when energy prices were affected by scarcity conditions.

Energy prices vary across the ERCOT market because of congestion costs that are incurred as power is delivered over the network. Figure 3 shows the monthly load-weighted average prices in the four geographic ERCOT zones during 2016 and 2015. These prices are calculated by weighting the real-time energy price for each interval and each zone by the total load in that interval. Load-weighted 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 or other forward contract prices.

Figure 3: Average Real-Time Energy Market Prices by Zone

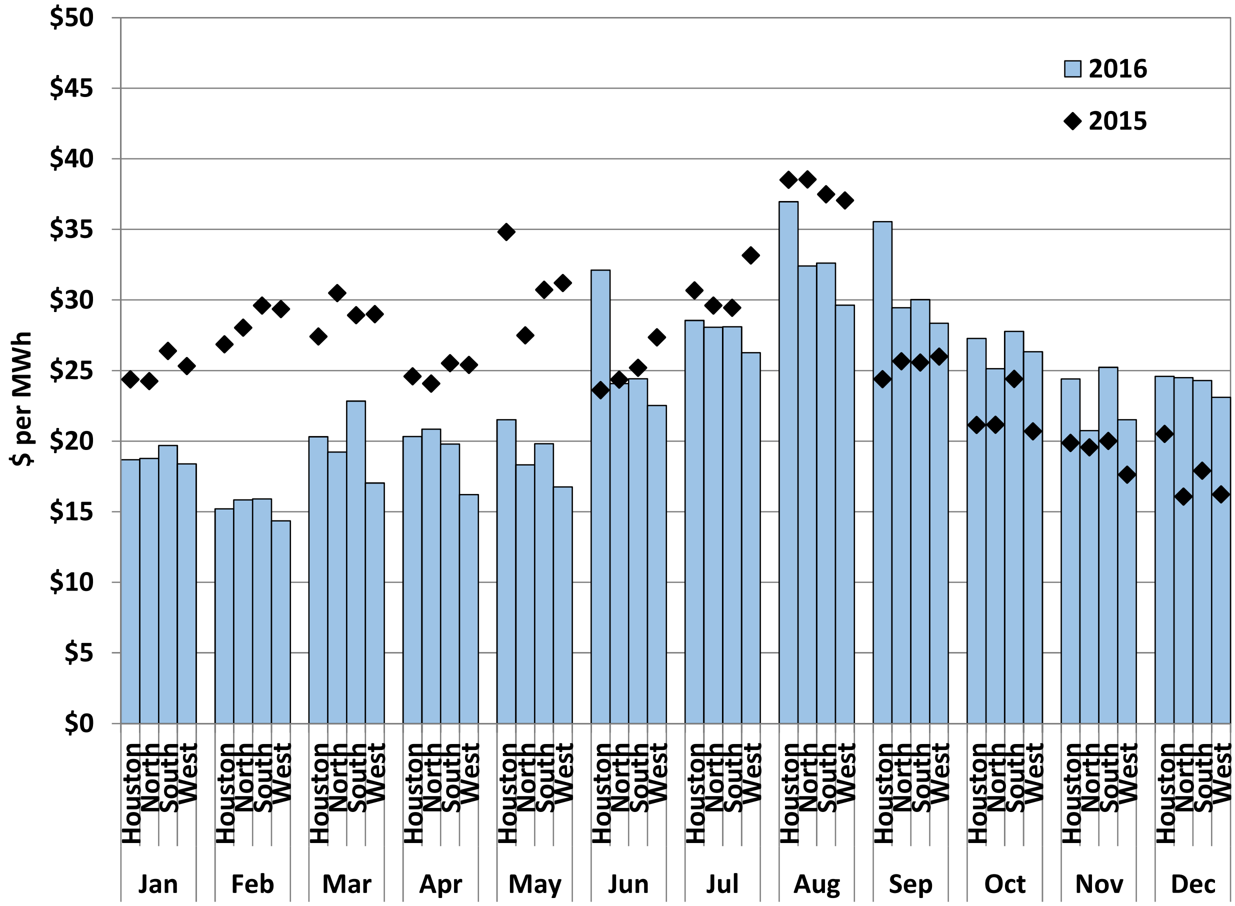


Table 1 provides the annual load-weighted average price for each zone for the past six years, and includes the annual average natural gas price for reference.

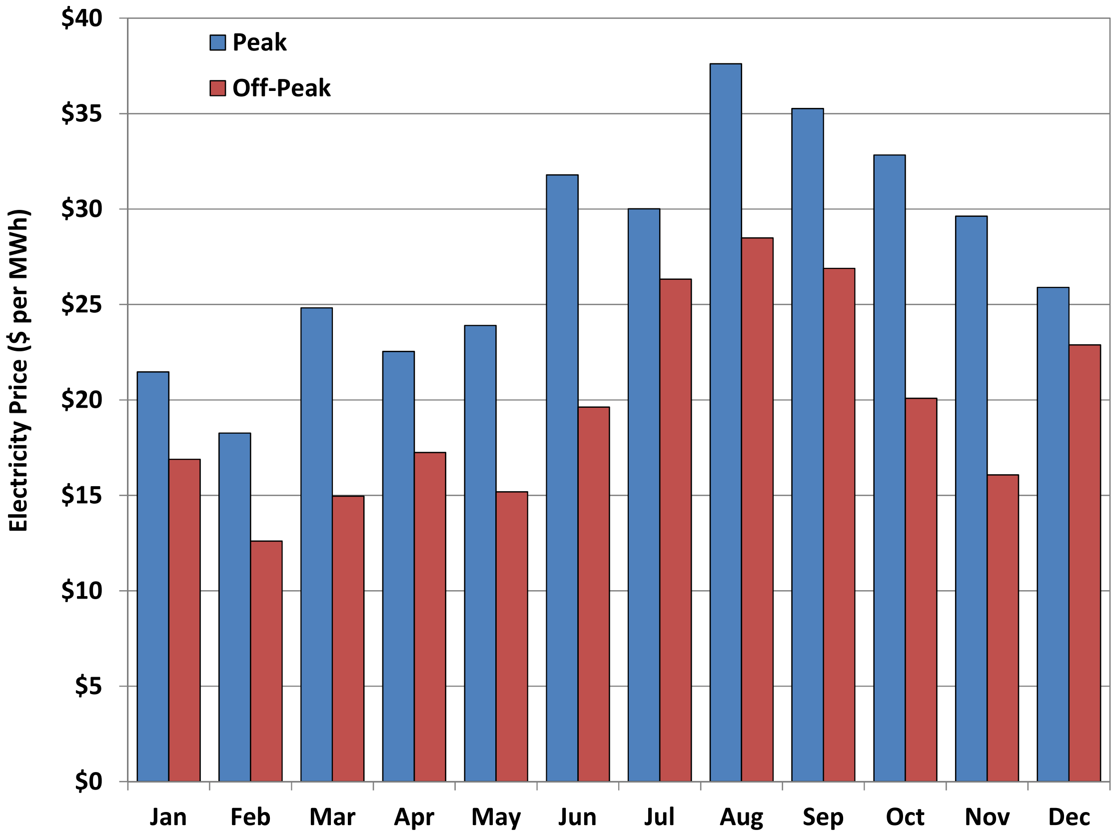
Table 1: Average Annual Real-Time Energy Market Prices by Zone

	2011	2012	2013	2014	2015	2016
ERCOT	**$53.23**	**$28.33**	**$33.71**	**$40.64**	**$26.77**	**$24.62**
Houston	$52.40	$27.04	$33.63	$39.60	$26.91	$26.33
North	$54.24	$27.57	$32.74	$40.05	$26.36	$23.84
South	$54.32	$27.86	$33.88	$41.52	$27.18	$24.78
West	$46.87	$38.24	$37.99	$43.58	$26.83	$22.05
Natural Gas ($/MMBtu)	$3.94	$2.71	$3.70	$4.32	$2.57	$2.45

The zonal prices in 2016 show greater disparities than 2015 because of congestion in the West and Houston. Prior to 2012, average prices in the West zone were lower than average ERCOT wide prices. This changed in 2012 when demand in the West rose because of increased oil and gas production activity. The West zone average annual price remained higher than the ERCOT average until 2016 when increased congestion caused by high levels of wind output in the West pushed the average prices in the West lower than other zones. Additionally, transmission congestion related to power flows in Houston caused that zone to exhibit the highest average prices and reduced the average prices in the North zone.

Figure 4 shows the load-weighted average real-time prices in ERCOT for the categories of Peak and Off-Peak for each month in 2016. The Peak block includes hours ending 7-22 on weekdays; the Off-Peak block includes hours ending 1-6 and 23-24 on weekdays and all hours on weekends. These pricing blocks align with the categories traded on the InterContinental Exchange (ICE) forward markets.

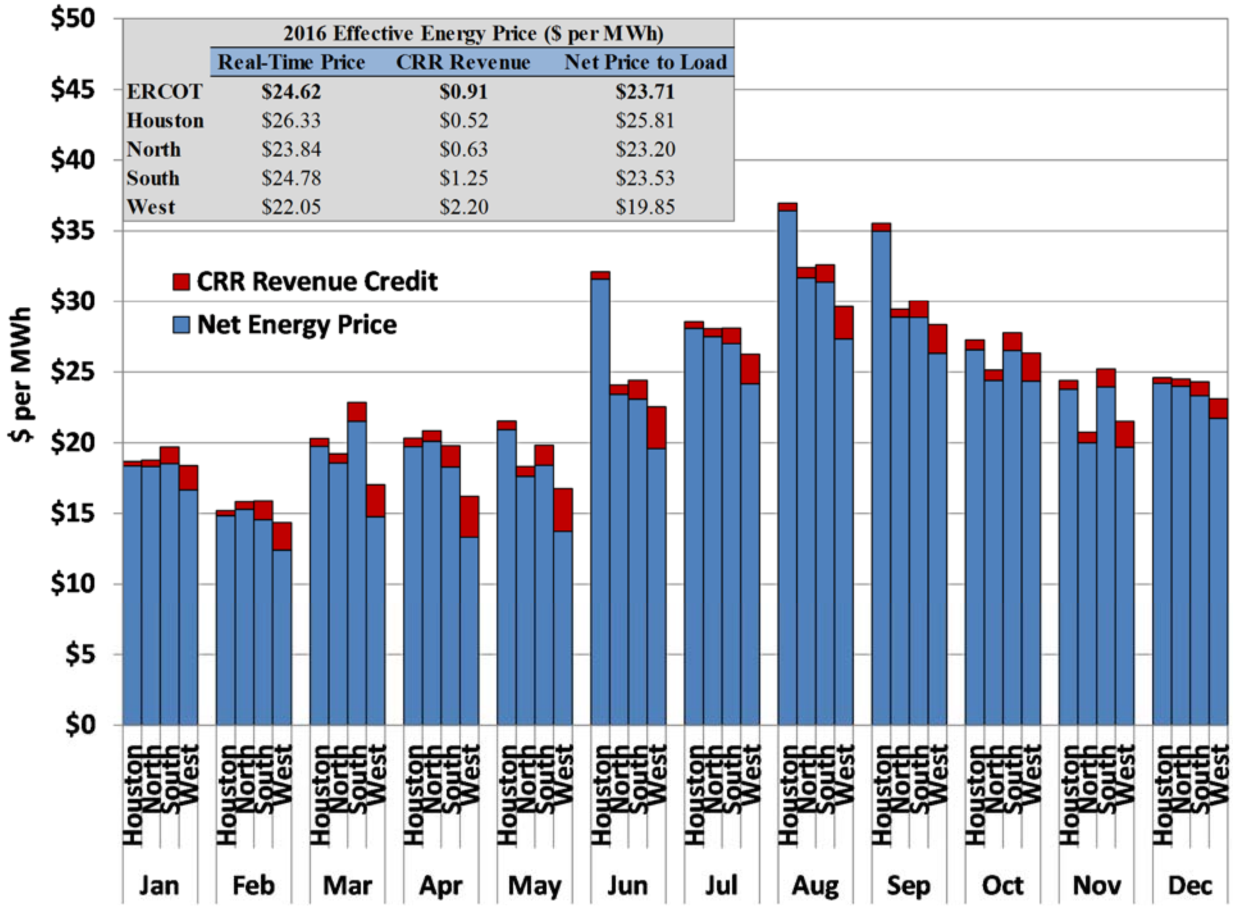
Figure 4: Peak and Off-Peak Pricing



As would be expected, Peak hours were higher priced than Off-Peak hours for every month in 2016. The monthly difference ranged from a minimum of $3.00 per MWh in December to a maximum of $13.55 per MWh in November. The average difference between monthly Peak and Off-Peak pricing was $8 per MWh.

Congestion Revenue Right (CRR) Auction Revenues are distributed to Qualified Scheduling Entities (QSEs) representing load, based on a zonal and ERCOT-wide monthly load-ratio share. The CRR Auction Revenues have the effect of reducing the total cost to serve load borne by a QSE. Figure 5 below shows the effect that this reduction has on a monthly basis, by zone.

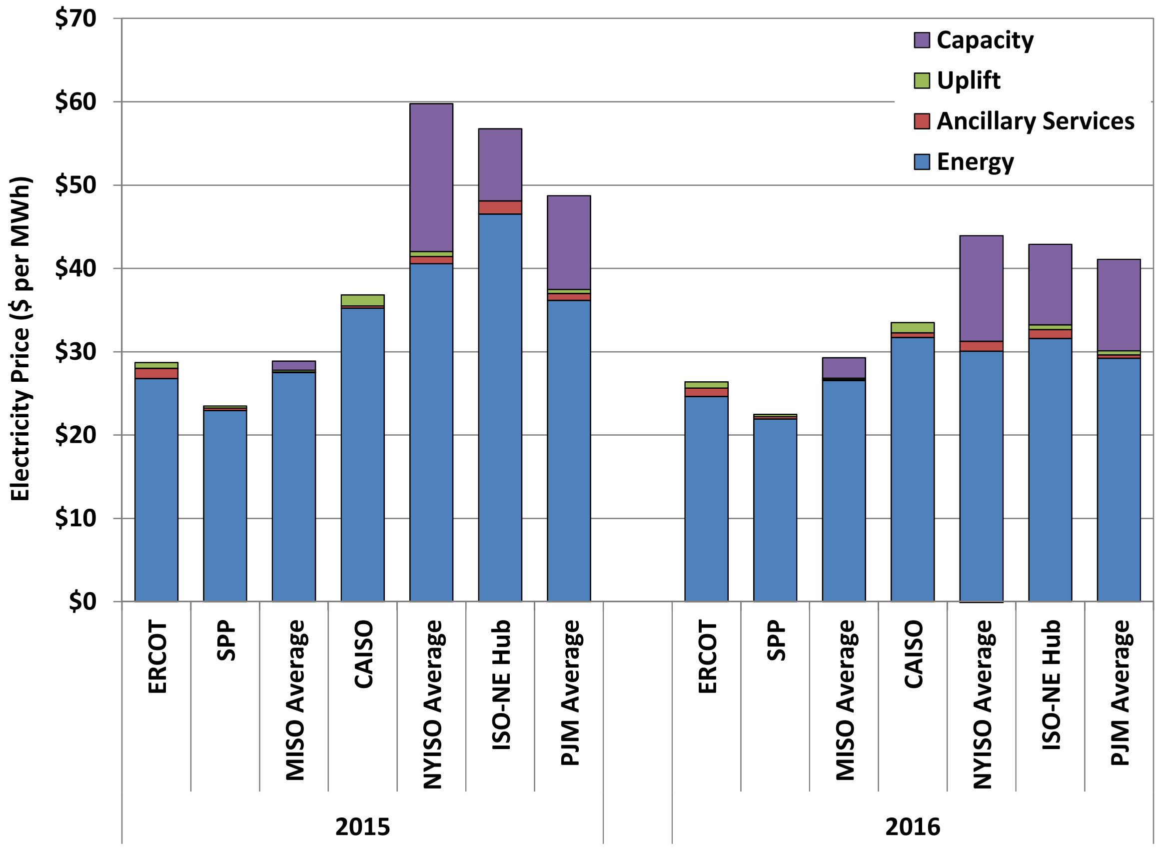
Figure 5: Effective Real-Time Energy Market Prices



With the CRR Auction Revenue offset included, the ERCOT-wide load-weighted average price was reduced by $0.91 per MWh to $23.71 per MWh in 2016. Focusing on zonal differences, a smaller credit in Houston relative to the ERCOT-wide CRR Auction Revenue credit and a larger credit in the West resulted in a net price difference between the two zones being even higher.

To provide additional perspective on the outcomes in the ERCOT market, Figure 6, below compares the all-in price in ERCOT with other organized electricity markets in the United States: Southwest Power Pool (SPP), Midcontinent ISO (MISO), California ISO, New York ISO, ISO New England, and the Pennsylvania-New Jersey-Maryland (PJM) Interconnection.

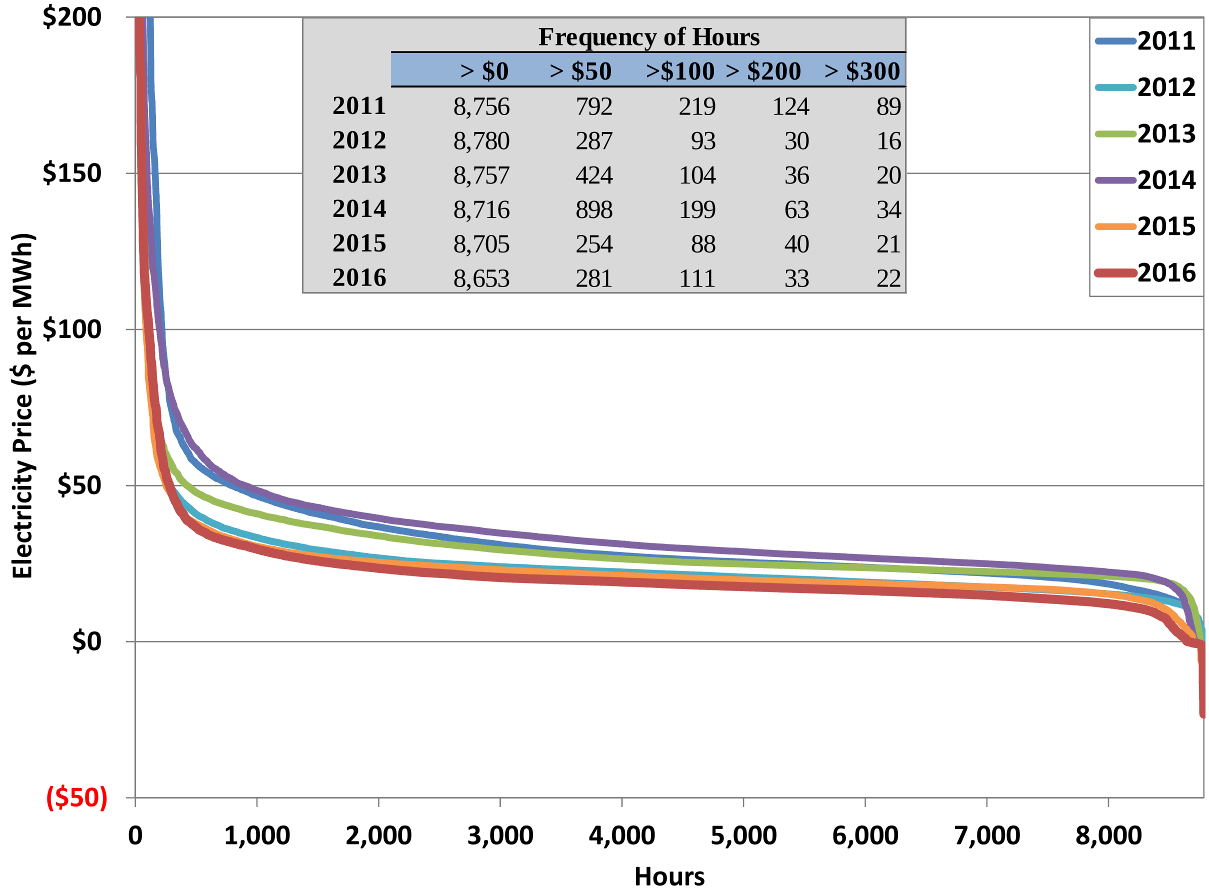
Figure 6: Comparison of All-in Prices Across Markets



The figure reports each market's average cost (per MWh of load) for energy, ancillary services (reserves and regulation), capacity markets (if applicable), and uplift. Figure 6 shows that, with the exception of a small increase in MISO, all-in prices were lower across U.S. markets in 2016. This highlights the pervasive effects of much lower natural gas prices across the nation.

Figure 7 below shows price duration curves for the ERCOT energy market in each year from 2011 to 2016. 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 ERCOT average prices derived by load weighting the zonal settlement point prices.

Figure 7: ERCOT Price Duration Curve



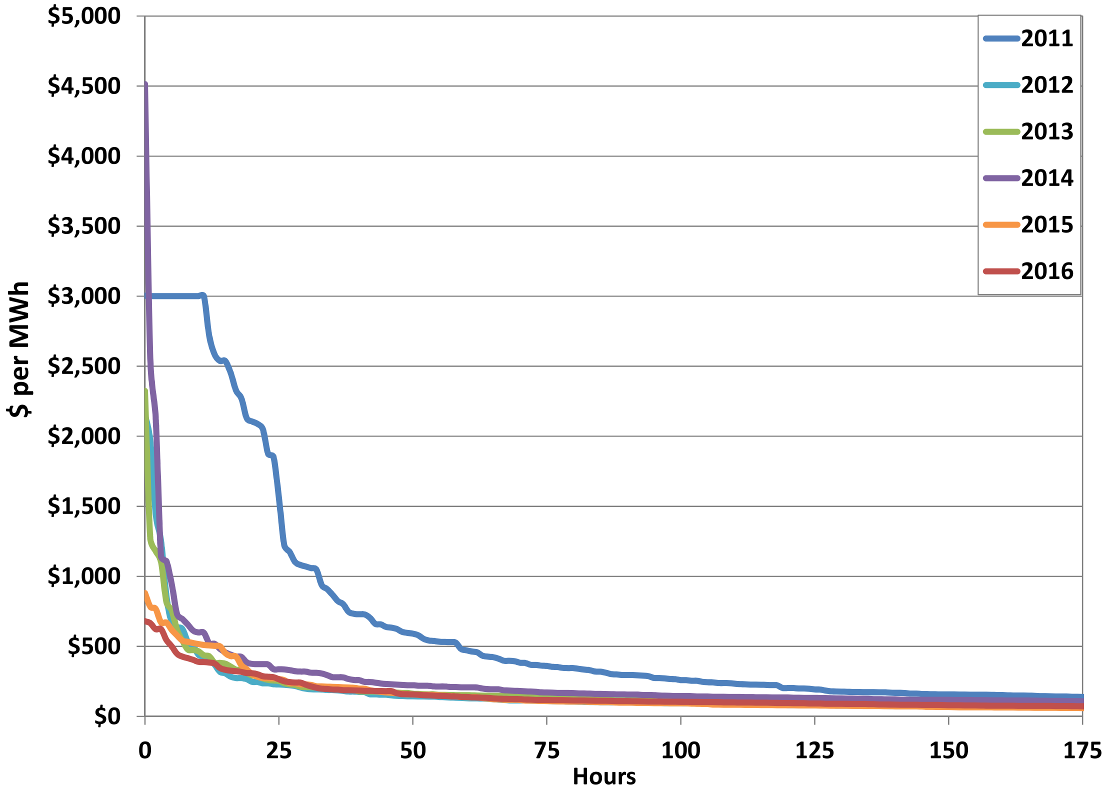
Frequency of Hours

	> $0	> $50	>$100	> $200	> $300
2011	8,756	792	219	124	89
2012	8,780	287	93	30	16
2013	8,757	424	104	36	20
2014	8,716	898	199	63	34
2015	8,705	254	88	40	21
2016	8,653	281	111	33	22

The number of hours with prices less than zero has been increasing in the past five years. In 2016, there were 131 hours of prices at or below zero, compared with 55 in 2015 and 44 in 2014. Negative ERCOT-wide prices may occur when wind is the marginal generation. More installed wind generation and additional transmission infrastructure has led to increased occurrences of negative prices.

To see where the prices during 2016 diverged from prior years, Figure 8 compares prices for the highest-priced two percent of hours in each year. In 2011, energy prices for the top 100 hours were significantly higher. These higher prices were due to high loads leading to more shortage conditions. Although the peak load in 2011 was exceeded in 2015 and 2016, generation additions during the intervening years have meant that shortage conditions continue to be rare.

Figure 8: ERCOT Price Duration Curve – Top 2% of Hours



To better observe the effect of the highest-priced hours on the average real-time energy price, the following analysis focuses on the frequency of price spikes in the real-time energy market. For this analysis, price spikes are defined as intervals when the load-weighted average energy price in ERCOT is greater than 18 MMBtu per MWh multiplied by the prevailing natural gas price. Prices at this level typically exceed the marginal costs of virtually all on-line generators in ERCOT.

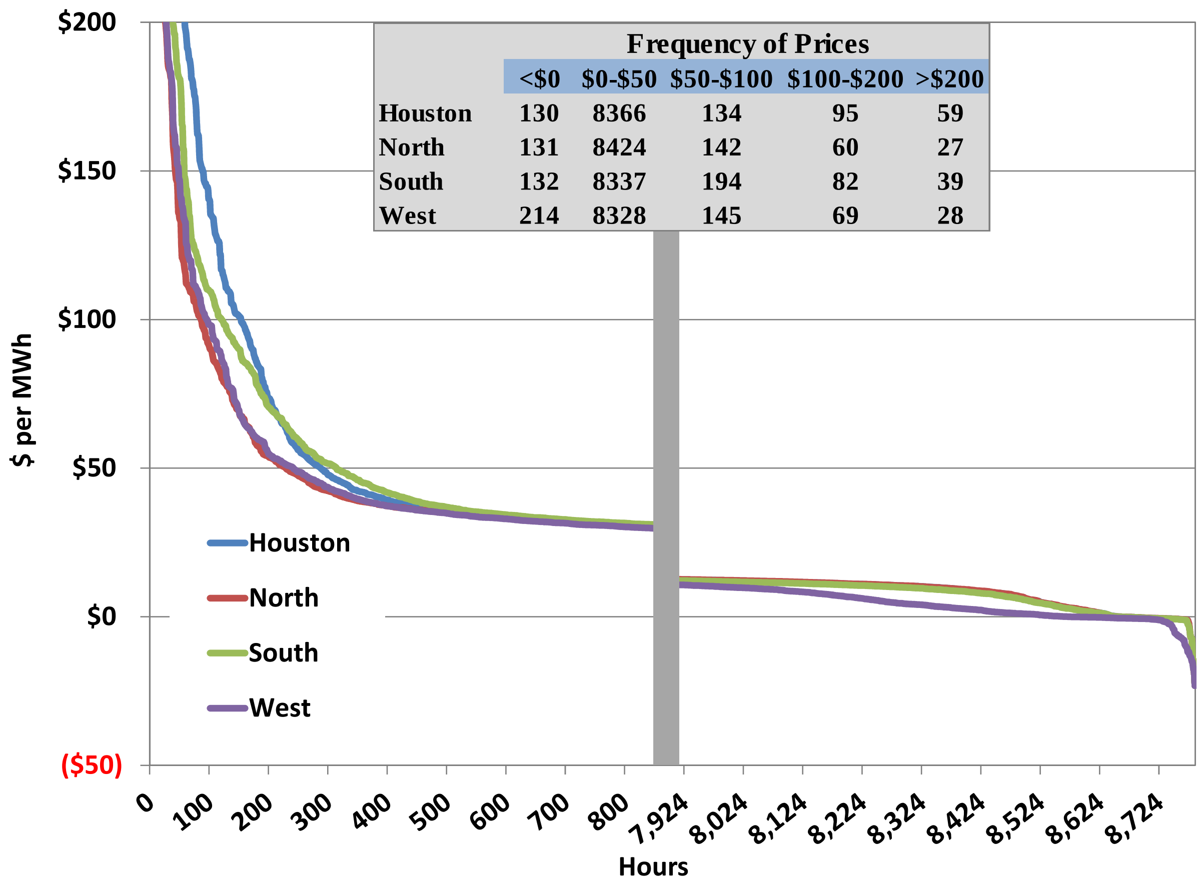
Table 2: Number and Impacts of Price Spikes on Average Real-Time Energy Prices

Year	Spikes Per Month	Magnitude (per MWh)	Price Impact
2011	83	$14.09	48%
2012	94	$3.63	16%
2013	54	$3.43	12%
2014	74	$5.28	16%
2015	89	$3.35	16%
2016	99	$3.53	19%

The overall impact of price spikes in 2016 was $3.53 per MWh. This result is generally consistent with the pricing impact of price spikes in past years. Of this price spike impact, $0.24 per MWh was due to the effects of the operating reserve adder.

To depict how real-time energy prices vary by hour in each zone, Figure 9 shows the top and bottom 10 percent of the hourly average price duration curve in 2016 for the four zones.

Figure 9: Zonal Price Duration Curves

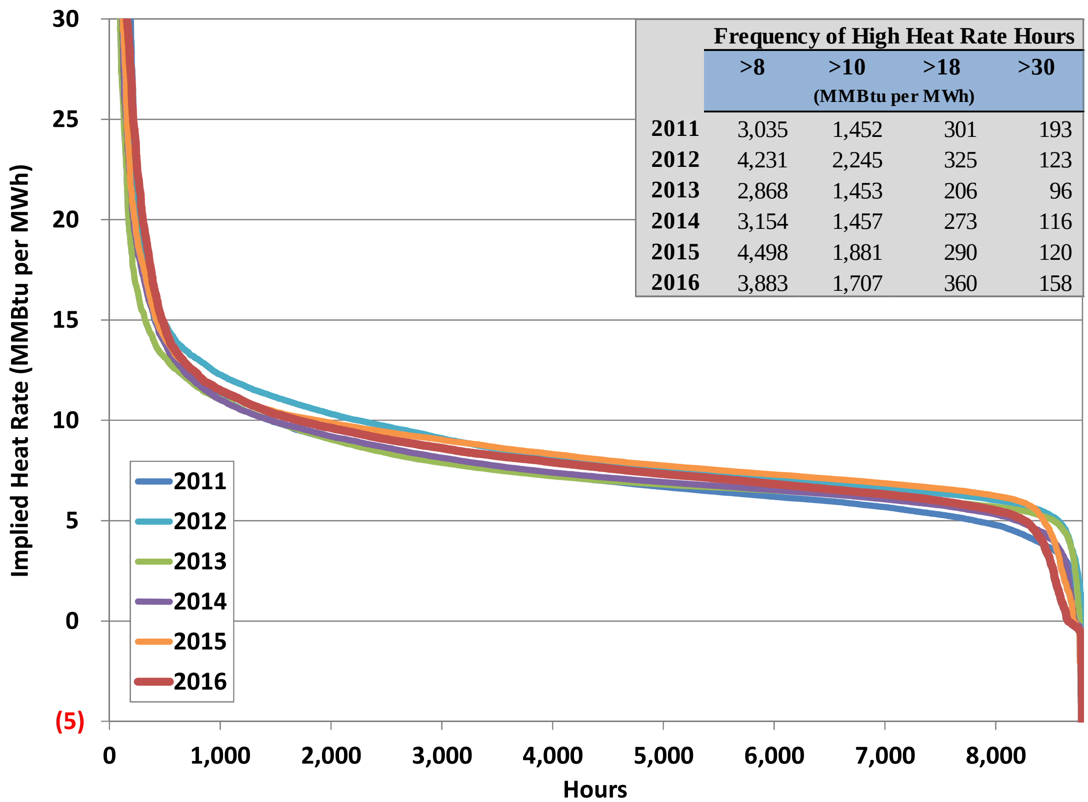


Negative prices occurred more frequently in 2016 for all zones and the West zone continued to experience more negative prices in 2016 than the other zones. Between 2012 and 2015 there had been a general trend toward fewer negative price intervals in the West zone as transmission additions reduced the frequency of negative West zone prices caused by transmission congestion during times of high wind output. This trend reversed in 2016 with 214 hours of negative prices in the West zone, compared to 121 hours in 2015. Negative prices in the other zones also occurred more frequently in 2016. The higher frequency of prices greater than $50 per MWh in the Houston and South zones is explained by North to Houston congestion, which had higher impacts than in 2015. More details about the transmission constraints influencing zonal energy prices are provided in Section III: Transmission Congestion and Congestion Revenue Rights.

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 summarize the changes in energy price that were related to these other factors, an "implied heat rate" is calculated by dividing the real-time energy price by the natural gas price. Figure 10 and Figure 11 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.

Figure 10: Implied Heat Rate Duration Curve – All Hours

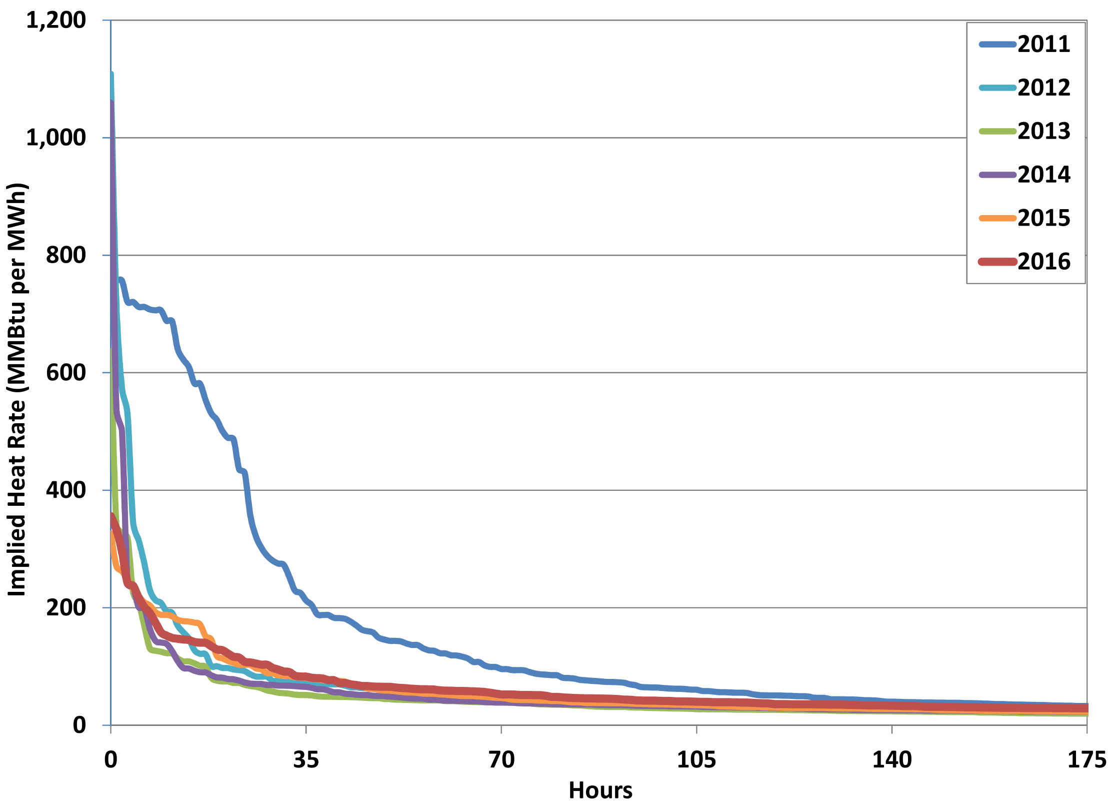


	Frequency of High Heat Rate Hours			
	>8	>10	>18	>30
	(MMBtu per MWh)			
2011	3,035	1,452	301	193
2012	4,231	2,245	325	123
2013	2,868	1,453	206	96
2014	3,154	1,457	273	116
2015	4,498	1,881	290	120
2016	3,883	1,707	360	158

Implied heat rates in 2016 were similar to those in 2015. This can be explained by the very low natural gas prices experienced in 2015 and 2016.

Figure 11 shows the implied marginal heat rates for the top two percent of hours for years 2011 through 2016. The implied heat rate duration curve for the top 2 percent of hours in 2016, closely resembles that for 2015. Among all years presented, 2011 remains an outlier.

Figure 11: Implied Heat Rate Duration Curve – Top 2 Percent of Hours



To further illustrate these differences, the next figure shows the implied marginal heat rates on a monthly basis in each of the ERCOT zones for 2015 and 2016. This figure is the fuel price-adjusted version of Figure 3 in the prior subsection. Implied heat rates in 2016 were very similar to those in 2015. This is expected given continued low natural gas prices and modest impacts from shortage conditions.

Figure 12: Monthly Average Implied Heat Rates

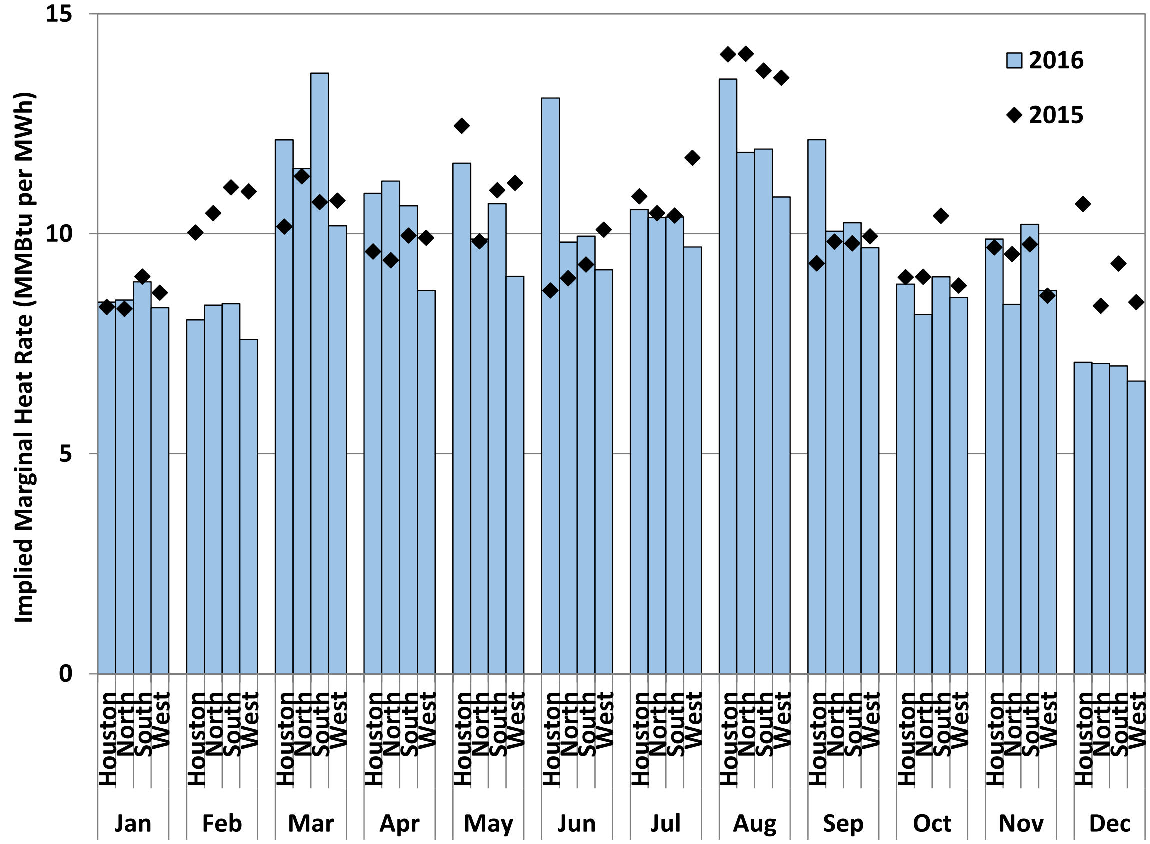


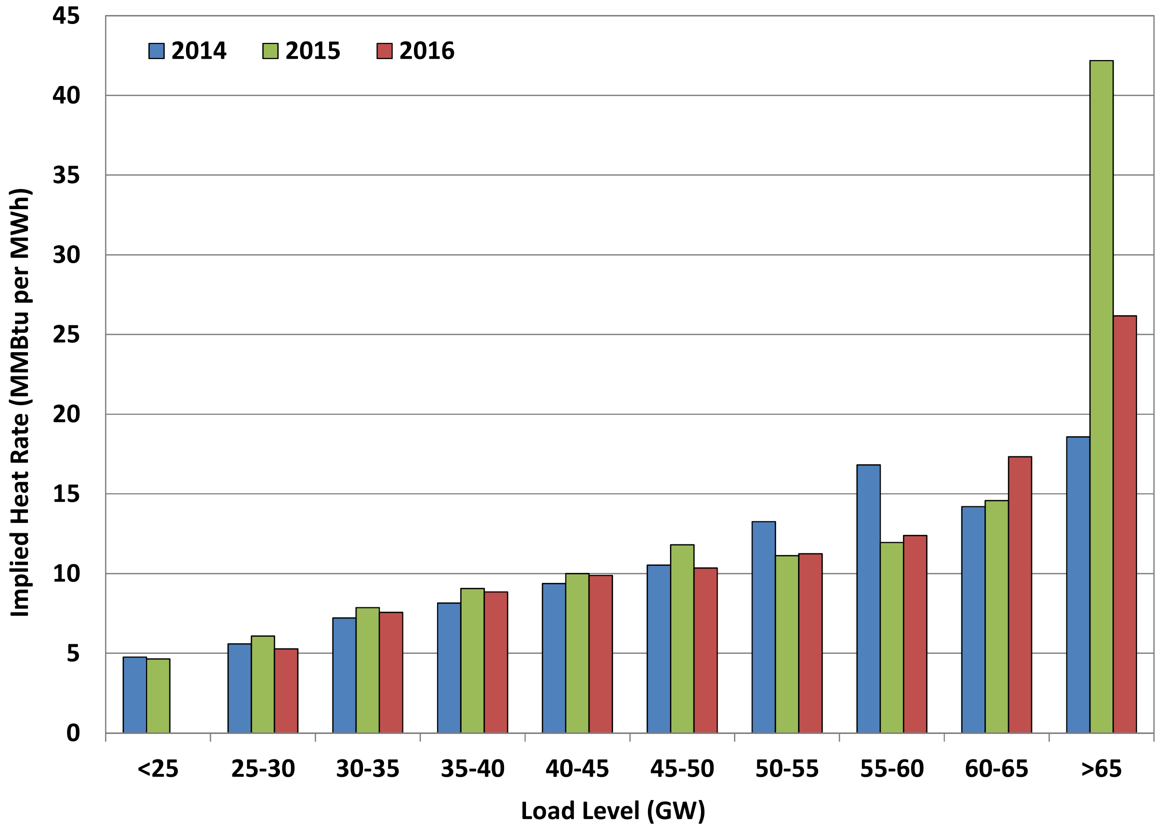
Table 3 displays the annual average implied heat rates by zone for 2011 through 2016. Adjusting for natural gas price influence, Table 3 shows that the annual, system-wide average implied heat rate decreased in 2016 compared to 2015. Zonal variations in the implied heat rate were greater in 2016, due to the increased influence of transmission congestion.

Table 3: Average Implied Heat Rates by Zone

	2011	2012	2013	2014	2015	2016
ERCOT	**13.5**	**10.5**	**9.1**	**9.4**	**10.4**	**10.1**
Houston	13.3	10.0	9.1	9.2	10.5	10.8
North	13.7	10.2	8.9	9.3	10.2	9.7
South	13.8	10.2	9.2	9.6	10.6	10.1
West	11.9	14.1	10.3	10.1	10.4	9.0
Natural Gas	$3.94	$2.71	$3.70	$4.32	$2.57	$2.45

The examination of implied heat rates from the real-time energy market concludes by evaluating them at various load levels. Figure 13 below provides the average implied heat rate at various system load levels from 2014 through 2016.

Figure 13: Implied Heat Rate and Load Relationship

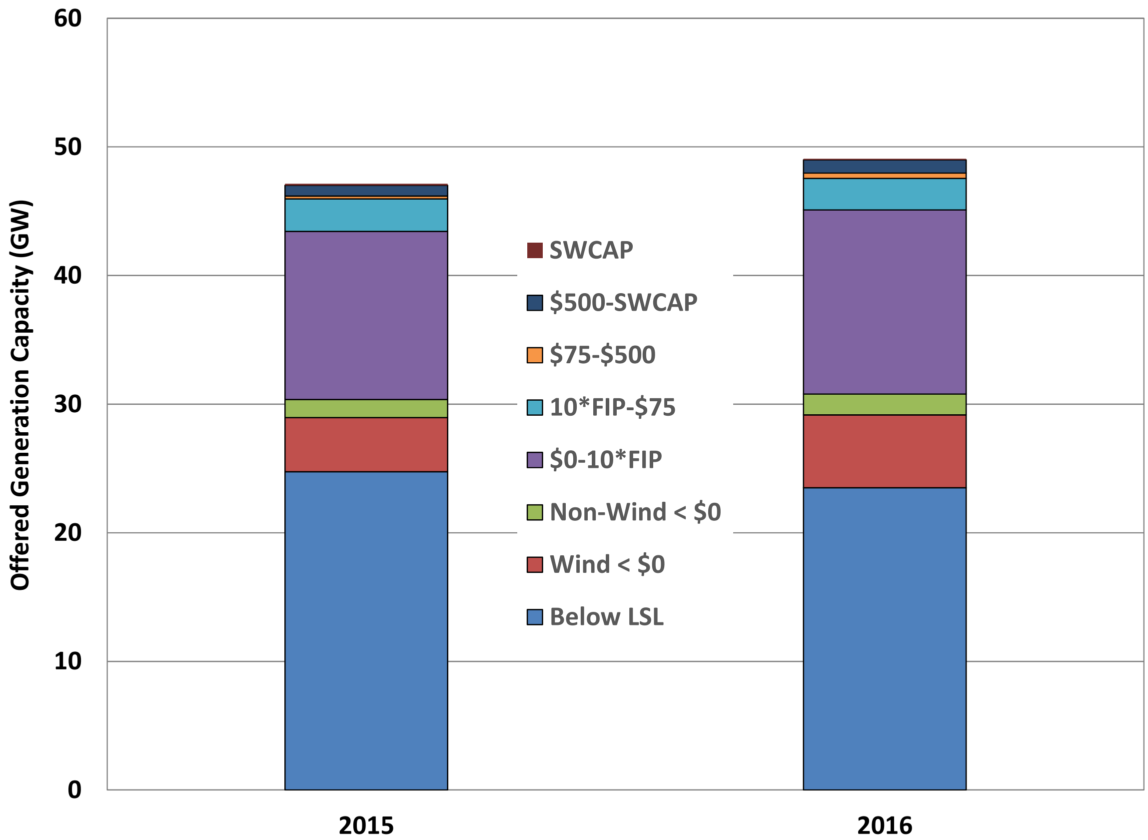


In a well-performing market, a clear positive relationship between these two variables is expected since resources with higher marginal costs are dispatched to serve higher loads. This relationship continues to exist in 2016.

C. Aggregated Offer Curves

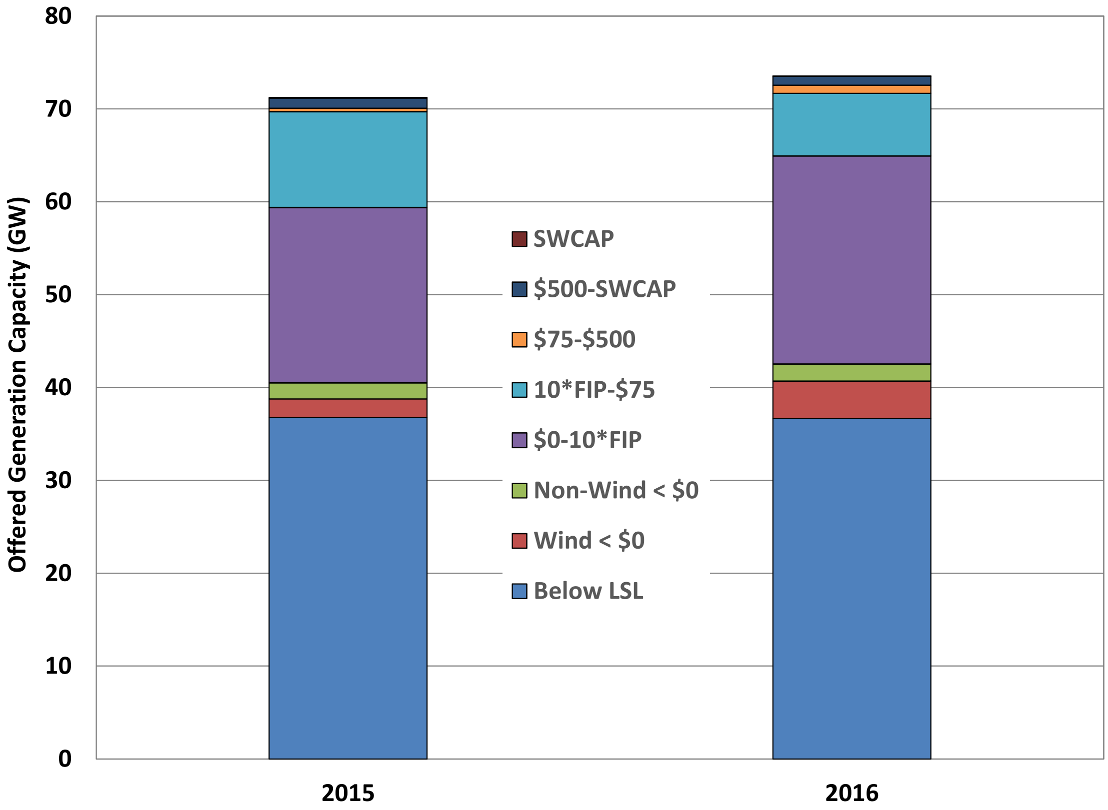
The next analysis compares the quantity and price of generation offered in 2016 to that offered in 2015. By averaging the amount of capacity offered at selected price levels, an aggregated offer stack can be assembled. Figure 14 provides the aggregated generator offer stacks for the entire year. Compared to 2015, more capacity was offered at lower prices in 2016. Specifically, continuing a trend from 2013, there was approximately 450 MW of additional capacity offered at prices less than zero. The greater capacity at prices less than zero was offered from wind generators (1,400 MW) and non-wind units (250 MW) with an off-setting decrease (1,200 MW) in capacity from below generators' low operating limits. There was an increase of approximately 1,250 MW of additional capacity offered in 2016 at prices between zero and ten multiplied by the daily natural gas price. The amount of capacity offered at prices between ten multiplied by the daily natural gas price and $75 per MWh decreased by 1,000 MW from 2015 to 2016. With a small, net increase (350 MW) to the quantities of generation offered at prices above $75 per MWh, the resulting average aggregated generation offer stack was roughly 2,000 MW greater in 2016 than in 2015.

Figure 14: Aggregated Generation Offer Stack – Annual



The next analysis provides a similar comparison focused on the summer season. As shown below in Figure 15, the changes in the aggregated offer stacks between the summer of 2015 and 2016 were similar to those just described. Comparing 2016 to 2015, there were approximately 700 MW additional capacity offered at prices less than zero, with a decrease of 900 MW of capacity below generators' low sustained limits (LSLs) and an increase of 1,600 MW in energy offered at prices less than zero but above the generators' LSLs. There was 1,900 MW more energy offered at prices between zero and ten multiplied by the daily natural gas price, but 350 MW less energy offered at prices between ten multiplied by the daily natural gas price and $75. With a small increase to the quantity of generation offered at prices above $75 per MWh, the resulting average aggregated generation offer stack for the summer season was approximately 2,400 MW greater than in 2015.

Figure 15: Aggregated Peak Hour Generation Offer Stack



D. ORDC Impacts and Prices During Shortage Conditions

The Operating Reserve Demand Curve (ORDC) is a scarcity pricing mechanism that reflects the loss of load probability (LOLP) at varying levels of operating reserves multiplied by the deemed value of lost load (VOLL).[14] Selected as an easier to implement alternative to real-time co-optimization of energy and ancillary services, the ORDC places an economic value on the reserves being provided, with separate pricing for online and offline reserves. The ORDC curves for 2016 are shown in Figure 16 below. The curves are determined in advance for four-hour blocks that vary across seasons. This depiction shows the breadth of distribution of the ORDC values across the year. The methodology leads to some large discontinuities between the curves applicable for adjacent time blocks. The largest such change occurs in the spring season between 5:59 a.m. and 6:00 a.m. where the value of the ORDC curve changes almost $1,200 per MWh. Once available reserve capacity drops to 2,000 MW price will rise to $9,000 per MWh for all the ORDC curves.

[14] At the September 12, 2013 Open Meeting, the PUCT Commissioners directed ERCOT to move forward with implementing ORDC, including setting the Value of Lost Load at $9,000

Figure 16: Seasonal Operating Reserve Demand Curves, by Four-Hour Blocks

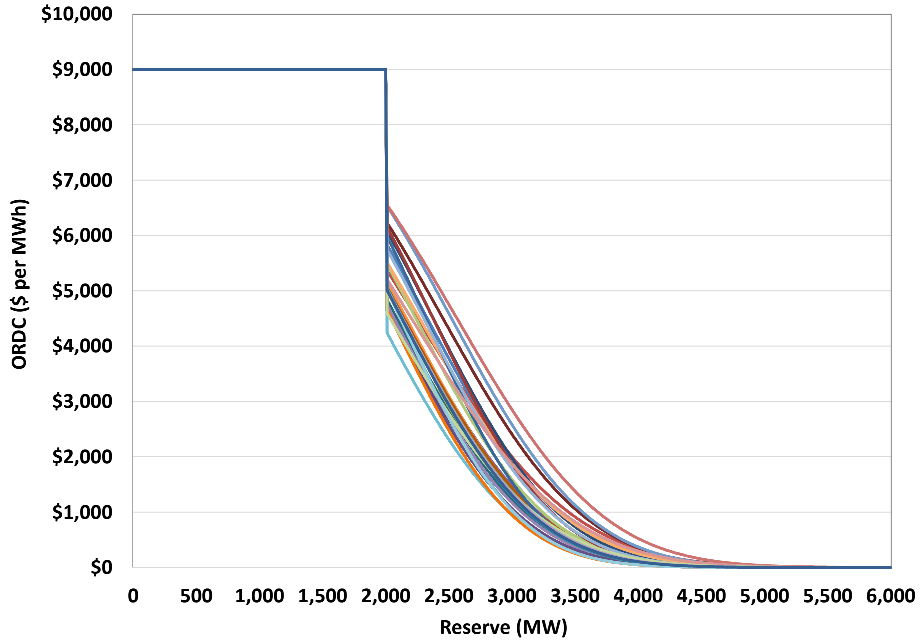
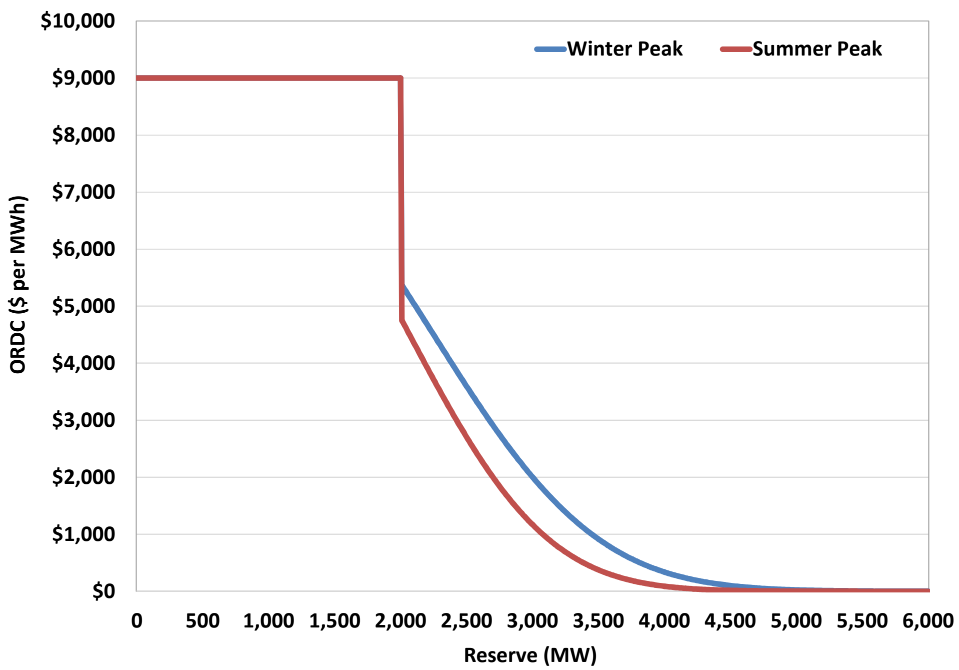


Figure 17 provides another depiction of the peak Operating Reserve Demand Curves applicable during winter and summer peak hours.

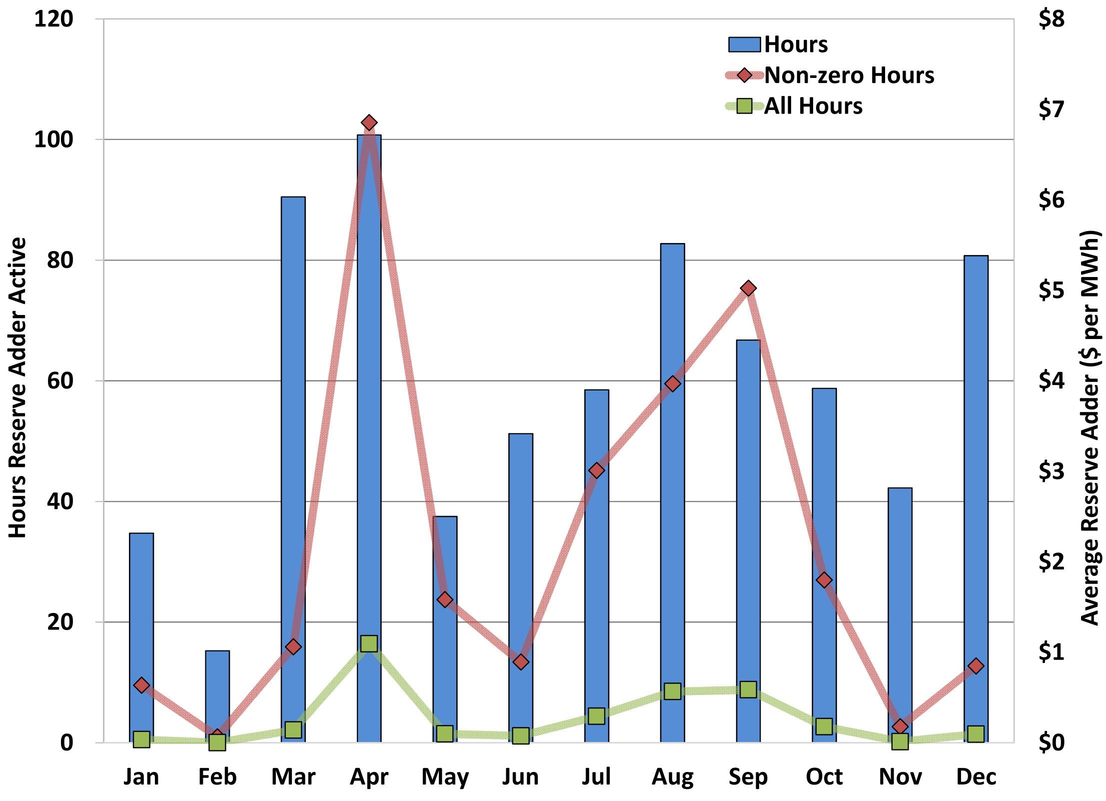
Figure 17: Winter and Summer Peak Operating Reserve Demand Curves



The following two analyses illustrate the contributions of the operating reserve adder and the reliability adder to shortage pricing. As described above, the contributions of the energy price adders were relatively small in 2016. The first of the two adders is the operating reserve adder, which is based on the loss of load probability, considering online and offline reserve levels, multiplied by the deemed value of lost load.

Figure 18 shows the number of hours in which the adder affected prices, and the average price effect in these hours and all hours. This figure shows that the operating reserve adder had the largest impacts during April and September, rather than during the summer months as observed in 2015. Overall, the operating reserve adder contributed $0.27 per MWh or 1 percent to the annual average real-time energy price of $24.62 per MWh. These results do not indicate that ORDC has been ineffective or that it should be modified. The effects of the operating reserve adder are expected to vary substantially from year to year, and to have the largest effects when poor supply conditions and unusually high load conditions occur together and result in sustained shortages.

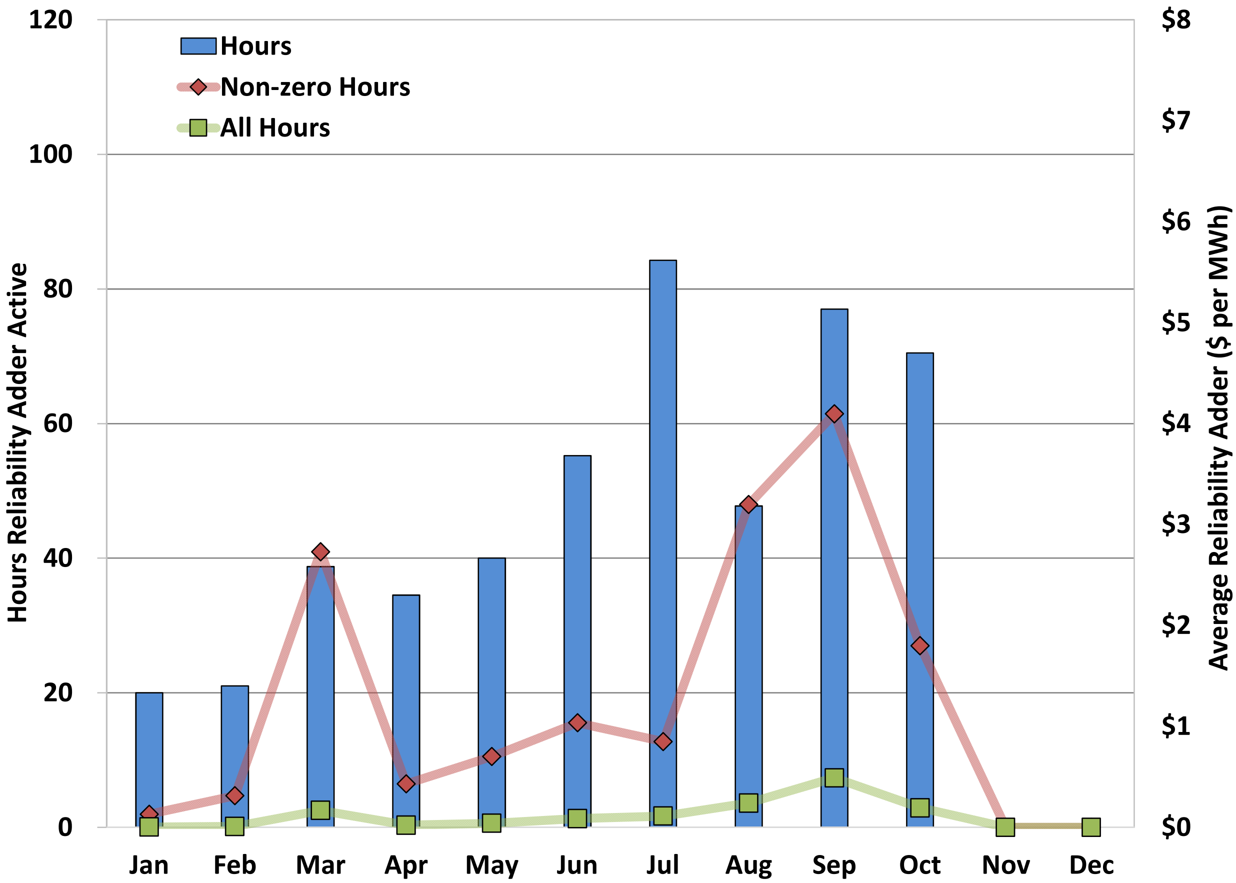
Figure 18: Average Operating Reserve Adder



In addition to the operating reserve adder, a reliability adder was implemented at the end of June 2015 and thus 2016 is the first full calendar year in which the effect of the adder can be observed. The reliability adder is intended to allow prices to reflect the costs of reliability actions taken by ERCOT, including RUC commitments and deployed load capacity. Absent this adder, prices will generally fall when these actions are taken.

Figure 19 below shows the impacts of the reliability adder. When averaged across the active hours, the largest price impacts of the reliability adder occurred in August and September. The reliability adder is zero in most hours. The reliability adder was non-zero for only 407 hours or 5 percent of the hours in 2016. There were no reliability adders in November and December. The contribution from the reliability adder to the annual average real-time energy price was $0.13 per MWh.

Figure 19: Average Reliability Adder

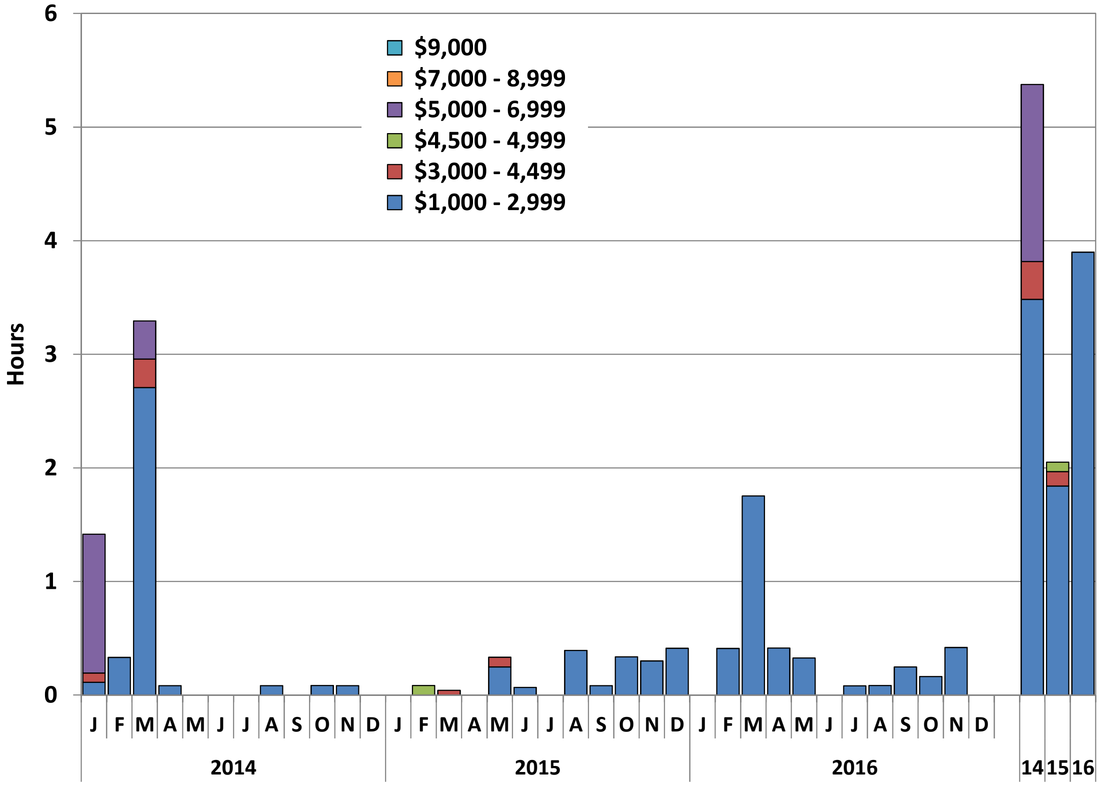


As an energy-only market, the ERCOT market relies heavily on high real-time prices that occur during shortage conditions. These prices provide key economic signals that provide incentives to build new resources and retain existing resources. However, the frequency and impacts of shortage pricing can vary substantially from year-to-year.

To summarize the shortage pricing that occurred from 2013 to 2016, Figure 20 below shows the aggregate amount of time when the real-time system-wide energy price, including the operating reserve adder and reliability adder during the times they were in effect, exceeded $1,000 per

MWh, by month. This figure shows that like in 2015, energy prices did not rise to the system-wide offer cap in 2016. In fact, prices in 2016 never exceeded $2,000 per MWh. Prices during 2015 exceeded $3,000 per MWh for a total of 0.21 hours, or less than 15 minutes. Prices during 2014 exceeded $3,000 per MWh for a total of 1.89 hours and were at the system-wide offer cap then in effect for 1.56 hours.

Figure 20: Duration of High Prices

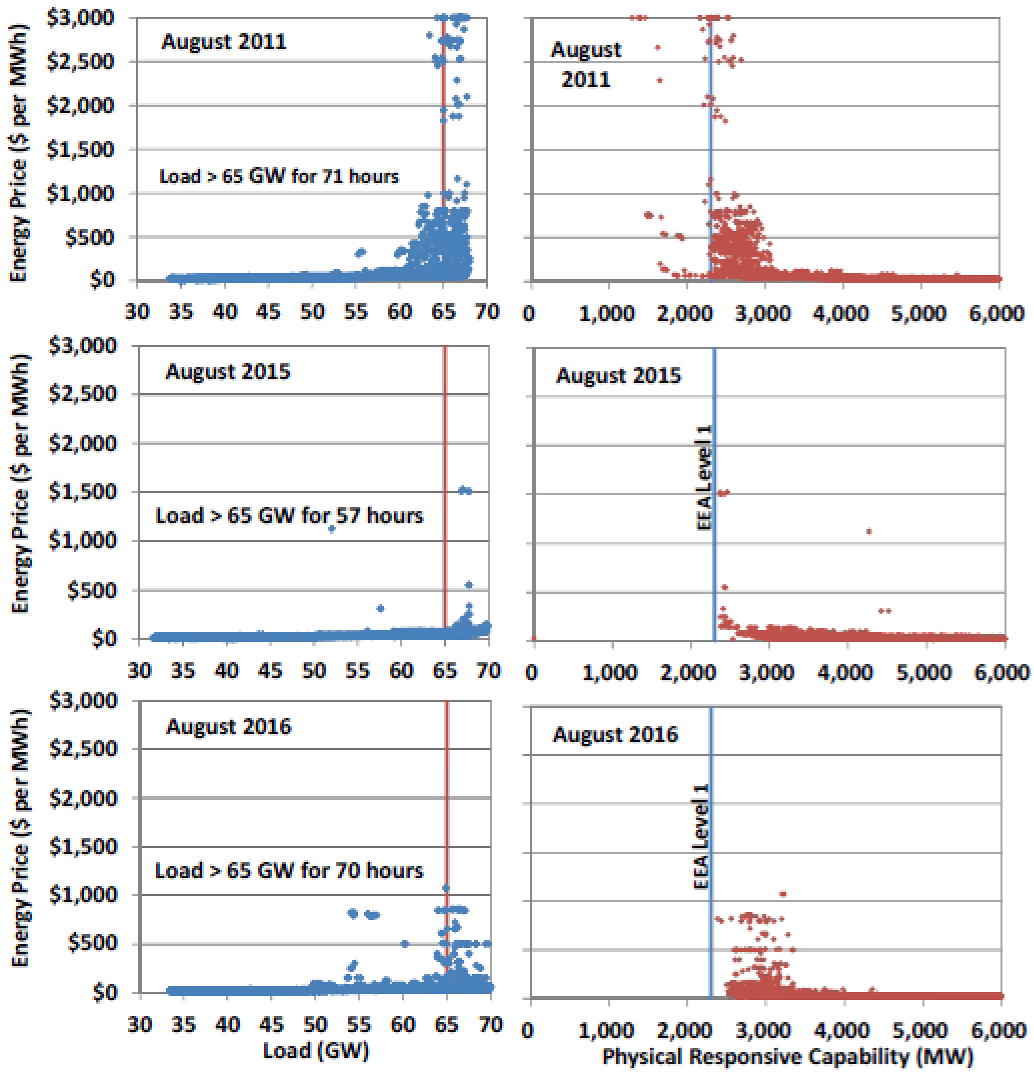


As a comparison, market prices cleared at the then in effect cap of $3000 per MWh for 28.44 hours in 2011. Extreme cold in February of 2011 and unusually hot and sustained summer temperatures led to much more frequent shortages in that year. Shortages in years with normal weather should be infrequent. As capacity margins fall, the frequency of shortages is likely to increase but will still vary substantially year-to-year.

Figure 21 provides a detailed comparison for the month of August in 2011, 2015, and 2016 showing load levels, required reserve levels, and real-time energy prices (excluding adders).[15] There were very few dispatch intervals when real-time energy prices approached $3,000 per MWh in 2015 and none in 2016, compared to the relatively high frequency in 2011.

[15] For purposes of Figure 21, the real-time energy prices excludes the operating reserve and reliability adders. This provides a better comparison between the years since the adders were not in effect in 2011.

Figure 21: Load, Reserves and Prices in August



The left side of Figure 21 shows the relationship between real-time energy price and load level for each dispatch interval for the months of August in the years 2011, 2015 and 2016. Load levels in August of 2016 were greater than 65 gigawatts (GW) for 70 hours, approaching the 71 hours observed in 2011. As previously discussed, a strong positive correlation between higher load and higher prices is expected in a well-functioning energy market, and our analysis shows such a relationship. Higher prices observed at non-peak load levels are typically due to transitory situations where there is insufficient generator ramping capability.

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 ERCOT declaring Energy Emergency Alert (EEA) Level 1 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 and the associated value of loss of load.

The right side of Figure 21 shows the relationship between real-time energy prices and the quantity of available operating reserves for each dispatch interval during August for the years 2011, 2015, and 2016. This figure shows a strong correlation between diminishing operating reserves and rising prices. Operating reserves did get within 100 MW of the minimum required level on one day in August 2016, but remained just above the level at which ERCOT would declare EEA Level 1. In contrast, there were numerous dispatch intervals in August 2011 when the minimum operating reserve level was approached or breached, and prices reached the system-wide offer cap in 17.4 hours.

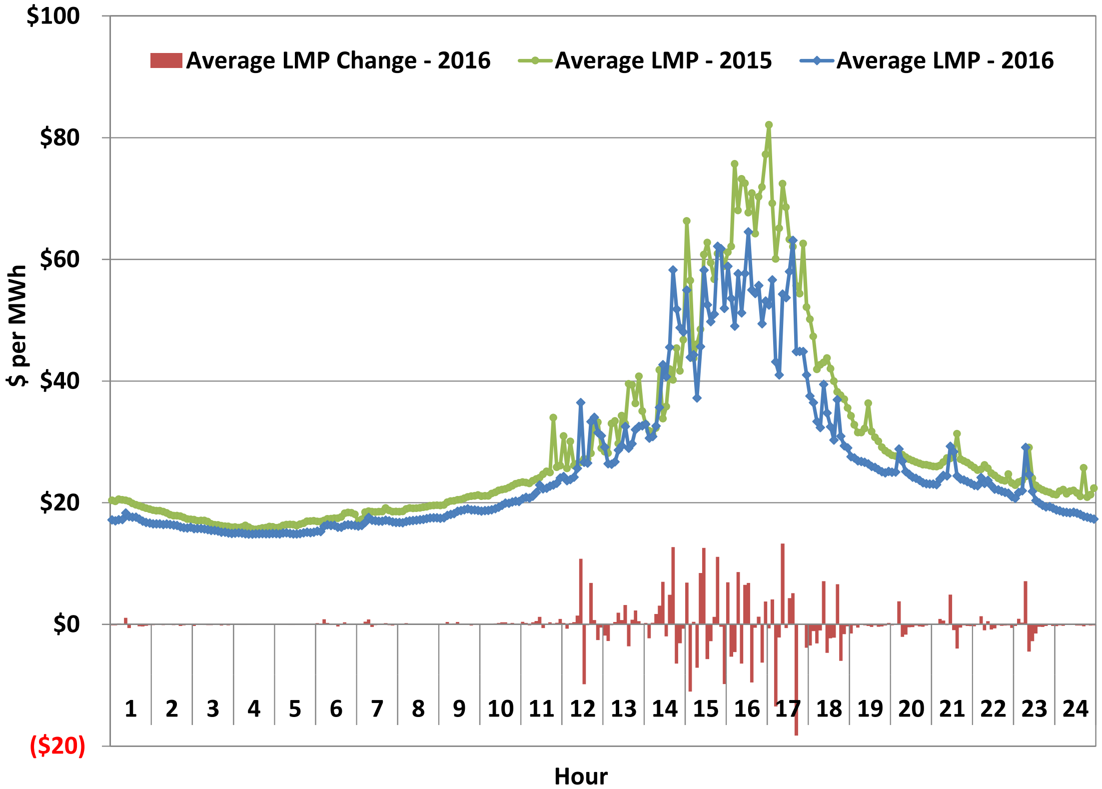
Concerns have been expressed that real-time prices were not higher during the infrequent intervals of low operating reserves in 2015 and 2016. A review of the ORDC parameters was undertaken in response to those concerns. There also have been changes to the reserve discount factor and how non-frequency responsive capacity is counted as reserve capacity. These changes, along with capacity additions and changes to the ancillary services requirements have all had an impact, some countervailing, on the levels of physical responsive capacity available during 2015 and 2016.

Prices in August 2016, even at lower operating reserve levels were set by generator offers. This is to be expected when operating reserve levels remain above minimum requirements.

E. Real-Time Price Volatility

Volatility in real-time wholesale electricity markets is expected because system load can change rapidly and the ability of supply to adjust can be restricted by physical limitations of the resources and the transmission network. Figure 22 below 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 real-time energy prices for 2016 are presented along with the magnitude of change in price for each five-minute interval. Average real-time energy prices from the same period in 2015 are also presented. Comparing average real-time energy prices for 2016 with those from 2015 shows greater volatility during peak hours.

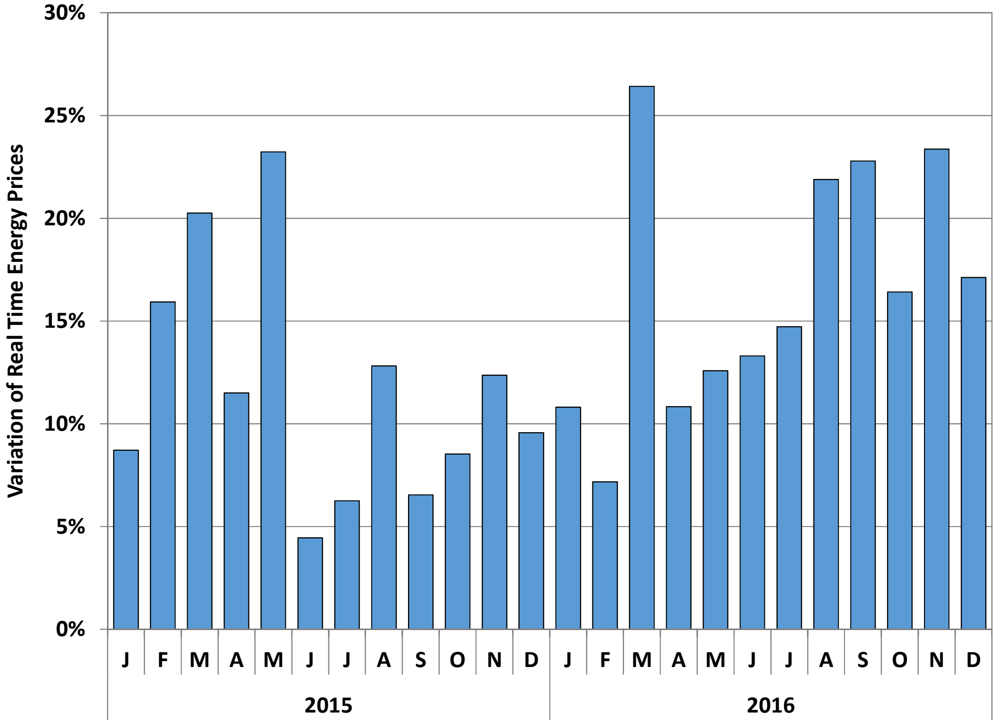
Figure 22: Real-Time Energy Price Volatility (May – August)



The average of the absolute value of changes in five-minute real-time energy prices during the months of May through August, expressed as a percent of average price, was 5.4 percent in 2016, compared to 5.0 percent in 2015. The percent of average price change from 2012 to 2014 ranged from 3.0 percent to 3.6 percent. In 2011, the absolute value of five-minute price changes was 6.2 percent.

Expanding the view of price volatility, Figure 23 below shows monthly average changes in five-minute real-time prices by month for 2016 and 2015. Without any prices at or close to the system-wide offer cap, the highest price variability occurs during spring and fall months when wind generation variations and load and wind generation forecast errors are the highest.

Figure 23: Monthly Price Variation



To show how the price volatility has varied by location, Table 4 below, shows the volatility of 15-minute settlement point prices for the four geographic zones for years 2012-2016.

Table 4: 15-Minute Price Changes as a Percentage of Annual Average Prices

	2012	2013	2014	2015	2016
Houston	13.0	14.8	14.7	13.4	20.8
South	13.1	15.4	15.2	14.6	19.9
North	13.9	13.7	14.1	11.9	15.5
West	19.4	17.2	15.4	12.9	16.8

These results show that price volatility is higher in 2016 than in the prior four years for all Load Zones, except the West Load Zone. Increased percentage variation in prices is expected given the lower annual average prices in 2016. While the West Load Zone had shown a continual decline in price volatility, an increase occurred in the West Load Zone in 2016, likely due to the increase in wind generation related congestion. Nonetheless, the volatility in the West Load Zone was lower than historically observed. The Load Zone with the highest volatility in 2016 was the Houston Load Zone. At greater than 20 percent, Houston Load Zone price volatility in 2016 was the highest of any Load Zone over the past five years. More costly and more frequent congestion related to power flows into the Houston area is the primary driver for the increased volatility.

II. Day-Ahead Market Performance

ERCOT's day-ahead market 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 allows 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. In addition to the purchase and sale of power, the day-ahead market also includes ancillary services and Point-to-Point (PTP) obligations. PTP obligations 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 for the ability to reliably flow on the transmission network, there are no operational obligations resulting from the day-ahead market. Day-ahead transactions are made for a variety of reasons, including satisfying the participant's own demand, managing risk by hedging the participant's exposure to real-time prices or congestion, or arbitraging with the real-time prices. 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, energy pricing outcomes from the day-ahead market are reviewed and convergence with real-time energy prices is examined. The volume of activity in the day-ahead market, including a discussion of PTP obligations, is also reviewed. This section concludes with a review of the ancillary service markets.

A. Day-Ahead Market Prices

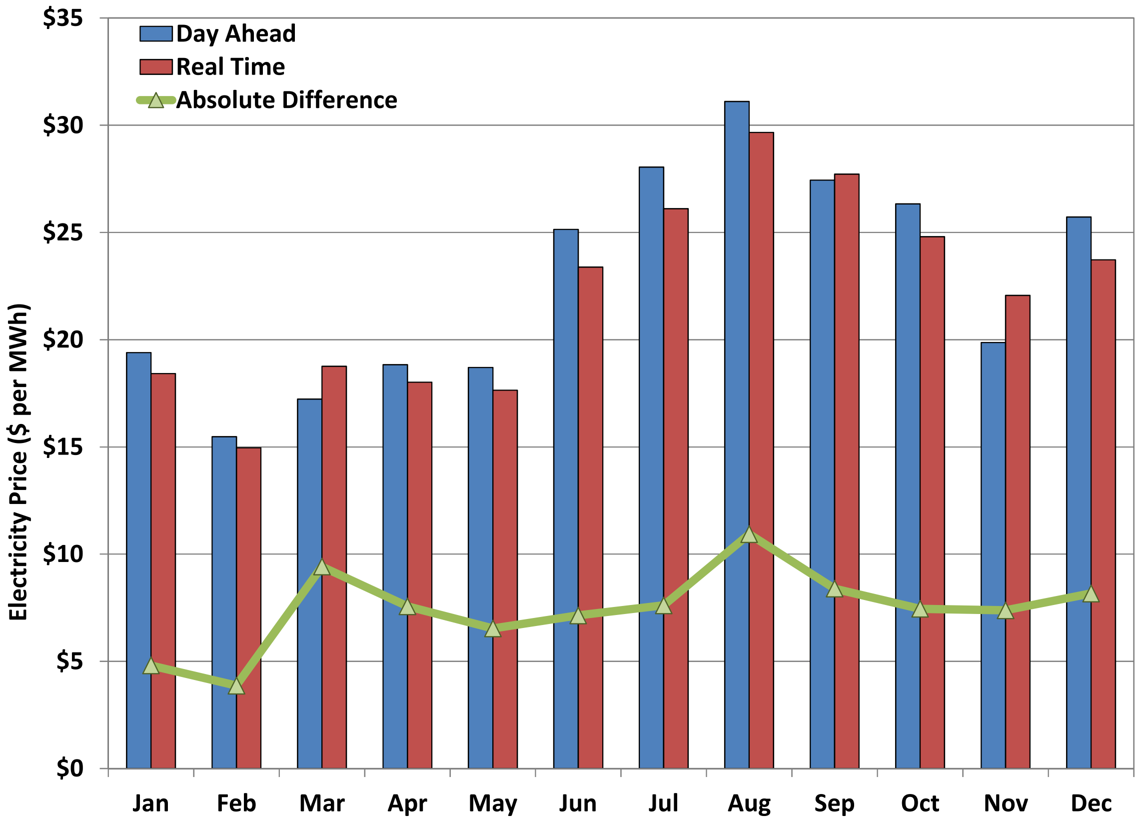
One indicator of market performance is the extent to which forward and real-time spot prices converge over time. Forward prices will converge with real-time prices when: (1) there are low barriers to shifting purchases and sales between the forward and real-time markets; and (2) sufficient information 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. This improves the convergence of forward and real-time prices, which should lead to improved commitment of resources needed to satisfy the system's real-time needs.

In this subsection, price convergence between the day-ahead and real-time markets is evaluated. This average price difference 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, the average of the absolute value of the difference between the day-ahead and real-time price are calculated 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.[16]

Figure 24 summarizes the price convergence between the day-ahead and real-time markets, by month in 2016. Price convergence was good in 2016. Day-ahead prices averaged $23 per MWh in 2016 compared to an average real-time price of $22 per MWh.[17] This day-ahead 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 market and higher for generators selling day ahead. The higher risk for generators is associated with the potential of incurring a forced outage and having to buy back energy at real-time prices. This explains why the highest premiums occurred during the summer months in 2016 with the highest relative demand and highest prices.

Figure 24: Convergence Between Day-Ahead and Real-Time Energy Prices



16 For instance, if day-ahead prices are $30 per MWh on two consecutive days while real-time prices are $20 per MWh and $40 per MWh respectively, the absolute price difference between the day-ahead market and the real-time market would be $10 per MWh on both days, while the difference in average prices would be $0 per MWh.

17 These values are simple averages, rather than load-weighted averages as presented in Figures 1 and 2.

Table 5 displays the average day-ahead and real-time prices, showing the convergence for years 2011-2016. The overall day-ahead premium decreased slightly in 2016 compared to 2015. The average absolute difference between day-ahead and real-time prices was $7.44 per MWh in 2016, down slightly from $8.08 per MWh in 2015.

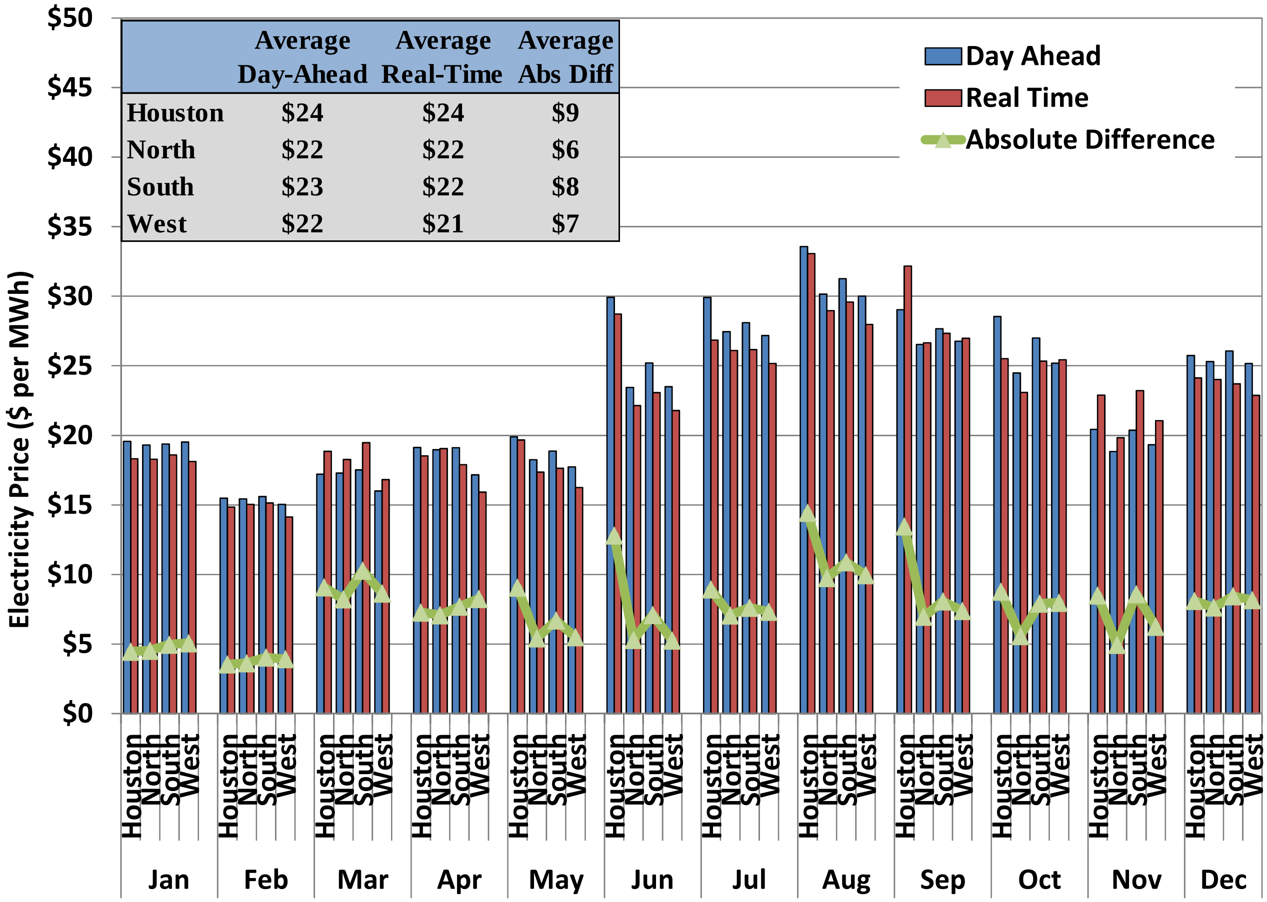
Table 5: Historic Average Day-Ahead and Real-Time Prices

Year	Average Day-Ahead Price	Average Real-Time Price
2016	$23	$22
2015	$26	$25
2014	$40	$38
2013	$33	$32
2012	$29	$27
2011	$46	$43

Real-time energy prices in ERCOT are allowed to rise to levels that are much higher than the shortage pricing in other organized electricity markets, which increases risk and helps to explain the higher day-ahead premiums regularly observed in ERCOT. Although most months experienced a day-ahead premium in 2016, it should not be expected that every month will produce a day-ahead premium. The real-time risks that lead to the premiums will materialize unexpectedly on occasion, resulting in real-time prices that exceed day-ahead prices (*e.g.*, in March, September and November).

In Figure 25 below, monthly day-ahead and real-time prices are shown for each of the geographic zones. Of note is that the volatility in the West zone has decreased and more closely resembles the relative stability of the other zones. The larger difference between day-ahead and real-time prices previously observed in the West zone was likely associated with the uncertainty of forecasting wind generation output and associated transmission congestion.

Figure 25: Day-Ahead and Real-Time Prices by Zone



	Average Day-Ahead	Average Real-Time	Average Abs Diff
Houston	$24	$24	$9
North	$22	$22	$6
South	$23	$22	$8
West	$22	$21	$7

B. Day-Ahead Market Volumes

The next figure summarizes the volume of day-ahead market activity by month, which includes both the purchases and sales of energy, as well as the scheduling of PTP obligations that represent the system flows between two locations. Figure 26 below shows that the volume of day-ahead purchases provided through a combination of generator-specific and virtual energy offers was approximately 53 percent of real-time load in 2016, which was a slight increase compared to 51 percent in 2015. Although it may appear that many loads are subjecting themselves to greater risk by not locking in a day-ahead price, other transactions are being used for this purpose.

Point to Point (PTP) obligations are financial transactions purchased in the day-ahead market. Although PTP obligations do not themselves involve the direct supply of energy, PTP obligations allow a participant to buy the network flow from one location to another.[18] When

18 PTP obligations are equivalent to scheduling virtual supply at one location and virtual load at another.

coupled with a self-scheduled generating resource, the PTP allows a participant to service its load while avoiding the associated real-time congestion costs between the locations. Other PTPs are scheduled by financial participants seeking to arbitrage locational congestion differences between the day-ahead and real-time markets.

To provide a volume comparison, all of these "transfers" are aggregated with other energy purchases and sales, netting location-specific injections against withdrawals to arrive at a "net system flow." The net system flow in 2016 was more than 5 percent lower than in 2015. However, it exceeded real-time load by approximately 22 percent. This does not necessarily suggest that the real-time load is fully hedged by day-ahead purchases and PTP obligations since some of the PTP obligations are scheduled by financial participants that do not serve load. Nonetheless, it is likely that a much higher share of the real-time load is hedged in the day-ahead than the 53 percent scheduling level discussed above.

Figure 26: Volume of Day-Ahead Market Activity by Month

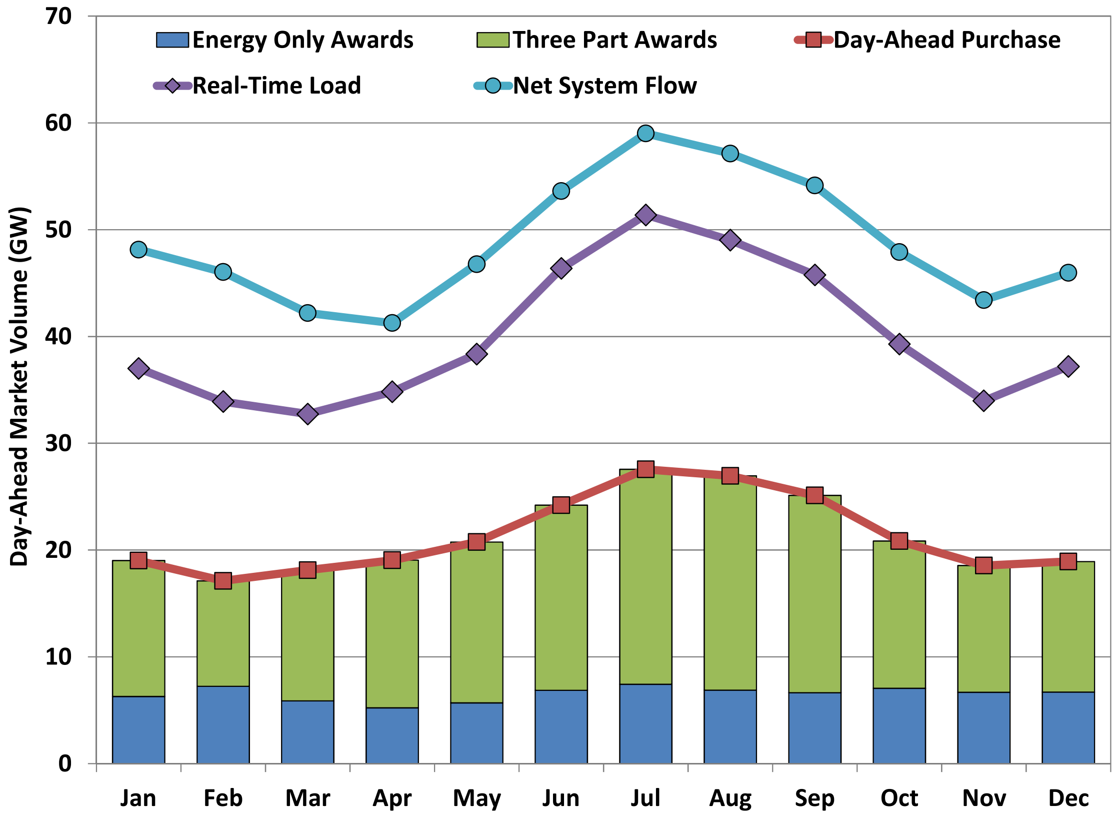
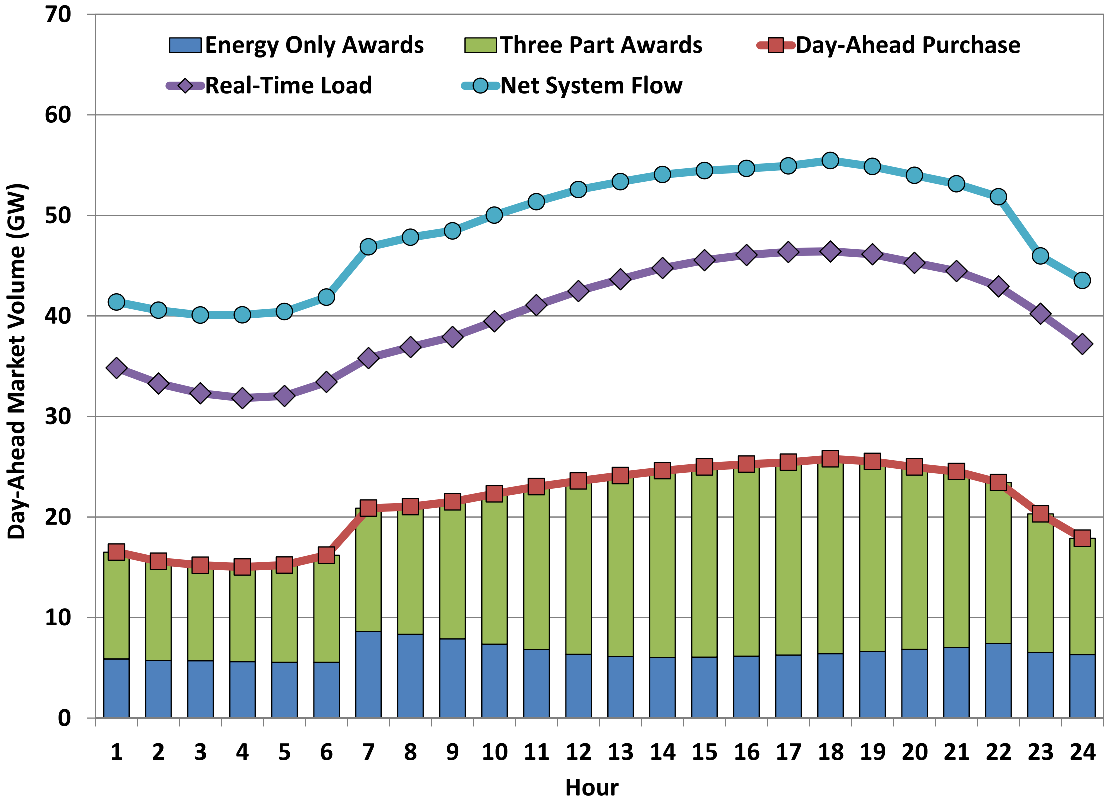


Figure 27 below, presents the same day-ahead market activity data summarized by hour of the day. In this figure the volume of day-ahead market transactions is disproportionate with load levels between the hours of 7 and 22 (hour ending). Since these times align with common bilateral transaction terms, the results in this figure are consistent with market participants using the day-ahead market to trade around those positions.

Figure 27: Volume of Day-Ahead Market Activity by Hour



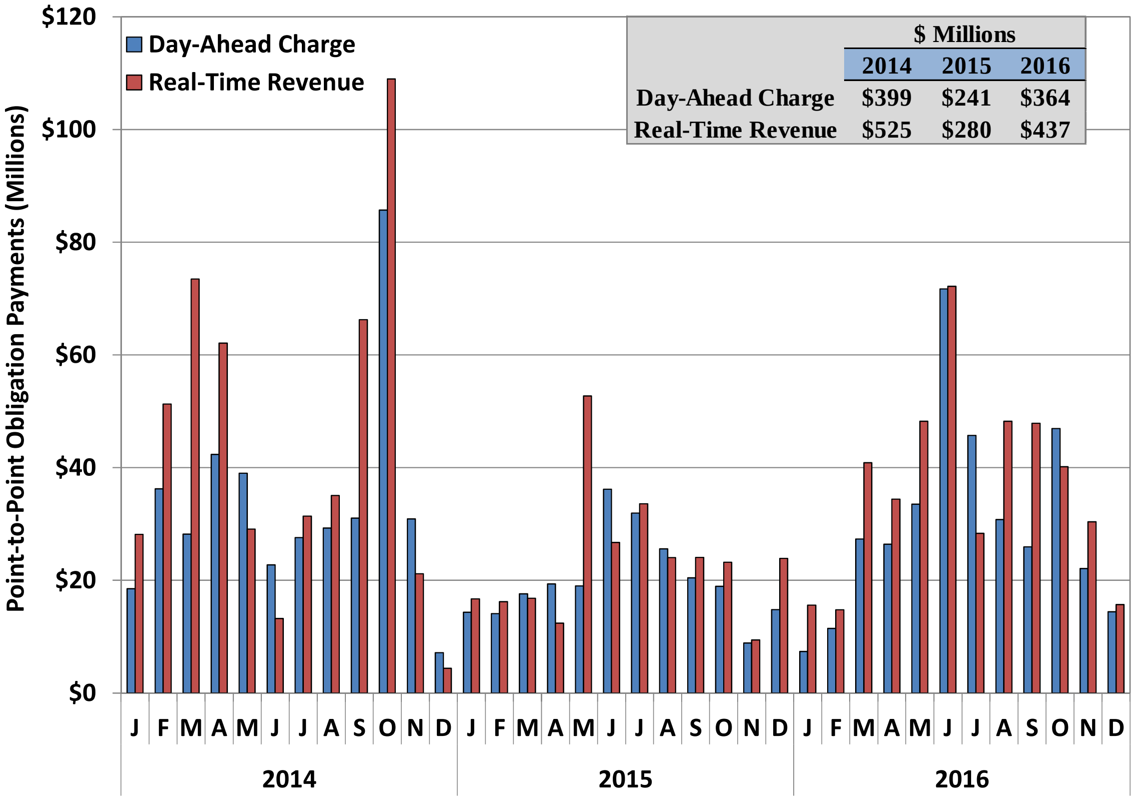
C. Point-to-Point Obligations

Purchases of PTP obligations comprise a significant portion of day-ahead market activity. They are similar to, and can be used to complement, Congestion Revenue Rights (CRRs). CRRs, as more fully described in Section III: Transmission Congestion and Congestion Revenue Rights, 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.

Participants buy PTP obligations by paying the difference in prices between two locations in the day-ahead market. They receive the difference in prices between the same two locations in the real-time market. Hence, a participant that owns a CRR can use its CRR proceeds from the day-ahead market to buy a PTP obligation between the same two points in order to transfer its hedge to real time. Because PTP obligations represent such a substantial portion of the transactions in the day-ahead market, additional details about the volume and profitability of these PTP obligations are provided in this subsection.

The first analysis of this subsection, shown in Figure 28, compares the total day-ahead payments made to acquire these products, with the total amount of revenue received by the owners of PTP obligations in the real-time market.

Figure 28: Point-to-Point Obligation Charges and Revenues



As in prior years, the aggregated total revenues received by PTP obligation owners in 2016 was greater than the amount charged to the owners to acquire them. This indicates that, in aggregate, buyers of PTP obligation profited from the transactions. This occurs when real-time congestion is greater than day-ahead market congestion. Across the year, and in ten of twelve months, the acquisition charges were less than the revenues received, implying that expectations of congestion as evidenced by day-ahead purchases were less than the actual congestion that occurred in real-time. During July and October these expectations were reversed, as congestion anticipated in the day-ahead market did not materialize in real time.

The payments made to PTP obligation owners come from real-time congestion rent. 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 are assessed in Section III: Transmission Congestion and Congestion Revenue Rights.

Figure 29: Point-to-Point Obligation Volume

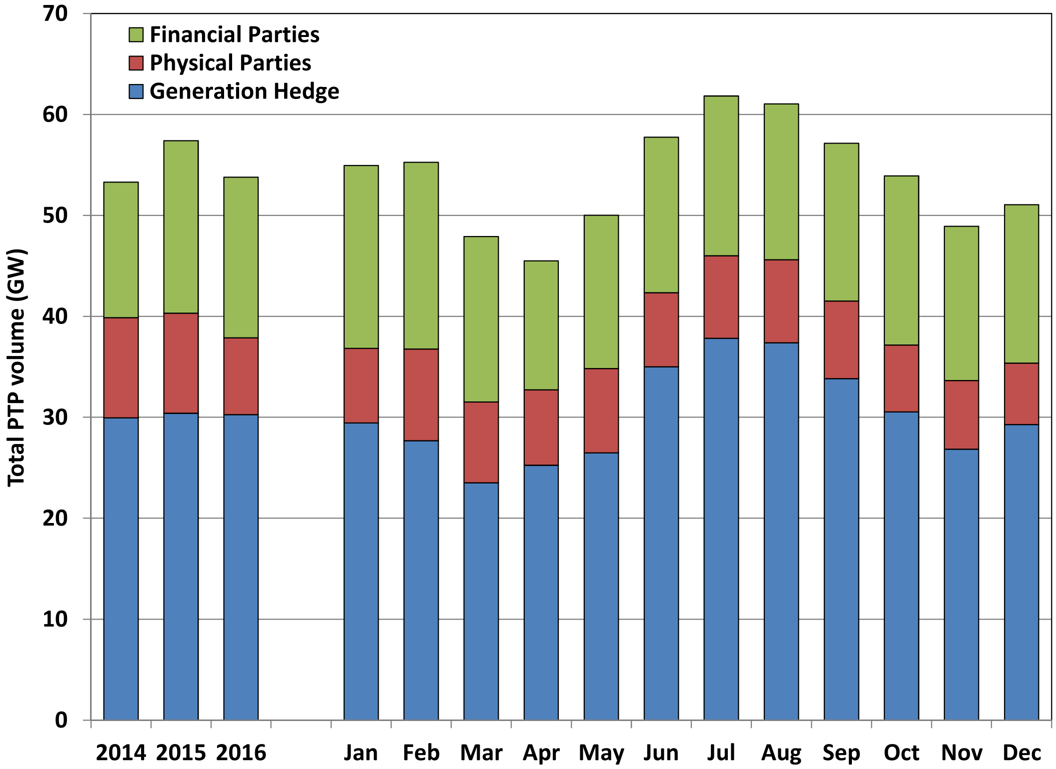
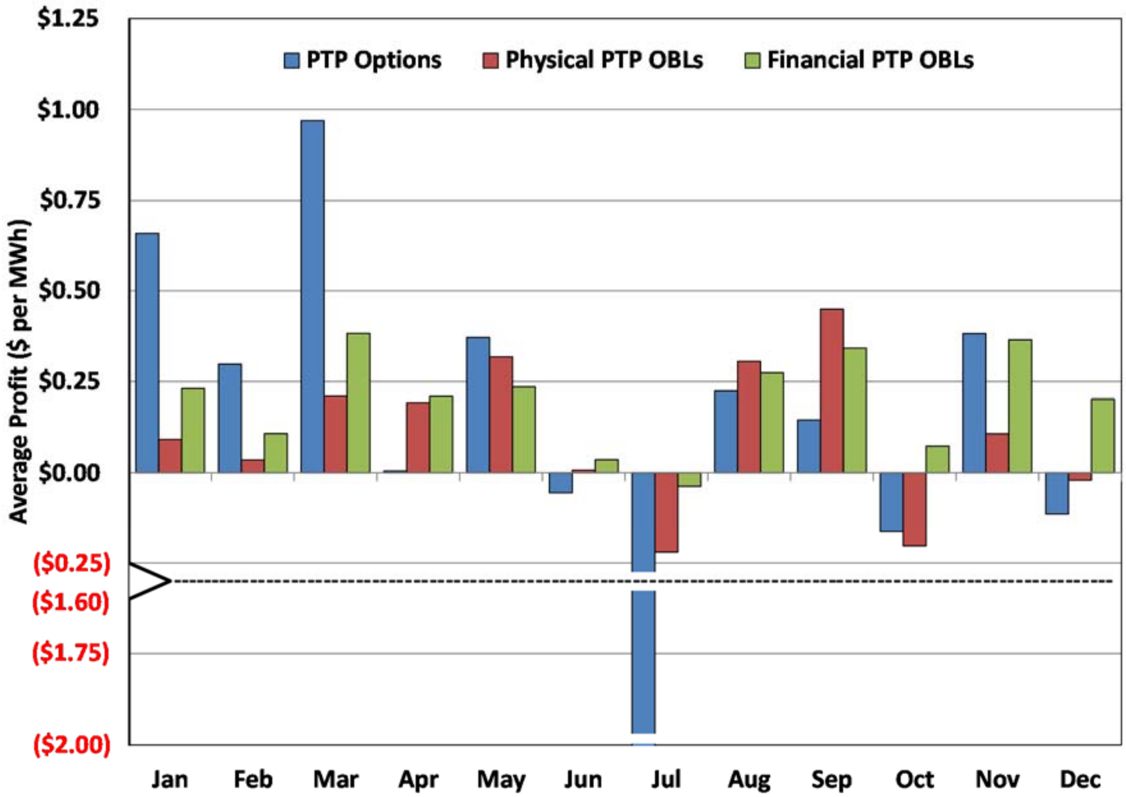


Figure 29 presents the total volume of PTP obligation purchases divided into three categories. Different from Figure 26 and Figure 27 above, the volumes in this figure do not net out the injections and withdrawals occurring at the same location. It presents average purchase volumes on both a monthly and annual basis.

For all PTP obligations that source at a generator location, the capacity up to the actual generator output is considered to be hedging the real-time congestion associated with generating at that location. The figure above shows 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 arbitrage anticipated price differences between two locations. This arbitrage activity is further separated by type of market participant. Physical parties are those that have actual real-time load or generation, whereas financial parties have neither.

To the extent the price difference between the source and sink of a PTP obligation is greater in real-time than it was in the day-ahead market, the owner will profit. Conversely, if the price difference does not materialize in real-time, the PTP obligation may be unprofitable. The profitability of PTP obligation holdings for all physical parties and financial parties are compared in Figure 30. Also shown are the profitability of instruments available only to NOIEs, referred to as PTP obligations settled as options.

Figure 30: Average Profitability of Point-to-Point Obligations



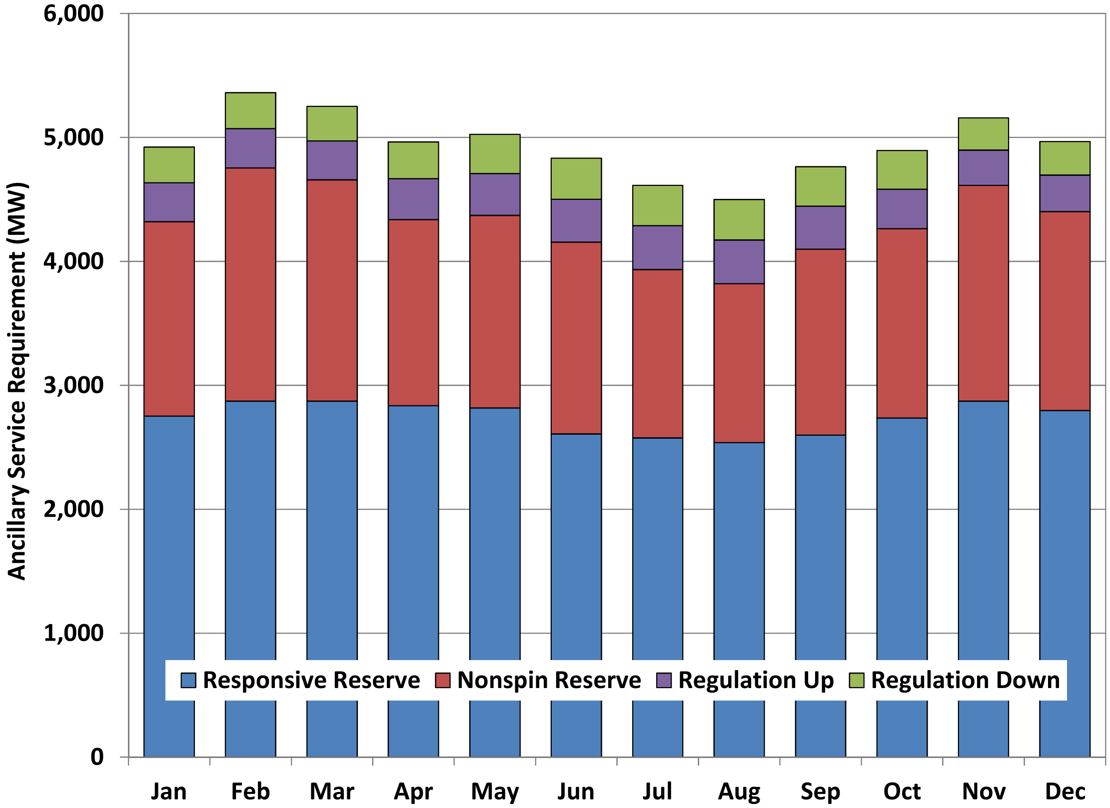
This analysis shows that in aggregate PTP obligation transactions in 2016 were profitable overall, yielding an average profit of $0.120 per MWh. Over the year, PTP obligations owned by physical parties, PTP obligations owned by financial parties, and PTP obligations settled as options were profitable in aggregate in 2016, with average profits of $0.103 per MWh, $0.200 per MWh, and $0.015 per MWh, respectively.

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 them through the ERCOT markets. 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.

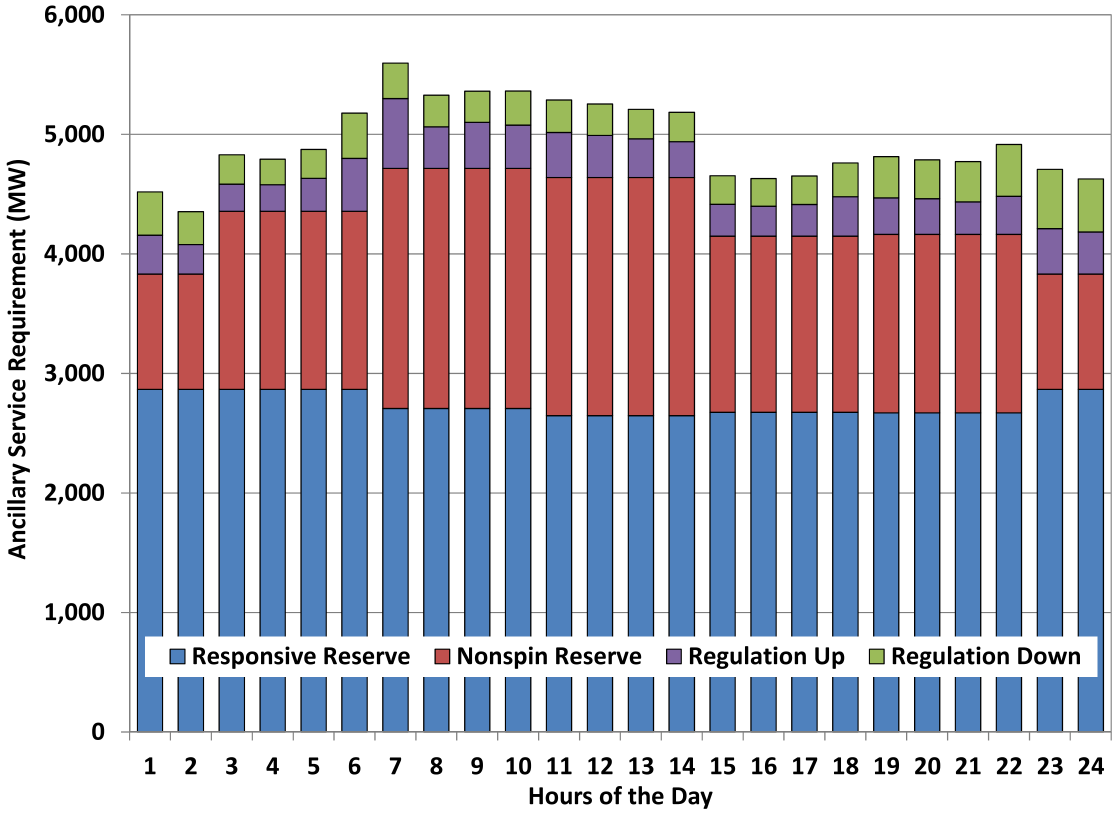
Since June 1, 2015, ERCOT has calculated the requirement for responsive reserves based on a variable hourly need. This requirement is determined and posted in advance for the year. ERCOT procures non-spinning reserves such that the combination of non-spinning reserves and regulation up will cover 95 percent of the calculated Net Load forecast error. ERCOT will always procure a minimum quantity of non-spinning reserves greater than or equal to the largest generation unit. Total requirements for ancillary services declined in 2016. The average total requirement in 2016 was approximately 4,900 MW, a reduction from the average total requirement of 5,300 MW in 2015. The reduction was spread fairly evenly across non-spinning reserves, and regulation up and down. Although the megawatt reduction was spread fairly evenly across these three services, the percentage reduction was much larger for the regulation services (26 percent) than non-spinning (7 percent). Figure 31 displays the hourly average quantities of ancillary services procured for each month in 2016.

Figure 31: Hourly Average Ancillary Service Capacity by Month



Another way to view the ancillary service requirements is by hour, averaged over the course of the year. Figure 32 presents this alternate picture of ancillary service procurement in 2016.

Figure 32: Yearly Average Ancillary Service Capacity by Hour



Ancillary services and energy are co-optimized in the day-ahead market. This means that market participants need not include expectations of forgone energy sales in ancillary service capacity offers. Because ancillary service clearing prices explicitly account for the opportunity costs of selling energy in the day-ahead market, ancillary service prices should generally be correlated with day-ahead energy prices. This correlation was not as obvious in 2016 as other factors contributed to changes in ancillary service prices. Monthly average prices for responsive reserve varied from $8 to $16 per MWh, with the most expensive month being December. One possible explanation is that high wind generation led to changes in unit commitment patterns and less online capacity capable of providing reserves.

Figure 33 below presents the average clearing prices of capacity for the four ancillary services. The absence of meaningful occurrences of scarcity conditions in 2016 resulted in relatively small variation in average energy prices and correspondingly stable ancillary service prices.

Figure 33: Ancillary Service Prices

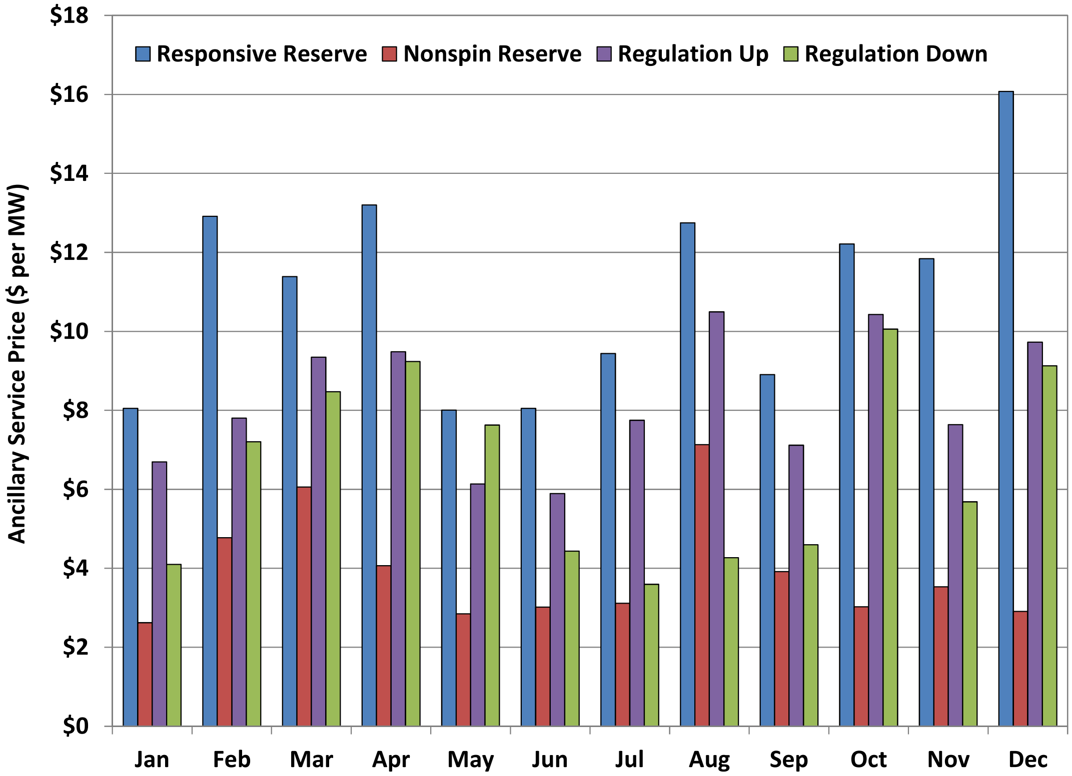


Table 6 compares the average annual price for each ancillary service in 2016 with 2015. The changes in total requirements for ancillary services in 2016 led to concomitant changes in ancillary service prices. The average price for responsive reserve remained about the same, as did the total requirements for the service. Reductions in the average price for non-spinning reserves and regulation up is consistent with the reduced requirements for each of those products.

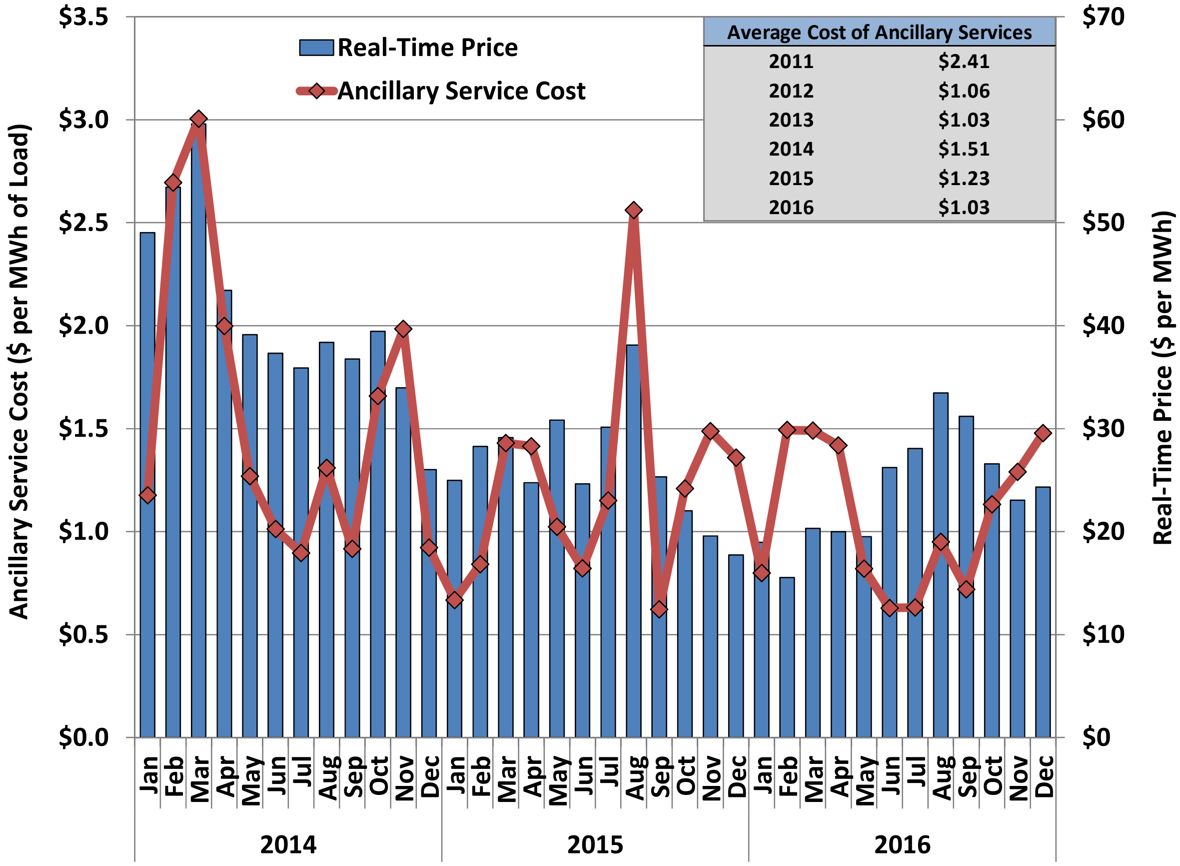
Table 6: Average Annual Ancillary Service Prices by Service

	2015 ($ per MWh)	2016 ($ per MWh)
Responsive Reserve	10.87	11.10
Non-Spinning Reserve	6.92	3.91
Regulation Up	10.59	8.20
Regulation Down	6.01	6.47

The prices for all of the ancillary service products remain modest in part due to the lack of shortages in 2016. When ERCOT experiences a shortage of operating reserves, real-time prices will rise to reflect the expected value of lost load embedded in the ORDC mechanism. The expectation of higher real-time prices will tend to drive up the day-ahead price for ancillary services. Hence, the lack of shortages contributed to the low average ancillary service prices shown in the table.

In contrast to the individual ancillary service prices, Figure 34 shows the monthly total ancillary service costs per MWh of ERCOT load and the average real-time energy price for 2014 through 2016. With no meaningful occurrences of scarcity conditions in 2016, the total cost for ancillary services was relatively low during summer months. The relatively higher costs observed during the other months may be explained by higher wind generation leading to changes in unit commitment patterns and less online capacity available to provide reserves.

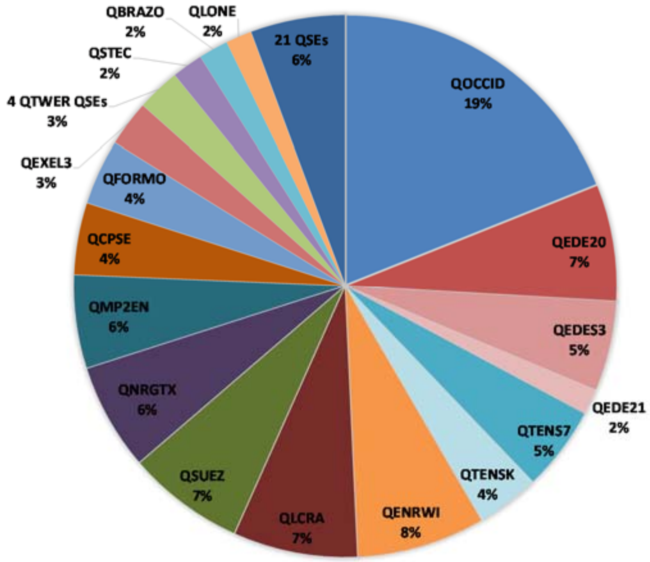
Figure 34: Ancillary Service Costs per MWh of Load



In absolute terms, the average ancillary service cost per MWh of load decreased to $1.03 per MWh in 2016 compared to $1.23 per MWh in 2015. Lower natural gas prices and smaller requirements for ancillary services led to the reduction in ancillary service prices in 2016. Total ancillary service costs were 4.2 percent of the load-weighted average energy price in 2016, similar to the 4.6 and 3.7 percent in 2015 and 2014.

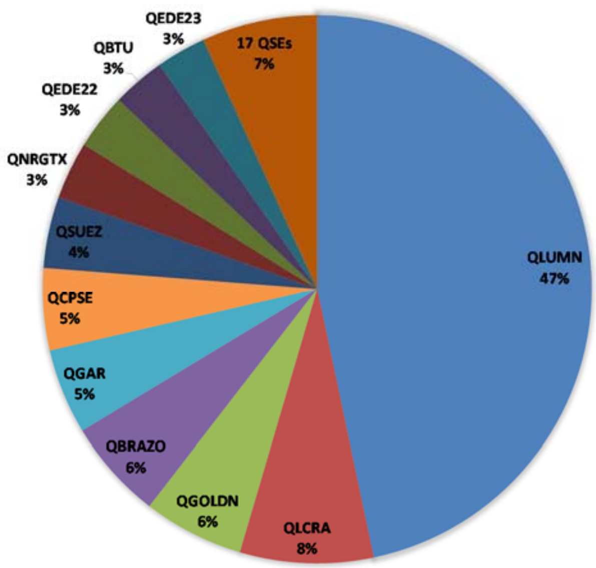
Responsive reserve service is the largest quantity purchased and typically the highest priced ancillary service product. Figure 35 below shows the share of the 2016 annual responsive reserve responsibility including both load and generation, displayed by Qualified Scheduling Entity (QSE). During 2016, 42 different QSEs self-arranged or were awarded responsive reserves as part of the day-ahead market; a slight decrease from 46 different providers in 2015.

Figure 35: Responsive Reserve Providers



In contrast, Figure 36 below shows that the provision of non-spinning reserves is much more concentrated, with a single QSE having nearly half the responsibility to provide non-spinning reserves.

Figure 36: Non-Spinning Reserve Providers



The non-spinning reserve provider concentration highlights the importance of modifying the ERCOT ancillary service market design to include real-time co-optimization of energy and ancillary services. Jointly optimizing all products in each interval would allow the market to substitute its procurements between units on an interval-by-interval basis to minimize costs and set efficient prices. Additionally, it could allow higher quality reserves (e.g., responsive reserves) to be substituted for lower quality reserves (e.g., non-spinning reserves), reducing the reliance upon a single entity to provide this type of lower quality reserves.

Figure 37: Regulation Up Reserve Providers

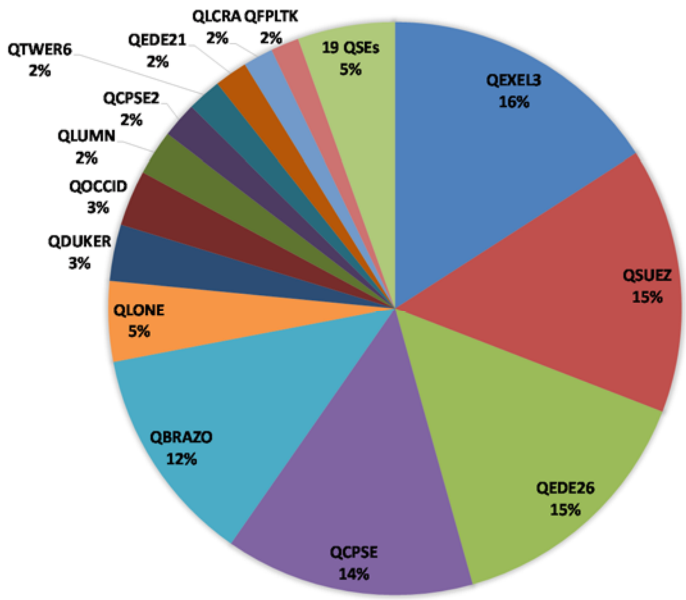
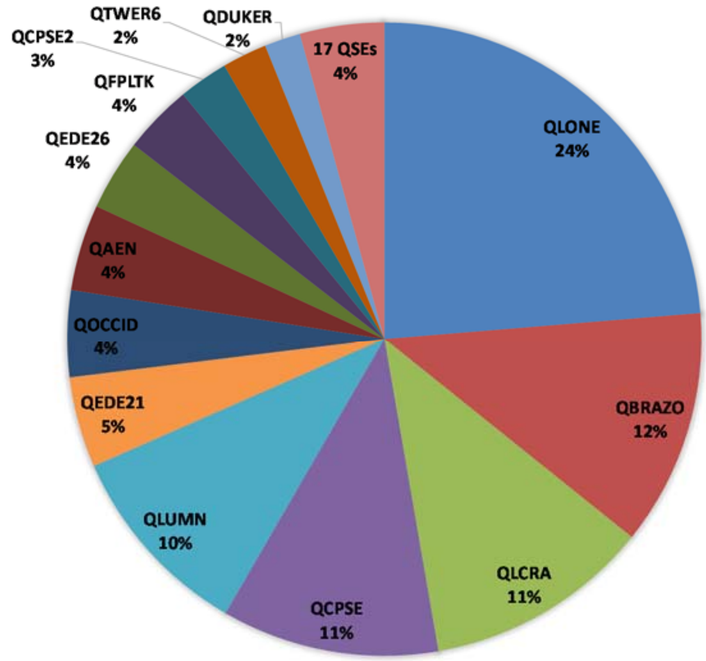


Figure 37 shows the distribution for regulation up reserve service providers and Figure 38 shows the distribution for regulation down reserve providers. These two figures show that the provision of regulation services is more concentrated than responsive reserves, but far less so than non-spinning reserves.

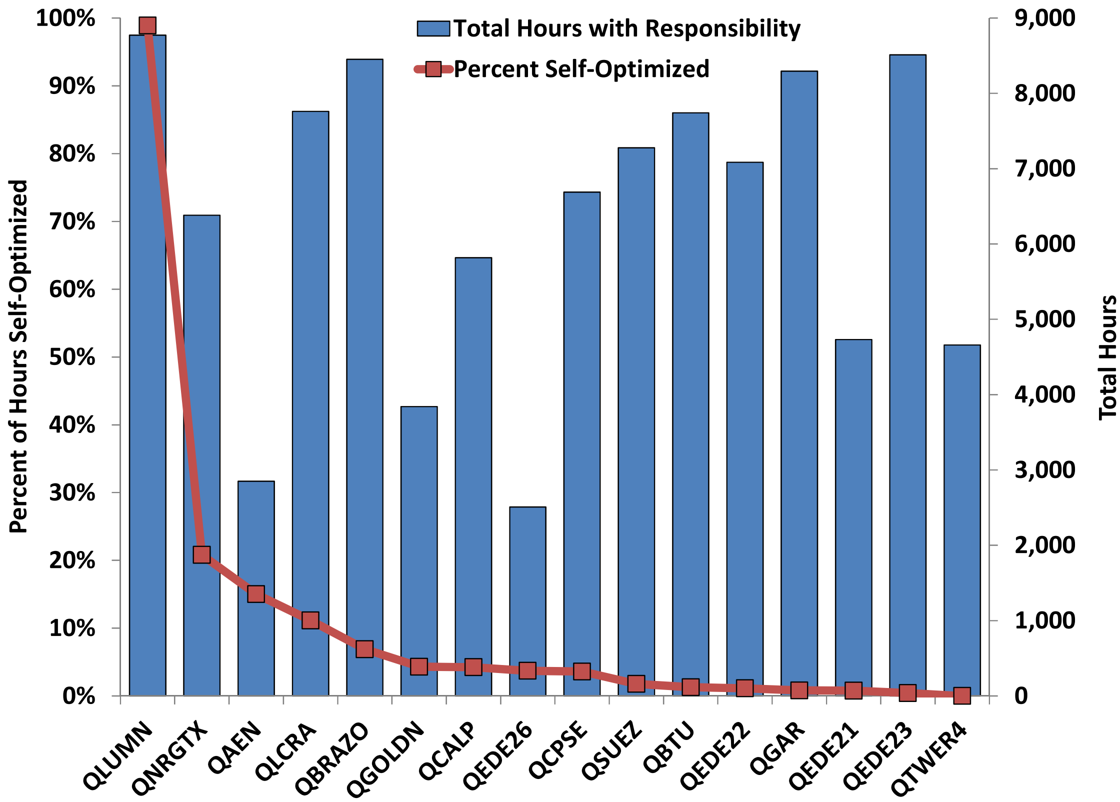
Figure 38: Regulation Down Reserve Providers



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 needed, changes often occur that prompt a QSE to move all or part of its ancillary service responsibility from one unit to another. These changes may be due to a unit outage or to other changes in market conditions affecting unit commitment and dispatch. In short, QSEs with multiple units are continually reviewing and moving ancillary service requirements, presumably to improve the efficiency of ancillary service provision, at least from the QSE's perspective.

The following two charts describe the frequency that each QSE with a unit-specific ancillary service responsibility at 16:00 day-ahead, moved any portion of its ancillary service responsibility to a different unit in its portfolio for real-time operations. Moving ancillary service responsibility is assumed to be in the QSE's self-interest and self-optimization information is shown with the total hours of ancillary service responsibility. Figure 39 shows this information for non-spinning reserves.

Figure 39: Internal Management of Non-Spinning Reserve Portfolio by QSE



The QSEs are listed in descending order based on the frequency of self-optimization. This figure, taken in conjunction with Figure 36, shows that the provider with the largest share of non-spinning reserve responsibility also most frequently moved the responsibility between its units. Luminant had a responsibility to provide non-spinning reserves in almost every hour of 2016, and for nearly all of those hours they moved at least a portion of their responsibility to a unit different from the one that initially received the award.

Figure 40 below provides a similar analysis for the percent of time when responsive reserve service was self-optimized by a QSE, that is, moving the day-ahead responsibility to a different unit before real-time.

Figure 40: Internal Management of Responsive Reserve Portfolio by QSE

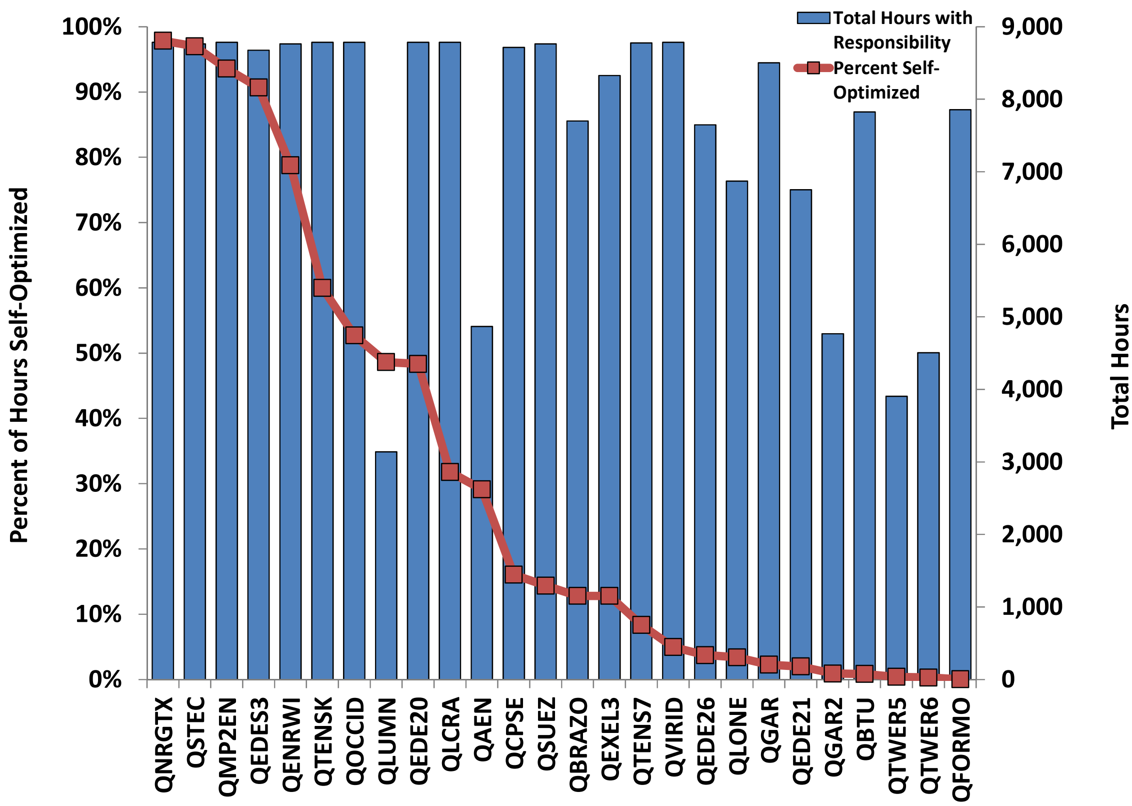


Figure 40 demonstrates that many QSEs moved responsive reserve responsibilities between units more routinely than QSEs providing non-spinning reserve service. For responsive reserve service, seven QSEs moved the responsibility more than 50 percent of the time; whereas only one QSE moved non-spinning reserve responsibility more than 50 percent of the time.

If all ancillary services could be continually reviewed and adjusted in response to changing market conditions, the efficiencies would flow to all market participants and would be greater than what can be achieved by QSEs acting individually. Since the initial consideration of ERCOT's nodal market design, the IMM has been recommending that ERCOT implement real-time co-optimization of energy and ancillary services because of this improved efficiency.

The ERCOT market appropriately reflects the tradeoff between providing capacity for ancillary services versus providing energy in its co-optimized day-ahead market. Those same tradeoffs exist in real time and without comprehensive, market-wide co-optimization, the ERCOT market will continue to be subject to the choices of individual QSEs. These choices are likely to be in the QSE's best interest. They are not likely to lead to the most economic provision of energy and ancillary services for the market as a whole. Further, QSEs without large resource portfolios are effectively precluded from participating in ancillary service markets due to the replacement risk they face having to rely on a supplemental ancillary services market (SASM). This replacement risk is substantial. Clearing prices for ancillary services procured in SASM are typically ten to thirty times greater than annual average clearing prices from the day-ahead market.

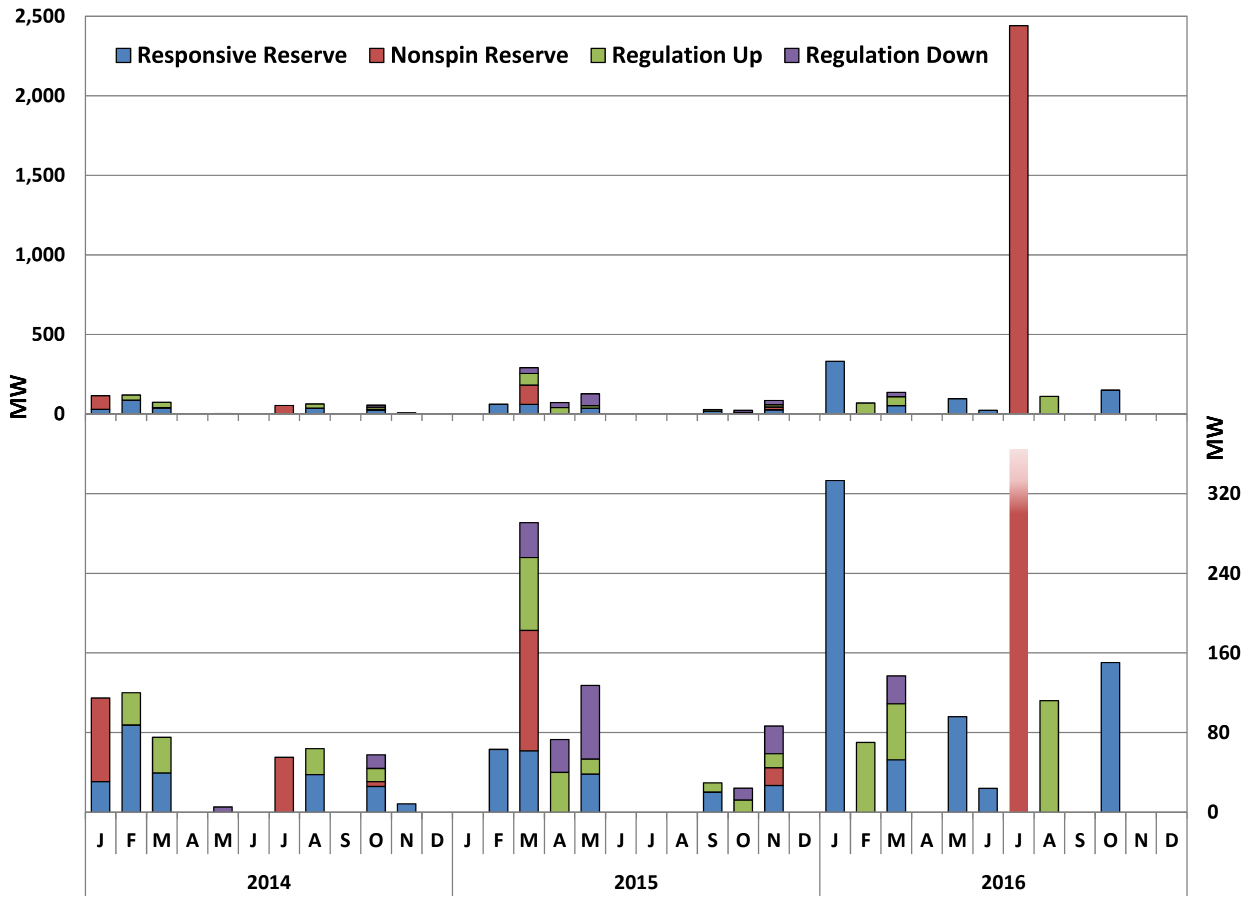
ERCOT uses SASMs to procure replacement ancillary service capacity when transmission constraints arise which make the capacity undeliverable, or when outages or limitations at a generating unit lead to failure to provide. A SASM may also be opened if ERCOT changes its ancillary service plan; this did not occur during 2016. ERCOT executed a SASM for 76 hours in 2016. This was slightly more frequent than in 2015, but still less than one percent of the time. The frequency of SASMs continues to be very low, declining from seven percent in 2012, three percent in 2013, and two percent in 2014.

The final analysis in this section, shown in Figure 41, summarizes the average quantity of each service that was procured via SASM. Identical data is shown on two different scales because of the very large SASM procurement of non-spinning reserves in July 2016.

The opportunity exists for market participants to use the SASM process as a re-configuration market. That is to move into or out of ancillary service positions awarded day ahead. SASMs were infrequent largely because of the dearth of ancillary service offers typically available throughout the operating day, limiting re-configuration opportunities. The SASM procurement method, while offer based, is inefficient and problematic.

Because ancillary services are not co-optimized with energy in the SASM, potential suppliers are required to estimate opportunity costs rather than have the auction engine calculate it directly, which leads to resources that underestimate opportunity costs being inefficiently preferred over resources that overestimate opportunity costs.

Figure 41: Ancillary Service Quantities Procured in SASM



Further, the need to estimate the opportunity costs, which change constantly and significantly over time as the energy price changes, provides a strong disincentive to SASM participation, contributing to the observed lack of SASM offers. The paucity of SASM offers frequently leaves ERCOT with two choices in response to ancillary service un-deliverability or failure to provide: (1) use an out-of -market ancillary service procurement action with its inherent inefficiencies; or (2) operate with a deficiency of ancillary services with its inherent increased reliability risk.

Real-time co-optimization of energy and ancillary services does not require resources to estimate opportunity costs, would eliminate the need for the SASM mechanism, and allow ancillary services to be continually shifted to the most efficient provider. Because co-optimization allows the real-time market far more flexibility to procure energy and ancillary services from online resources, it would also reduce ERCOT's need to use RUC procedures to acquire ancillary services. Its biggest benefit would be to effectively handle situations where entities that had day-ahead ancillary service awards were unable to fulfill that commitment, e.g. due to a generator forced outage. Thus, implementation of real-time co-optimization would provide benefits across the market.

III. Transmission Congestion and Congestion Revenue Rights

One of the most important functions of any electricity market is to manage the flows of power on the transmission network by not allowing additional power flow on transmission facilities that have reached 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 the output level of one or more generators to reduce the amount of electricity flowing on any transmission facility nearing its operating limit.[19] This leads to higher costs as a result of necessary changes to generation output to ensure that operating limits are not violated. This increase in more expensive generation and/or decrease in less expensive generation results in different prices at different nodes. The decision about which generator(s) will vary its output is based on the generator's energy offer curve and how much of its output will flow across the overloaded transmission element. This leads to the dispatch of the most efficient generation to reliably serve demand while providing locational marginal pricing reflective of the actions taken to ensure system security.

The locational difference in prices produced by congestion can provide challenges to parties that have transacted in long term power contracts; namely, if the production point (for a seller) or consumption point (for a purchaser) is different from the contracted delivery point, the party is subject to the risk that the prices will be different when settled. Congestion Revenue Rights (CRR) markets enable parties to purchase the rights to those price differences in seasonal and monthly blocks, and thus achieve some level of price certainty.

This section of the report summarizes transmission congestion in 2016, provides a review of the costs and frequency of transmission congestion in both the day-ahead and real-time markets, and concludes with a review of the activity in the CRR market.

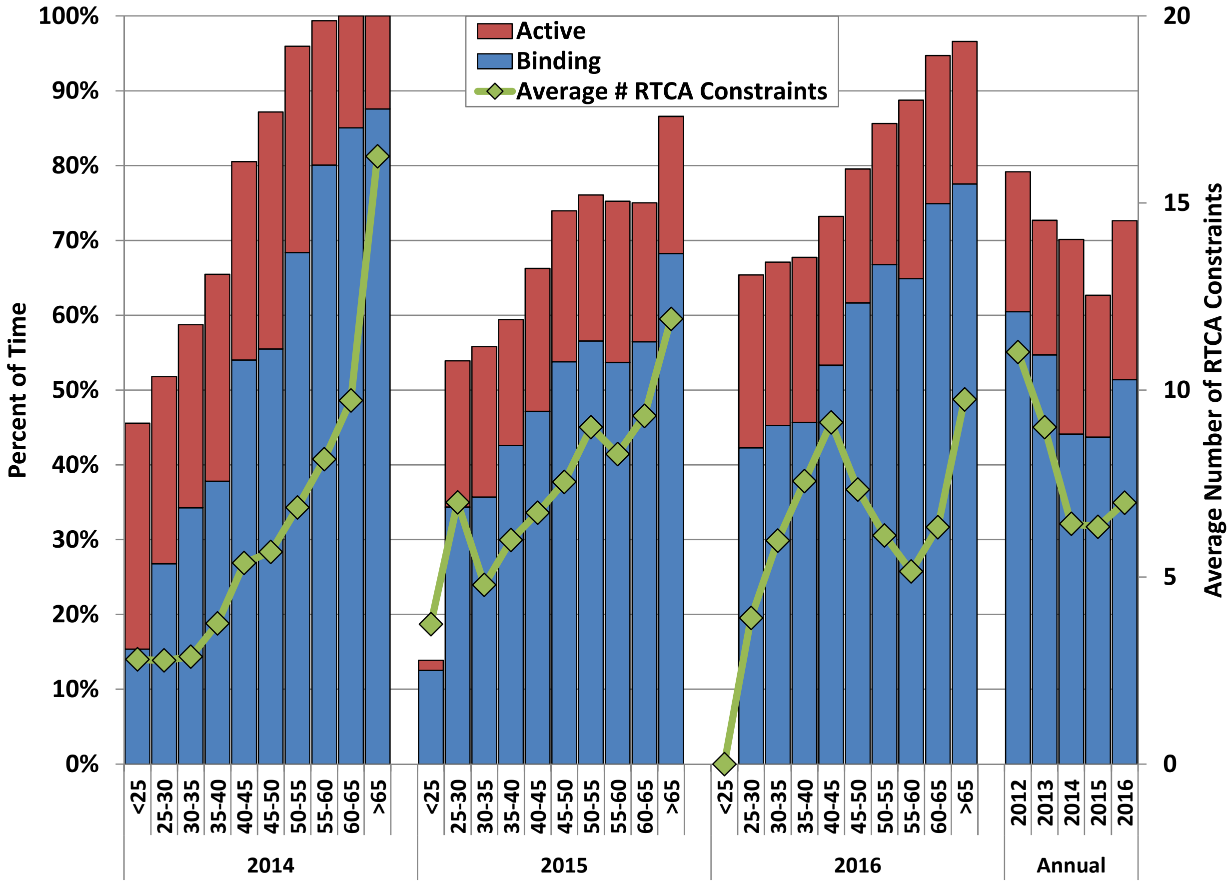
A. Summary of Congestion

The total congestion costs experienced in the ERCOT real-time market were $497 million in 2016, a 40 percent increase from 2015 values. This is a substantial increase, especially given the reduction in natural gas prices that would typically reduce transmission congestion. The increase in congestion occurred as constraints were binding in 8 percent more intervals in 2016. The North and Houston zones experienced an increase in price differences between the two zones and within each zone in 2016. The costs of congestion in the West and South zones in 2016 were similar to 2015.

[19] 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.

Figure 42 provides a comparison of the amount of time transmission constraints were active and binding for various load levels in 2014 through 2016. This figure also indicates the average number of constraints in a Real-Time Contingency Analysis (RTCA) execution for each load level. This is the process in which the resulting flows on the transmission system are evaluated after systematically removing elements of the transmission system. If the loss of a transmission element (contingency) results in a flow higher than the element rating, this is considered a thermal constraint. Binding transmission constraints are those for which the dispatch levels of generating resources are actually altered in order to maintain transmission flows at reliable levels. The costs associated with this re-dispatch are the system's congestion value and are included in nodal prices. Active transmission constraints are those which the dispatch software evaluated, but did not require a re-dispatch of generation.

Figure 42: Frequency of Binding and Active Constraints



Constraints were activated more frequently in 2016 – 73 percent of all hours compared to 63 percent in 2015. The percent of time with active constraints in 2016 is very similar to 2013. There was more constraint activity at nearly all load levels in 2016 except for load levels below 25 GW. The most notable difference between 2016 and 2015 is that, while RTCA on average showed fewer constraints in 2016, the percentage of time with an active constraint in each load level, except for the very lowest loads, was higher than 2015. This is explained by the number of SCED intervals with an active Generic Transmission Constraint (GTC) which increased by 66 percent in 2016 as compared to 2015.

GTCs are not derived from the real-time contingency analysis but rather are based on studies performed by ERCOT. GTC limits are calculated by ERCOT the day before the operating day. A GTC indicates a requirement for SCED dispatch to resolve a stability or a voltage condition. Certain GTCs are monitored in real-time. The North to Houston, Bakersfield, Panhandle, Laredo, Zorillo to Ajo, and Valley import are analyzed in real-time using the Voltage Stability Assessment or Transient Stability Assessment components of EMS. Using these tools to continuously evaluate these constraints in real time provides a more accurate limit than what was calculated by ERCOT in the day-ahead process. Actions taken to resolve a GTC may also benefit other potential congestion issues, resulting in fewer thermal constraints in RTCA. This could explain the lower number of RTCA constraints at certain load levels in 2016.

Shown below in Table 7 are the GTCs that were monitored in 2016. The highlighted GTC, Molina, was removed on July 8, 2016 when the stability issue was resolved.

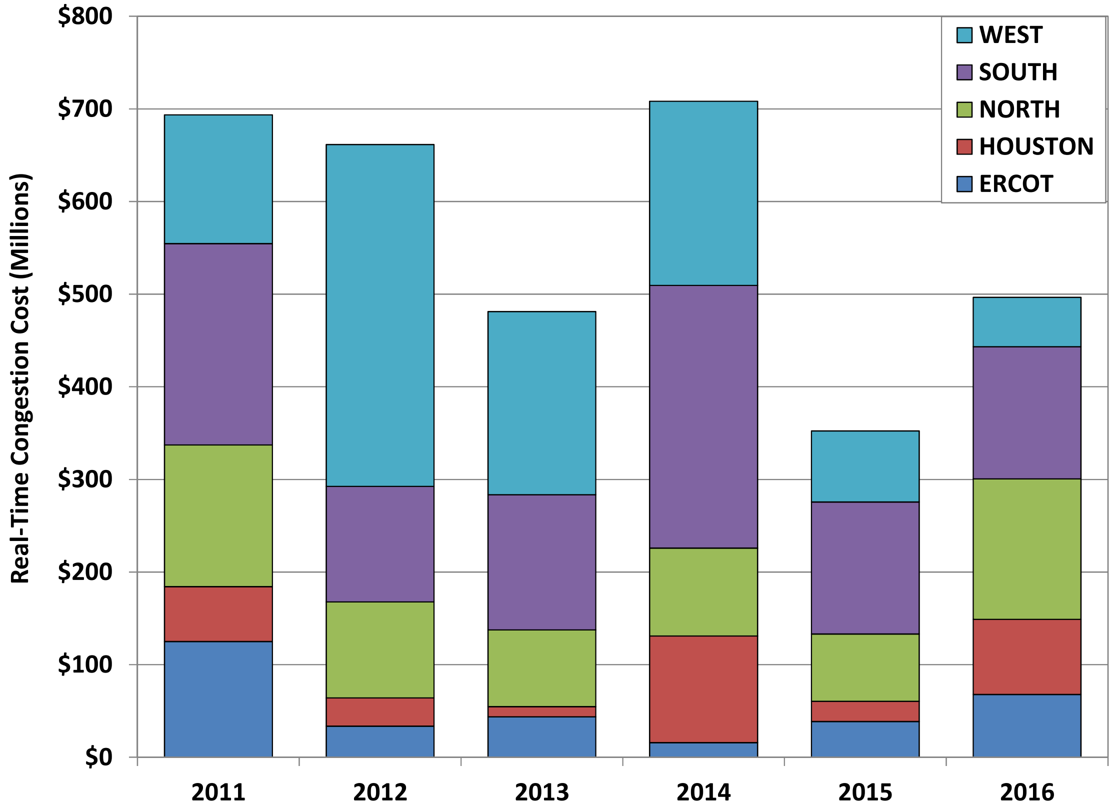
Table 7: Generic Transmission Constraints

Generic Transmission Constraint	Effective Date
North to Houston	December 1, 2010
Rio Grande Valley Import	December 1, 2010
Zorillo to Ajo	February 27, 2015
Panhandle	July 31, 2015
Laredo	September 9, 2015
Liston	November 12, 2015
Molina	December 1, 2015
Pomelo Tap	October 5, 2016
Red Tap	August 29, 2016

Except for the North to Houston and the Rio Grande Valley Import constraints, all GTCs resulted from issues identified during the generation interconnection process. In 2016, NPRR809 was introduced to allow the interconnection study results to become more transparent to market participants and provide earlier notification of an upcoming GTC.

Figure 43 displays the amount of real-time congestion costs associated with each geographic zone. Costs associated with constraints that cross zonal boundaries, for example, North to Houston, are shown in the ERCOT category.

Figure 43: Real-Time Congestion Costs



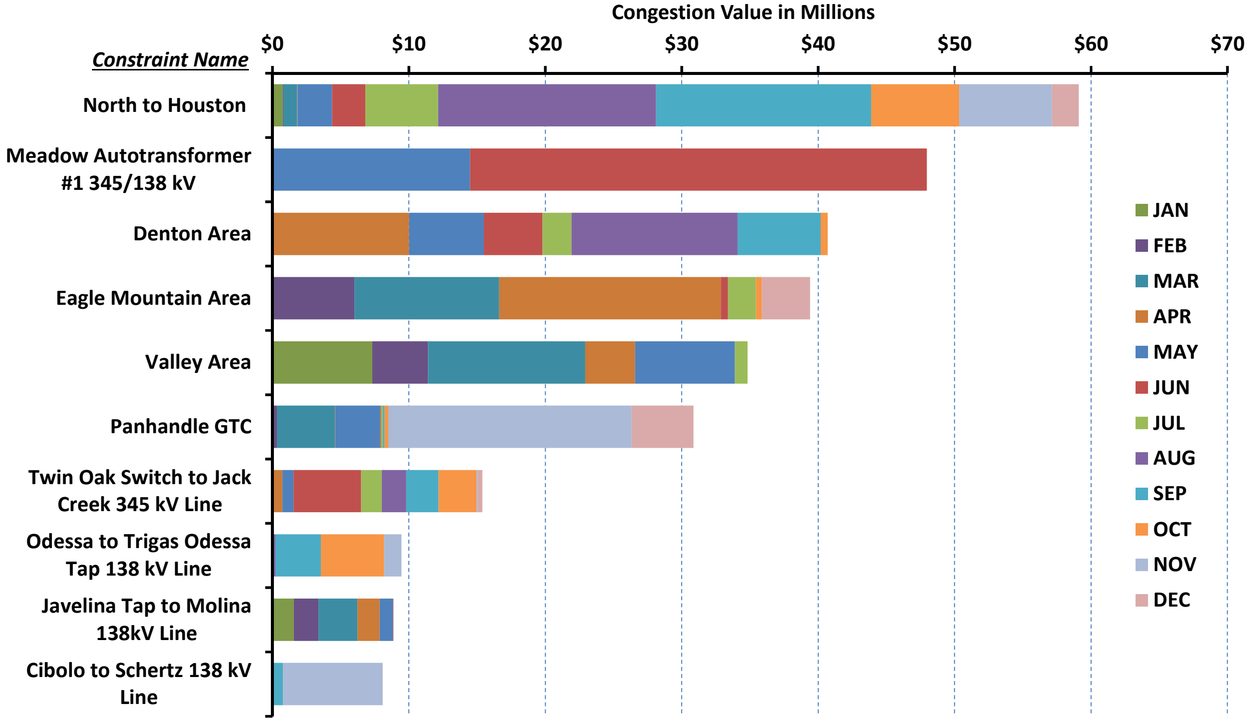
Cross zonal congestion in 2016 was the most costly since 2011 due to the increased frequency and cost associated with Houston import constraints. The North and Houston zones experienced an increase in price impacts between and within the two zones in 2016. Congestion costs for the West and South zones were very similar to 2015. Most of the increased congestion was attributable to a variety of transmission outages, some of which were taken to perform system upgrades. The completion of these upgrades is expected to reduce associated congestion.

B. Real-Time Constraints

The review of real-time congestion begins with describing the congested areas with the highest financial impact. For this discussion, a congested area is determined by consolidating multiple real-time transmission constraints that are determined to be similar due to their geographic proximity and constraint direction. There were 320 unique constraints that were binding at some point during 2016 with a median financial impact of approximately $150,000. In 2015 there were 350 unique constraints with a median financial impact of $162,000. The most expensive constraints in 2016 had a larger price impact than 2015 as evidenced by the decrease in the number of unique constraints.

Figure 44 displays the ten most costly real-time constraints as measured by congestion value. The North to Houston constraint, comprised of the GTC and multiple thermal constraints, most notably the double circuit Singleton to Zenith 345 kV lines and double circuit Jewett to Singleton 345 kV lines, was the most congested location in 2016 at $59 million. This area was also the most costly in 2015 at $38 million.

Figure 44: Most Costly Real-Time Constraints



The second-highest valued congested element was the Meadow 345/138 kV #1 autotransformer which feeds Houston from the south. Its impacts were $48 million and occurred solely in May and June. They were related to a 345/138 kV transformer replacement outage at the PH Robinson generation site.

The third most congested area was the Denton area, north of Dallas / Fort Worth, which includes the West Denton to Jim Christal 138 kV line and West Denton to Fort Worth 138 kV line constraints. Congestion in this area was due to outages taken to accommodate transmission and substation construction to support load growth in the Denton area.

Congestion due to planned and forced outages within the Eagle Mountain area includes the constraints of Eagle Mountain to Morris Dido 138 kV line, Morris Dido to Rosen Heights

138 kV line, as well as the Eagle Mountain 345/138 kV autotransformer #2. This congestion is located in the North zone and feeds Dallas / Fort Worth load from the west.

The constraints in the Valley area located on the east side of the lower Rio Grande Valley and include the Los Fresnos to Loma Alta 138 kV line ($27 million) and La Palma to Villa Cavazos 138 kV line ($8 million). These constraints were often in effect during the time that other transmission facilities in the area were taken out of service to accommodate construction of new transmission facilities. Examples of new construction include the 345 kV lines from Lobo to North Edinburg and North Edinburg to Loma Alta, which were built to improve the ability to reliably serve load in the Valley. Lobo is located close to Laredo, which allows a northwest feed into the Valley, while the other 345 kV line was built to help facilitate cross-Valley flows. The newly-constructed lines also impacted the Valley Import GTC, reducing the congestion value by 60 percent from $17 million in 2015 down to $7 million in 2017. The Valley Import GTC was binding primarily in March and May of 2016.

Further affecting congestion in the Valley at the end of 2016 was the permanent loss of the Frontera generating station in the Valley, which disconnected from the ERCOT grid to fully interconnect with Mexico. The 524 MW combined-cycle unit was built in 2000 with the capability for one of its gas turbines to switch grids. Frontera announced its intent to fully disconnect from ERCOT in 2014 and completed the disconnection in October 2016. The aforementioned 345 kV lines built in the Valley were built to strengthen the transmission system in anticipation of the departure of the Frontera unit.

The Panhandle GTC was implemented in July 2015 and had its largest impact of $18 million in November due to major 345 kV double-circuit line outages taken to repair tornado damage that occurred in May. By the end of 2016, there was almost 1,500 MW of wind and gas generation installed in the Panhandle. The Panhandle GTC is comprised of the eight 345 kV lines from northwest Texas where most of the Panhandle wind interconnects.

The next four constraints were due to planned outages and/or high loads in the area. The Twin Oak Switch to Jack Creek 345 kV line is located between the North and Houston zones and feeds into College Station. The Odessa to Trigas Odessa Tap 138 kV line is located in the far west and incurred congestion primarily in September and October. The Javelina Tap to Molina 138 kV line is east of Laredo and experienced more frequent congestion due to wind generation installed in the area. There is also a GTC associated with the area called Molina GTC, however its congestion costs were minor. And lastly, the Cibolo to Schertz 138 kV line is east of San Antonio and incurred congestion solely in November.

Irresolvable Constraints

The constraint shadow price is the value at which economic dispatch results in profit-maximizing for the generators while also meeting demand at the lowest overall production cost. However, if the dispatch cannot resolve a reliability problem with the available generators, the shadow price would continue to increase as the economic dispatch sought a solution. In situations where there is no generation solution the shadow price would theoretically rise to infinity. Therefore, the shadow price is capped. Shadow price caps are based on a reviewed methodology,[20] and are intended to reflect the level of reduced reliability that occurs when a constraint is not able to be resolved. Currently the shadow price caps are $5,000 per MW for base-case (non-contingency) or voltage violations, $4,500 per MW for 345 kV, $3,500 per MW for 138 kV, and $2,800 per MW for 69 kV thermal violations. A GTC shadow price cap is considered a voltage constraint and is set at $5,000 per MW.

When a constraint becomes irresolvable, chronically reaching the shadow price cap, ERCOT's dispatch software cannot find a dispatch combination to reduce the flows on the transmission element(s) of concern to a reliable operation level. A regional peaker net margin mechanism is used such that once local price increases accumulate to a predefined threshold due to an irresolvable constraint, the constraint's shadow price cap would be re-evaluated. The shadow price is recalculated based upon the mitigated offer cap of existing resources with a defined shift factor threshold consistent with the methodology.

---

[20] ERCOT Business Practice Manual, Setting the Shadow Price Caps and Power Balance Penalties in Security Constrained Economic Dispatch (ERCOT Board Approved 2/14/17), available at http://www.ercot.com/mktrules/obd/obdlist.

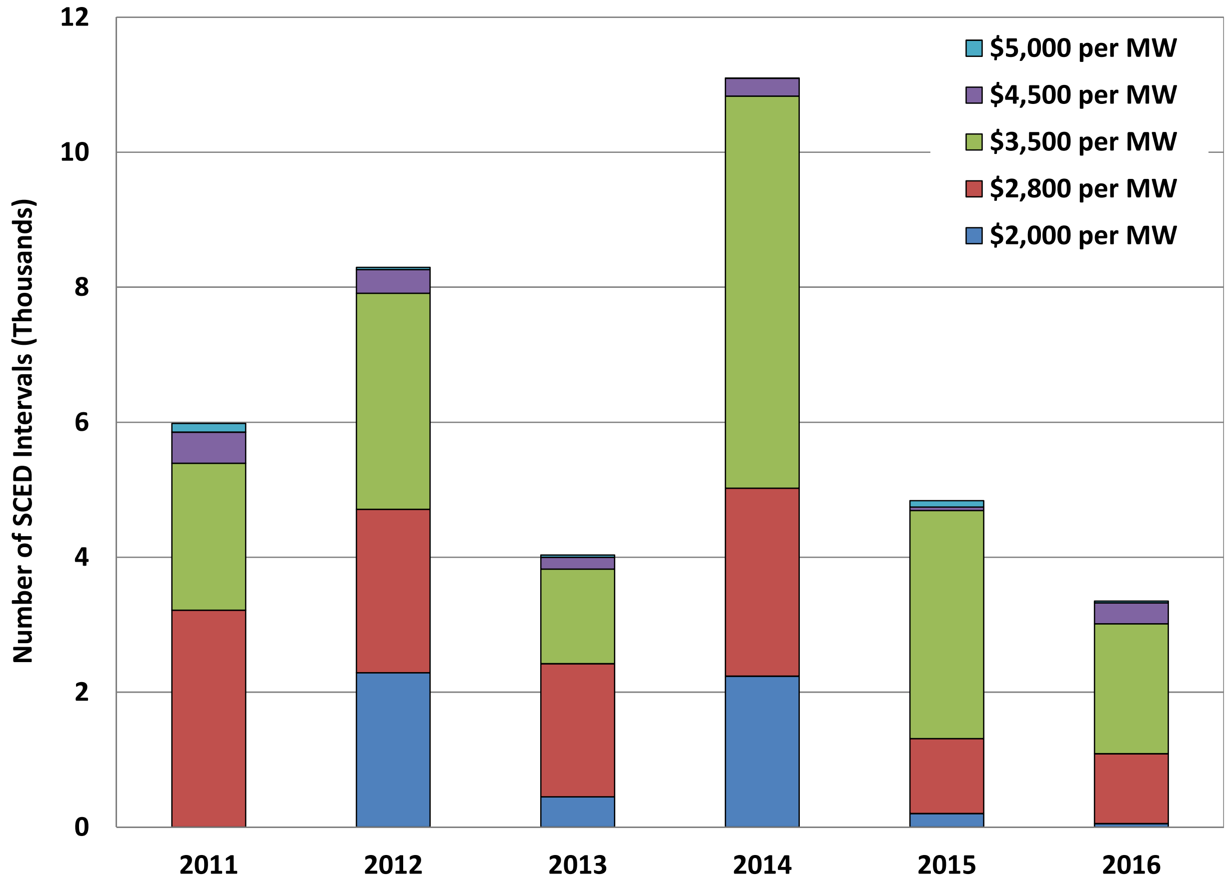
Table 8: Irresolvable Elements

Irresolvable Element	Original Max Shadow Price	2016 Adjusted Max Shadow Price	Effective Date	Termination Date	Load Zone
Valley Import	$5,000	$2,000	1/1/12	-	South
Heights TNP #1 138/69 kV Transformer	$3,500	$2,000	9/23/14	1/30/16	Houston
Abilene Northwest to Ely Rea Tap 69 kV Line	$2,800	$2,000	9/26/14	-	West
Harlingen to Oleander 69 kV Line	$2,800	$2,000	10/9/14	-	South
Rio Hondo to East Rio Hondo 138 kV line	$3,500	$2,000	10/10/14	-	South
Emma to Holt Switch 69 kV line	$2,800	$2,800	10/27/14	-	West
Heights TNP #2 138/69 kV transformer	$3,500	$2,000	10/28/14	1/30/16	Houston
San Angelo College Hills 138/69 kV autotransformer	$3,500	$2,000	7/22/15	-	West

As shown above in Table 8, eight elements were deemed irresolvable in 2016 and had a shadow price cap imposed according to the irresolvable constraint methodology. Two elements were deemed resolvable during ERCOT's annual review and were removed from the list. All three irresolvable constraints located in the South Load Zone are located in the Valley. This is the smallest annual list since the irresolvable methodology was implemented in 2012.

Constraints that are violated in SCED are noted by the flow being greater than the value of the constraint. In other words, SCED was not able to resolve the constraint with the available re-dispatch of generation. This is also noted by a shadow price that is equal to the designated maximum shadow price of the constraint. Figure 45 below shows the number of SCED intervals a constraint reached its maximum shadow price.

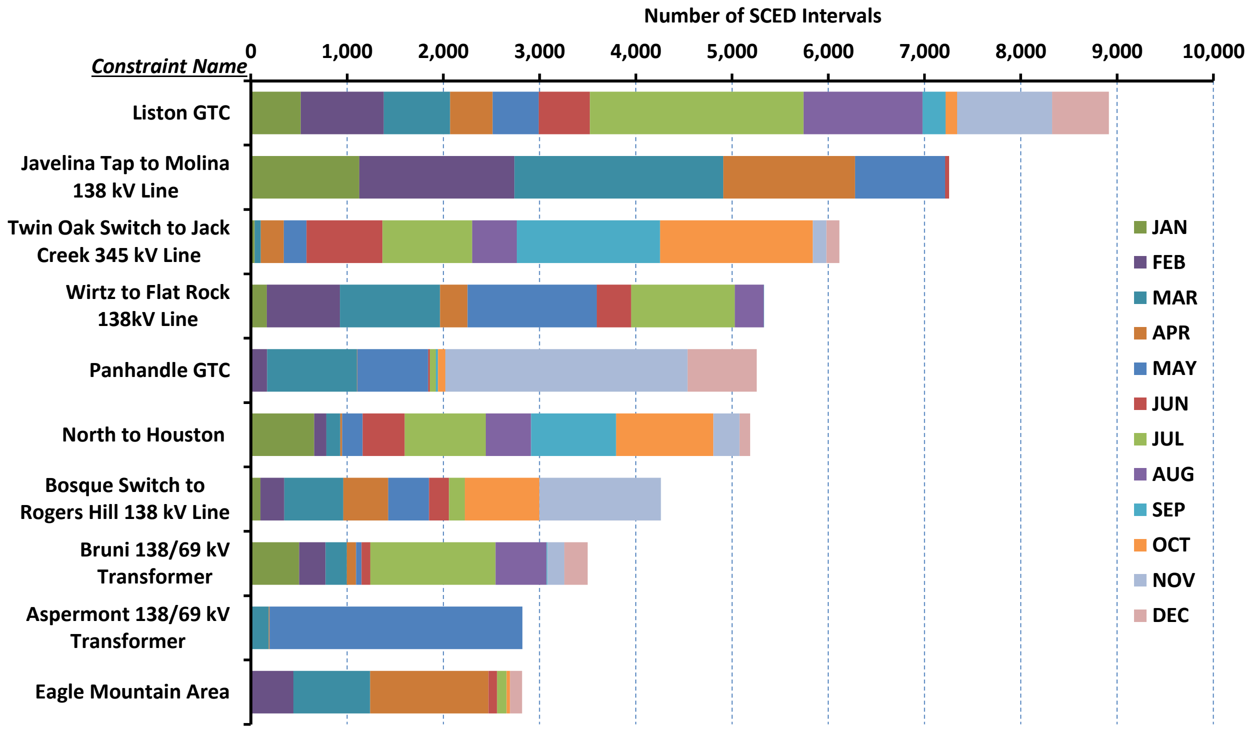
Figure 45: Frequency of Violated Constraints



The frequency of constraints being at maximum shadow prices in 2016 was the lowest since the start of the nodal market. However, the higher-priced constraints at $5,000 per MW and $4,500 per MW occurred in more SCED intervals in 2016 – 337 intervals – than in the years 2013 through 2015. This corresponds to the high congestion value experienced in 2016 and further highlights the impact of more North to Houston congestion, as well as the increase in violated GTCs. Even with the more frequent occurrence of base case and 345 kV contingency overloads, no new irresolvable constraints were identified during 2016. The majority of the irresolvable constraints have shadow price caps set at $2,000 per MW.

Figure 46 presents a slightly different set of real-time congested areas. Shown are the areas that were most frequently constrained.

Figure 46: Most Frequent Real-Time Constraints



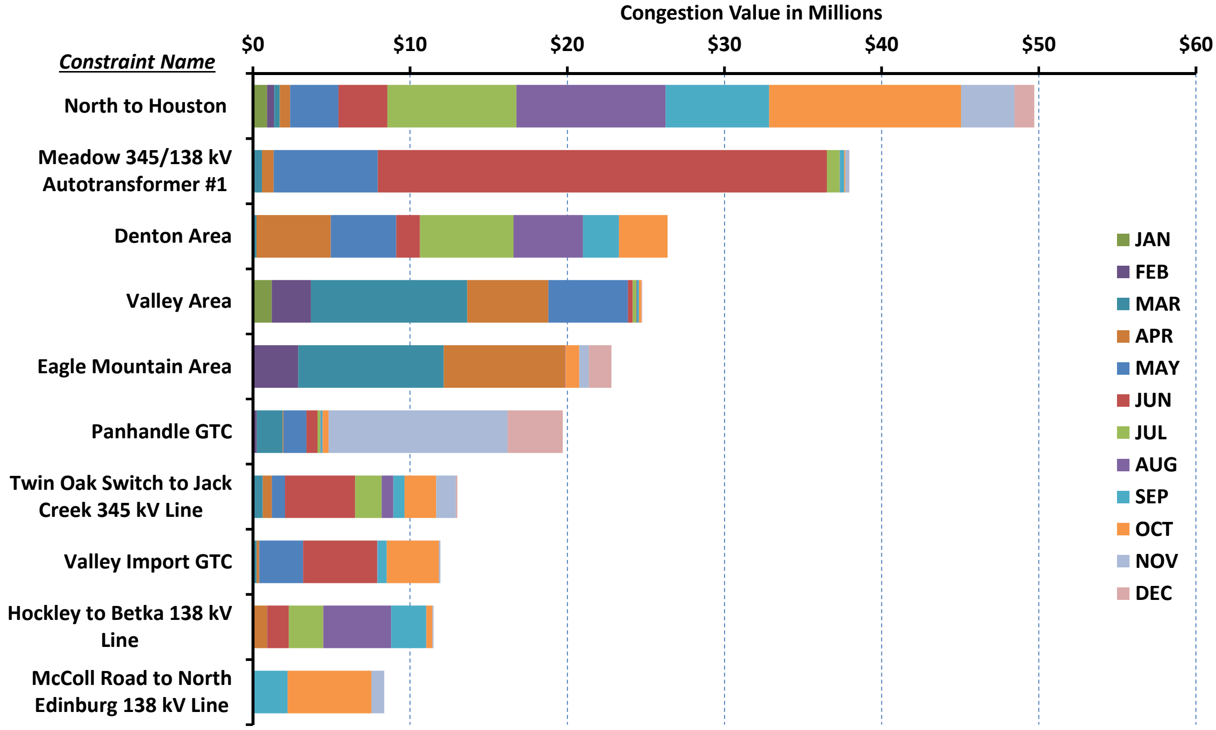
Five of the ten most frequently occurring constraints have already been described as costly. They are the Javelina Tap to Molina 138 kV line, Twin Oak Switch to Jack Creek 345 kV line, Panhandle GTC, North-to-Houston import, and the Eagle Mountain area. The rest of the constraints, although frequently occurring, had moderate financial impacts. This occurs if the generation to be re-dispatched is similarly priced.

The Liston GTC is a constraint defined to control the voltage stability limit in the valley near the Liston 138 kV substation, and is expected to be removed in March 2017. The Wirtz to Flat Rock 138 kV line is located northwest of Austin. The Bosque Switch to Rogers Hill 138 kV line feeds into Waco. The Bruni 138/69 kV transformer constraint frequently limits the output from two wind generators located east of Laredo. The Aspermont 138/69 kV transformer located just south of the Panhandle had frequent congestion in September due to outages taken to perform transmission upgrades in the area.

C. Day-Ahead Constraints

This subsection provides a review of the transmission constraints from the day-ahead market. Figure 47 presents the ten most congested areas from the day-ahead market, ranked by their value. Eight of the constraints listed here were previously described in the real-time subsection. 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 day-ahead market similarly to how they transact in real-time, the same transmission constraints are expected to appear in both markets.

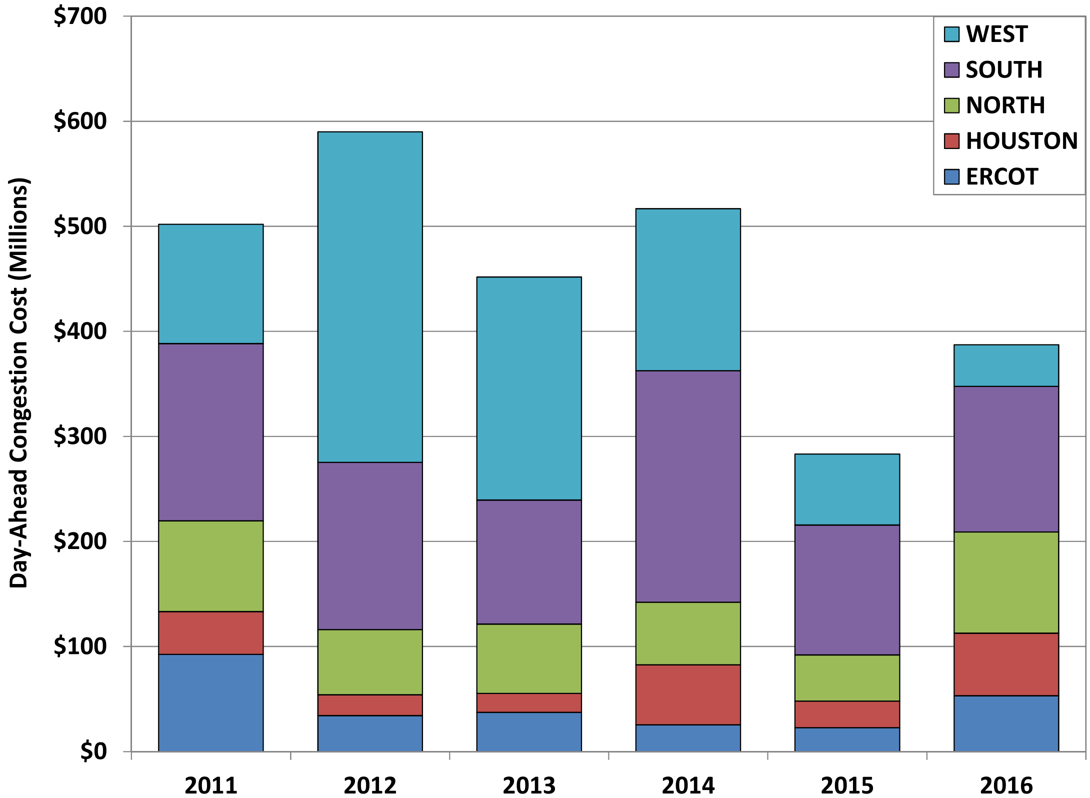
Figure 47: Most Costly Day-Ahead Congested Areas



Since the start of the nodal market, the day-ahead constraint list has contained many constraints that were unlikely to occur in real-time. This is the first year that the majority of the most costly day-ahead constraints were also costly real-time constraints. A contributing factor to this convergence is that ERCOT continually hones the constraint list to monitor which constraints should be included in the day-ahead market analysis to be consistent with market activities observed in real-time.

Located northwest of Houston, the Hockley to Betka constraint was the ninth most costly day-ahead constraint in 2016. While the constraint was not in the top ten real-time constraints, it still had a fairly large real-time price impact of $8 million. The McColl Road to North Edinburg 138 kV line is located on the west side of the Valley, therefore not included in the Valley area description, and was the tenth most costly day-ahead constraint.

Figure 48: Day-Ahead Congestion Costs by Zone



As they were in real-time, day-ahead congestion in the North and Houston zones and across zones (shown as ERCOT) was higher in 2016 than 2015. The total increase in day-ahead congestion costs was approximately 37 percent. The increase in North zone congestion can be explained by Denton-area transmission construction, while Houston import related congestion resulted in the increase in Houston zone and ERCOT congestion. With the completion of the Houston Import project, Houston congestion is expected to decrease in 2018.

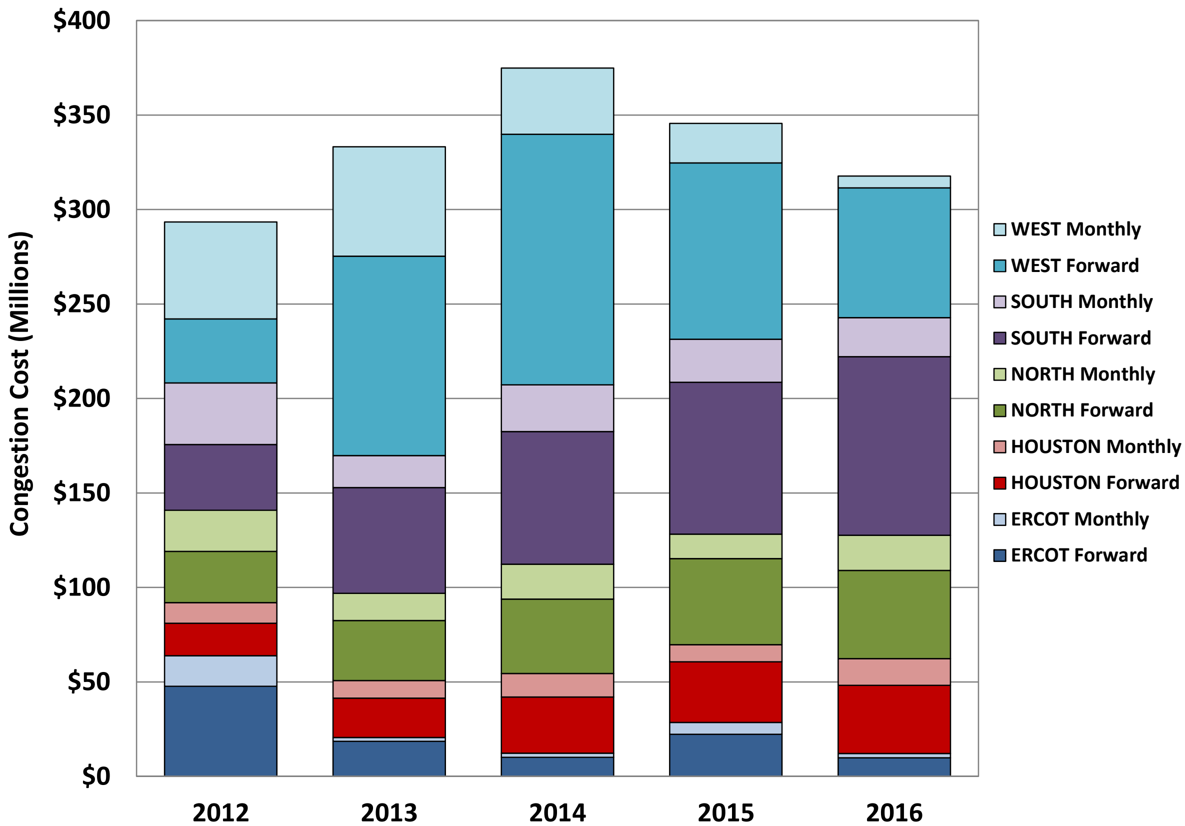
D. Congestion Revenue Rights Market

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 constraints. This causes different clearing prices for energy at different locations. 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.

CRRs may be acquired in semi-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 or charges corresponding to the difference in day-ahead locational prices of the source and sink.

Figure 49 details the congestion cost as calculated by shadow price and flow on binding constraints in the CRR auctions. Note that this calculation, based on the binding constraint location, is similar to the calculation used earlier in this report to display the zonal location of real-time and day-ahead congestion costs and is different from the method used by ERCOT to determine CRR revenue allocation. The costs are broken down by the zonal location of the constraint and whether they were incurred in a monthly auction (Monthly) or a seasonal or annual auction (Forward).

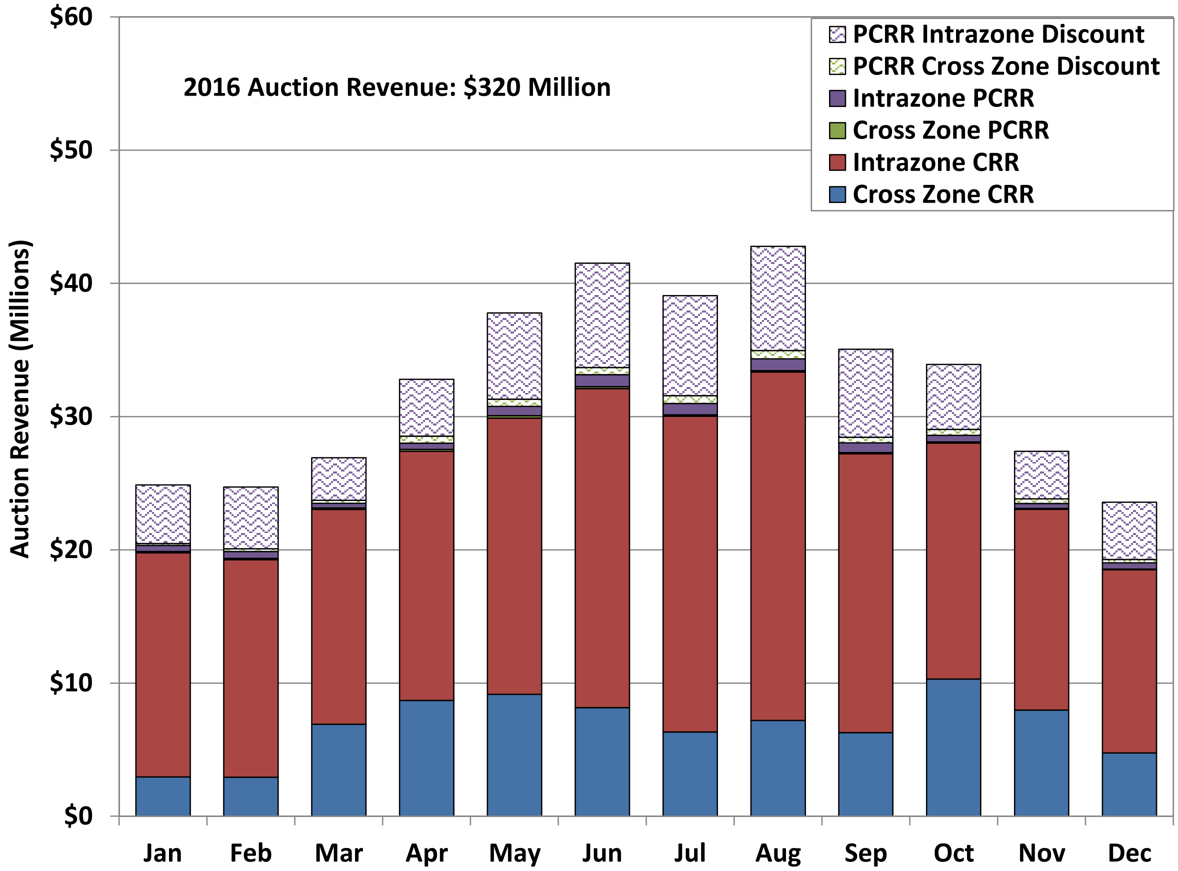
Figure 49: CRR Costs by Zone



Comparing the congestion trends indicated by Figure 49 to the trends seen in the real-time market and the day-ahead market shows that the CRR market did a poor job of predicting the increase of day-ahead (and real-time) congestion. Whereas the congestion costs increased for both the day-ahead and real-time markets compared to 2015, the total CRR congestion decreased. The North and Houston zones saw only slight increases in CRR congestion compared to very large increases in day-ahead and real-time congestion.

Figure 50 summarizes the revenues collected by ERCOT in each month for all CRRs, including both auctioned and allocated. Also shown is the amount of discount provided to the PCRR recipients.

Figure 50: CRR Auction Revenue

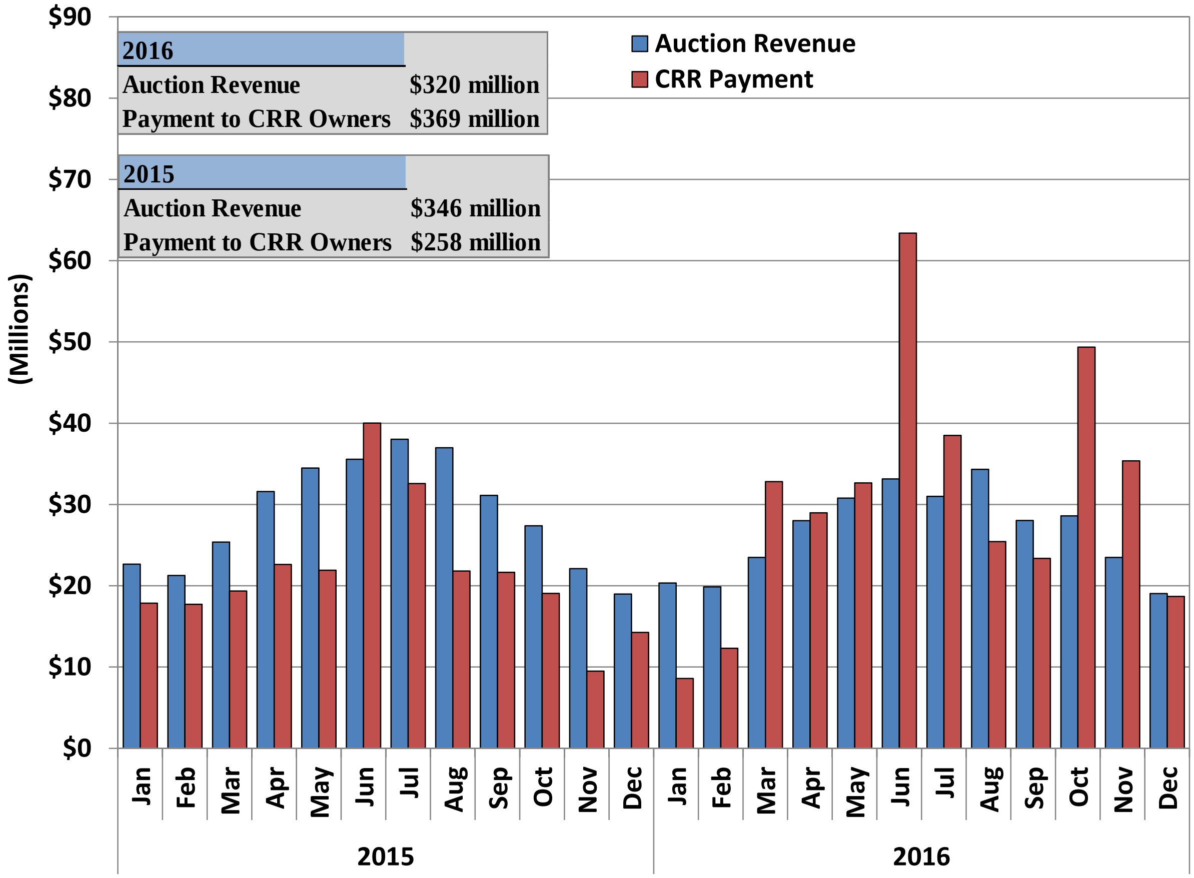


CRR auction revenues are distributed to loads in one of two ways. Revenues from cross-zone CRRs are allocated to loads ERCOT wide. Revenues from CRRs that have the source and sink in the same geographic zone are allocated to loads within that zone. Allocating CRR auction revenues in this manner reduces the net cost for load purchases in heavily-congested areas, but it does so whether the congestion had raised prices in the area or lowered prices in the area. As a case in point, congestion lowered prices in the West zone to below the ERCOT average, as seen above in Figure 5. However, because so many CRRs were purchased in the West zone to capture the value of this price lowering congestion, a higher than load-ratio share portion of the CRR revenue gets distributed to QSEs representing West zone load, thus further lowering the West zone prices.

As previously mentioned in this section, purchasers of PCRRs are only charged a fraction of the PCRR auction value. The difference between the auction value and the value charged to the purchaser is shown in Figure 50 as the PCRR Discount. Even as the total amount of CRR auction revenue dropped to $320 million in 2016 from $346 million in 2015, the total PCRR discount increased from $49 million in 2015 to $70 million in 2016.

Next, Figure 51 compares the value received by CRR owners (in aggregate) to the price paid to acquire the CRRs.

Figure 51: CRR Auction Revenue and Payment Received

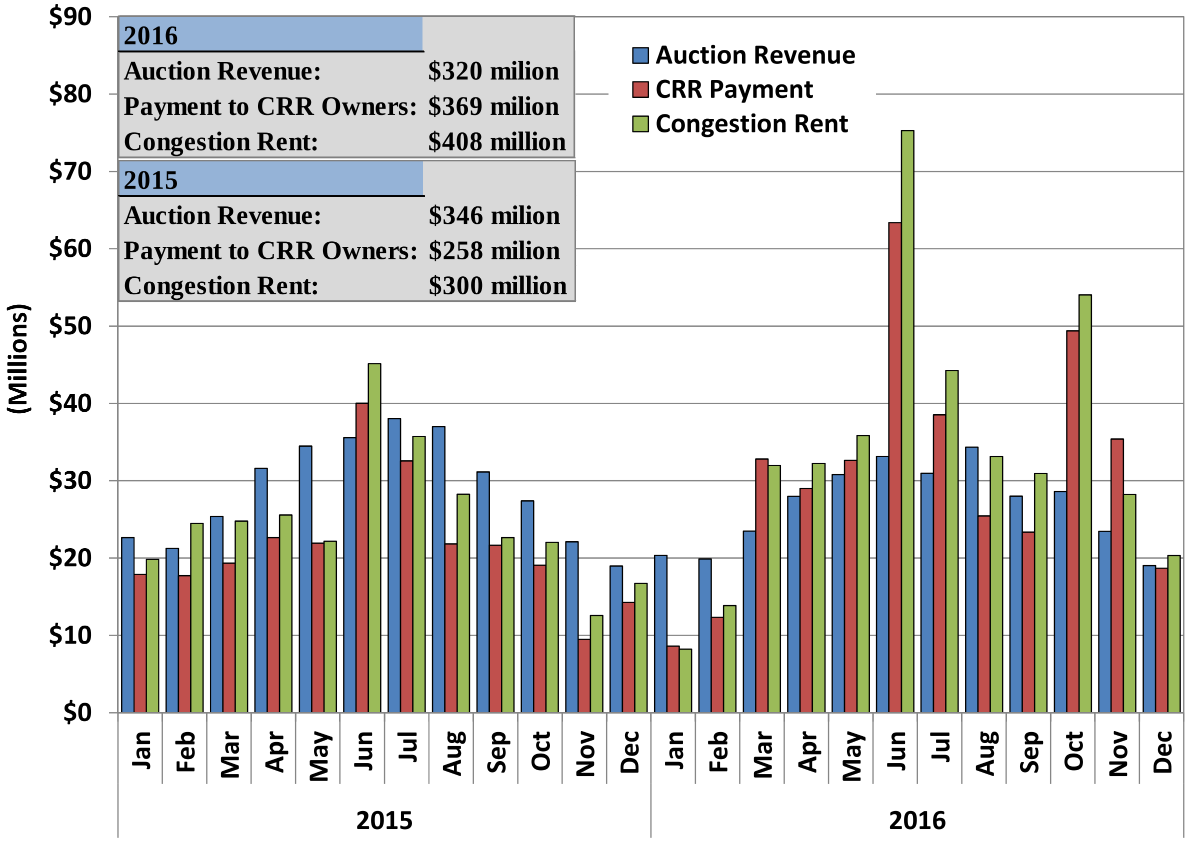


Although results for individual participants and specific source/sink combinations varied, the aggregated results for the year and in most months show that participants paid less for CRRs in 2016 than they received in payment from the day-ahead market, though it is worth noting that if NOIEs had paid full price for PCRRs the total net procurement cost would have exceeded the receipts. For the entire year of 2016 participants spent $320 million to procure CRRs and received $369 million.

The next analysis of aggregated CRR positions adds day-ahead congestion rent to the picture. Day-ahead congestion rent is the difference between the total costs that loads pay and the total revenue that generators receive in the day-ahead market. Day-ahead congestion rent creates the source of funds used to make payments to CRR owners. Figure 52 presents CRR auction revenues, payment to CRR owners, and day-ahead congestion rent in 2015 and 2016, by month.

Congestion rent for the year 2016 totaled $408 million and payment to CRR owners was $369 million. It is worth noting that, since the CRR network model uses line ratings that are 90 percent of the expected lowest line ratings for the month, one would expect that CRRs would be somewhat undersold and that day-ahead congestion rent would be higher than the payment to CRR owners.

Figure 52: CRR Auction Revenue, Payments and Congestion Rent



The target value of a CRR is the megawatt amount of the CRR multiplied by the locational marginal price (LMP) of the sink of the CRR less the LMP of the source of the CRR. While the target value is paid to CRR account holders most of the time, there are two circumstances where an amount less than the target value is paid. The first circumstance happens when the CRR is modeled on the day-ahead network and causes a flow on a transmission line that exceeds the line's limit. In this case, CRRs with a positive value that have a source and/or a sink located at a resource node settlement point are often derated, that is, paid a lower amount than the target value.

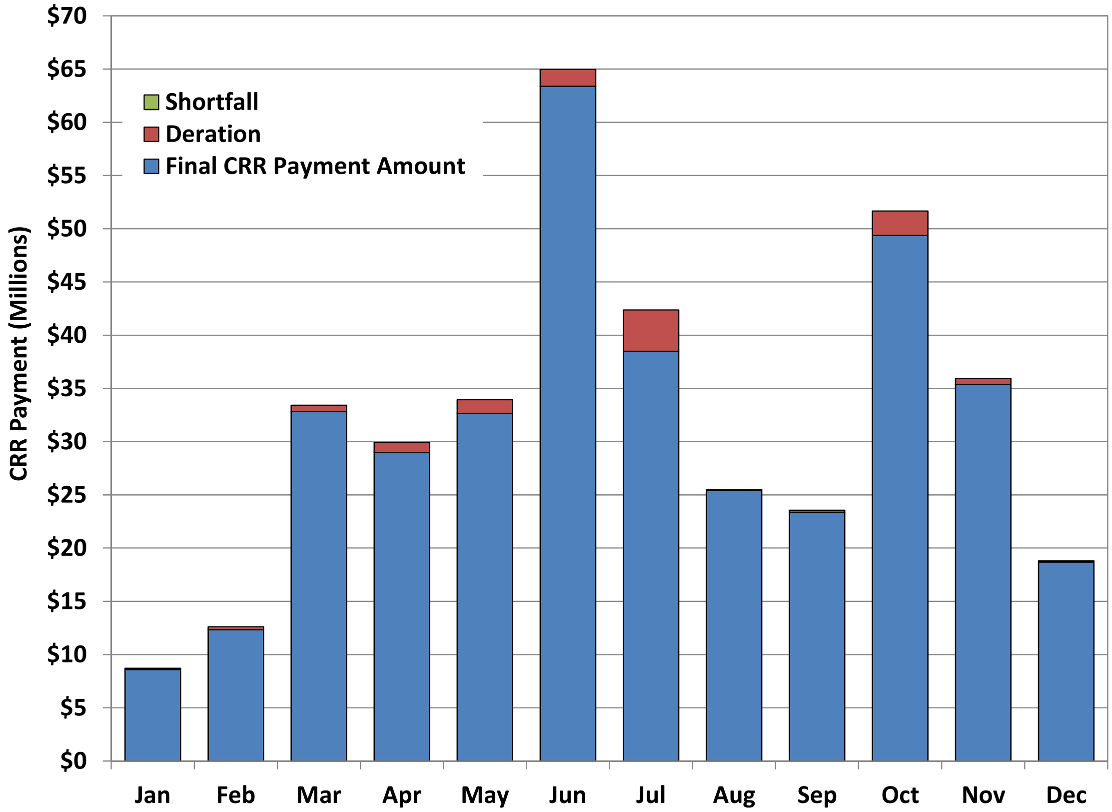
The second circumstance occurs when there is not enough day-ahead congestion rent to pay all the CRRs at target (or derated, if applicable) value. In this case, all holders of positively valued CRRs receive a prorated shortfall charge such that the congestion revenue plus the shortfall charge can pay all CRRs at target or derated value. This shortfall charge has the effect of lowering the net amount paid to CRR account holders; however, if at the end of the month there is excess day-ahead congestion rent that has not been paid out to CRR account holders, the excess congestion rent can be used to make whole the CRR account holders that received shortfall charges. If there is not enough excess congestion rent from the month, the rolling CRR balancing fund can be drawn upon to make whole CRR account holders that received shortfall charges.

The rolling CRR balancing fund began in December 2014, thus 2016 provides the second full year to review its performance.[21] The CRR balancing fund started the year at its capped value of $10 million and was drawn upon once to cover a shortfall of $5.7 million in November. With $762 thousand added back to the fund in December, it ended the year with a balance of $5.1 million.

Figure 53 shows the amount of target payment, deration amount, and net shortfall charges (after make whole payments) for 2016. In 2016 the total target payment to CRRs was $381 million; however, there were $12 million of derations and no shortfall charges leaving a final payment to CRR account holders of $369 million. This corresponds to a CRR funding percentage of 97 percent.

21 The CRR Balancing Fund was implemented with NPRR580.

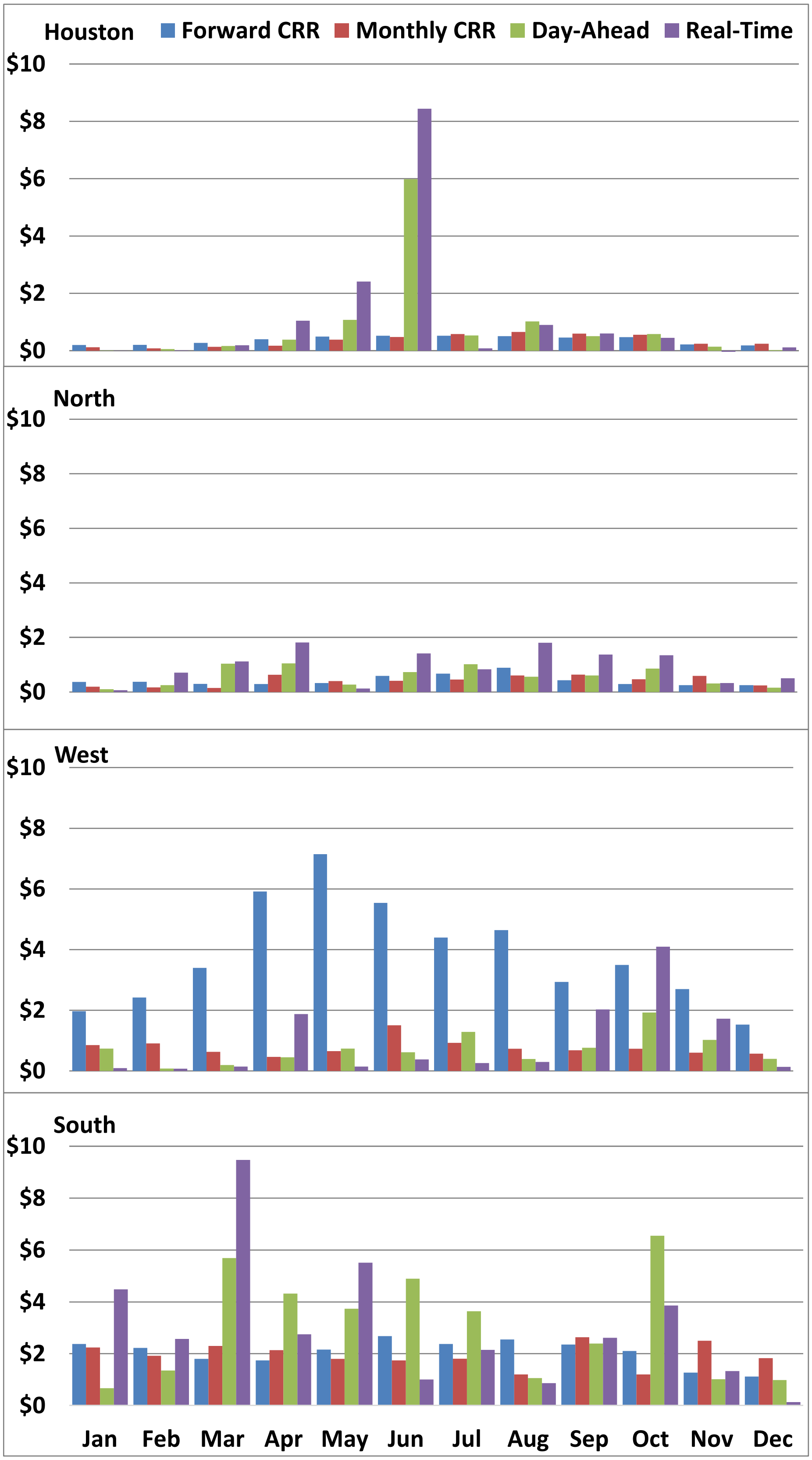
Figure 53: CRR Shortfalls and Derations



The last look at congestion examines the price spreads for each pair of hub and Load Zones in more detail. These price spreads are interesting as many loads may have contracts that hedge to the hub price and are thus exposed to the price differential between the hub and its corresponding Load Zone. Figure 54 presents the price spreads between all Hub and Load Zones as valued at four separate points in time – at the average of the four semi-annual CRR Auctions, monthly CRR auction, day-ahead and real-time.

Of note is the relatively poor convergence between the forward CRR price spreads for the West Load Zone and the actual price spreads. This may be due to the difficulty forecasting the price impacts of variable wind output. Also noteworthy is that the South Load Zone has overtaken the West Load Zone to become the Zone with the highest Hub to Zone price spread. This is likely due to congestion in the Valley area.

Figure 54: Hub to Load Zone Price Spreads

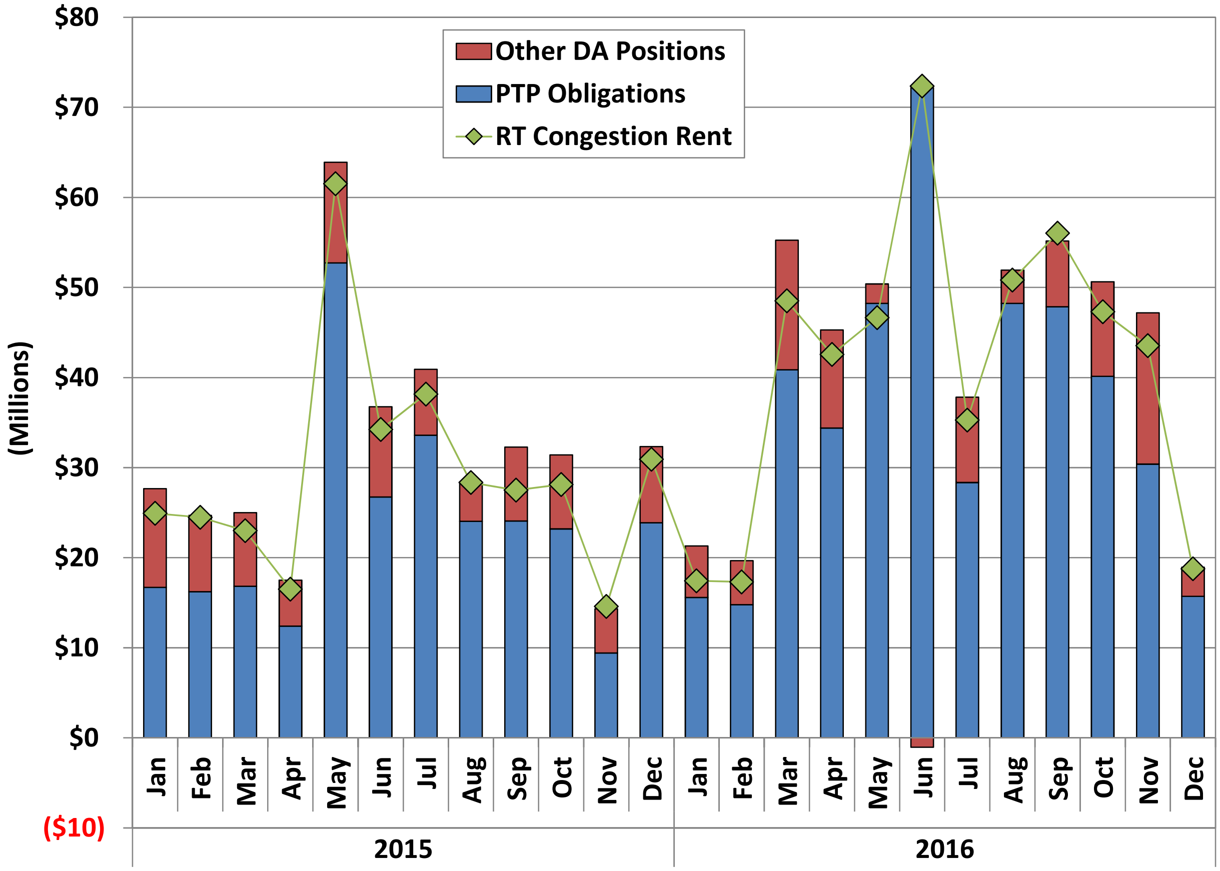


E. Revenue Sufficiency

In Figure 55 the combined payments to Point-to-Point (PTP) obligation owners and effective payments to other day-ahead positions are compared to the total real-time congestion rent. For 2016, real-time congestion rent was $497 million, payments for PTP obligations (including those with links to CRR options) were $437 million and payments for other day-ahead positions were $88 million, resulting in a shortfall of approximately $28 million for the year.

By comparison, the real-time congestion rent was $352 million in 2015. Payments for PTP obligations and real-time CRRs were $280 million and payments for other day-ahead positions were $95 million, resulting in a shortfall of approximately $23 million for the year.

Figure 55: Real-Time Congestion Rent and Payments



IV. Demand and Supply

This section reviews and analyzes the load patterns during 2016 and the existing generating capacity available to satisfy the load and operating reserve requirements. Specific analysis of the large quantity of installed wind generation is included, along with a discussion of the daily generation commitment characteristics. This section concludes with a discussion of demand response resources.

A. ERCOT Load in 2016

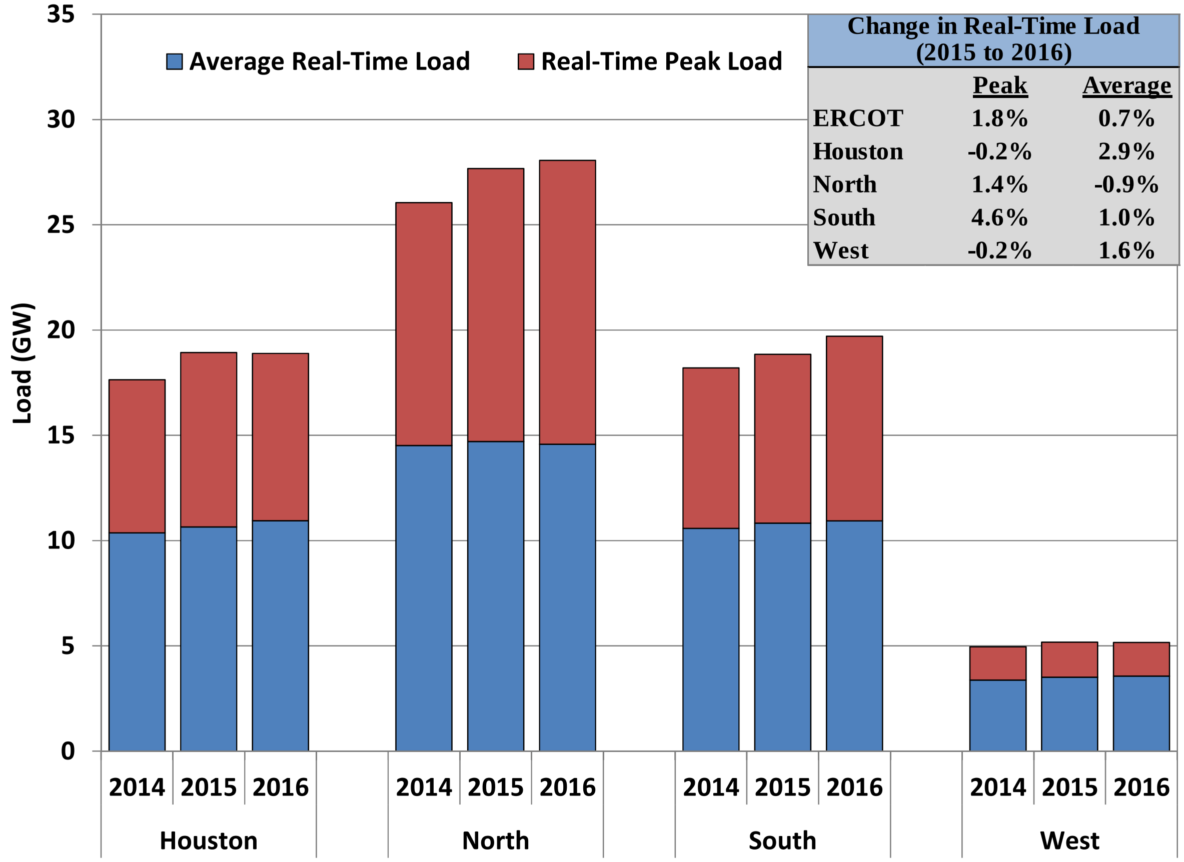
The changes in overall load levels from year to year can be shown by tracking the changes in average load levels. This metric tends to capture changes in load over a large portion of the hours during the year. Separately evaluating the changes in the load during the highest-demand hours of the year is also important. Significant changes in peak demand levels play a major role in assessing the need for new resources. The level of peak demand also affects 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 or inconvenience to customers that can result from interruptions to that service. Hence, both of these dimensions of load during 2016 are examined in this subsection and summarized in Figure 56.

This figure shows peak load and average load in each of the ERCOT geographic zones from 2014 to 2016.[22] In each zone, as in most electrical systems, peak demand significantly exceeds average demand. The North zone is the largest zone (with about 37 percent of the total ERCOT load); the South and Houston zones are comparable (27 percent) while the West zone is the smallest (9 percent of the total ERCOT load).

Figure 56 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 is greater than the annual ERCOT peak load.

22 For purposes of this analysis, Non-Opt In Entity (NOIE) Load Zones have been included with the proximate geographic zone.

Figure 56: Annual Load Statistics by Zone

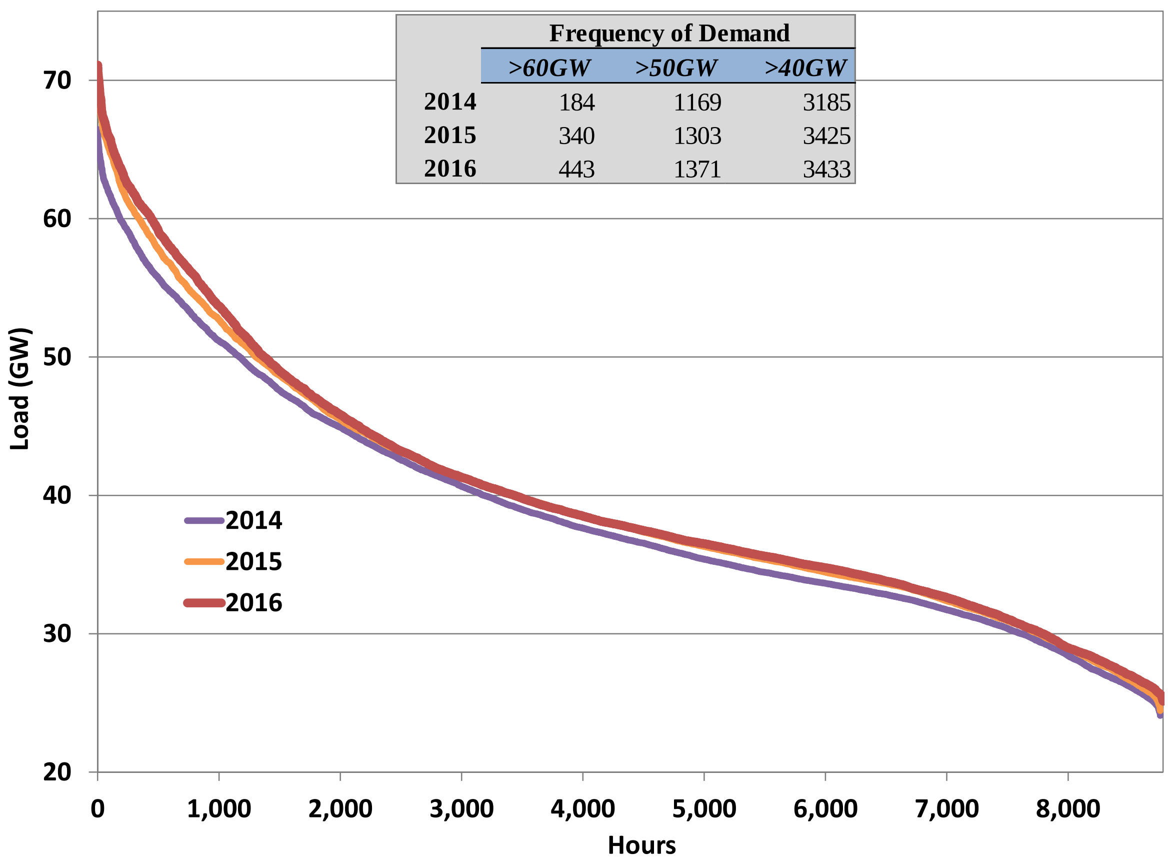


Total ERCOT load over the calendar year increased 1.1 percent (approximately 450 MW on average) to total 351.5 TWh in 2016. As 2016 was a leap year, the relative increase in the total load is higher than the increase in average load. With the exception of the North zone, all zones showed an increase in average real-time load in 2016. Houston saw the largest average load increase at 2.9 percent. Changes in average loads were largely explained by summer weather. Cooling degree days, a metric that is highly correlated with weather-related summer load, increased 4 percent on average from 2015 to 2016 in Houston and decreased 3 percent in Dallas. However, cooling degree days in 2016 were still 12 to 16 percent lower than ERCOT's hottest recent summer in 2011.

Summer conditions in 2016 also led to a new ERCOT-wide coincident peak hourly demand record of 71,110 MW on August 11, 2016. This broke the prior year's peak demand record of 69,877 MW that occurred on August 10, 2015. In fact, demand exceeded 70,000 MW five different times in 2016. The 2016 peak represents a 1.8 percent increase from the peak hourly demand of 2015. The zones experienced varying changes in peak load. Although the West zone had shown a prior trend of increasing load due to oil and gas production activity, that trend reversed in 2016 with a decrease in West zone peak load corresponding with a decline in oil and gas activity. Houston also showed a decrease in peak load. The South zone had the greatest increase in peak load at 4.6 percent.

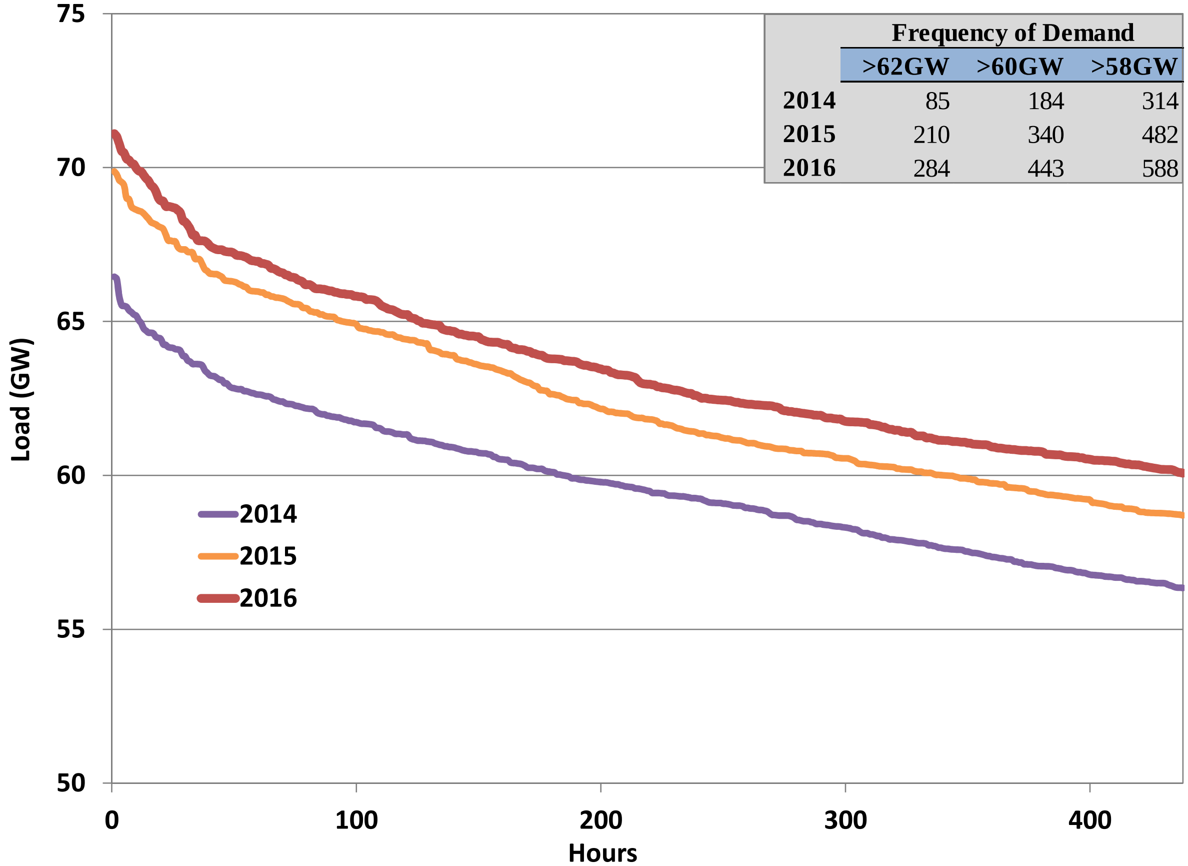
To provide a more detailed analysis of load at the hourly level, Figure 57 compares load duration curves for each year from 2014 to 2016. A load duration curve illustrates the number of hours (shown on the horizontal axis) that load exceeds a particular level (shown on the vertical axis). ERCOT has a 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. The load duration curve in 2016 is very similar to 2015, with a slight increase in the hours at the highest load levels.

Figure 57: Load Duration Curve – All Hours



To better illustrate the differences in the highest-demand periods between years, Figure 58 below shows the load duration curve for the 5 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 2011 to 2016, the peak load averaged 18 percent greater than the load at the 95th percentile. These load characteristics imply that a substantial amount of capacity – more than 10 GW – is needed to supply energy in less than 5 percent of the hours.

Figure 58: Load Duration Curve – Top Five Percent of Hours

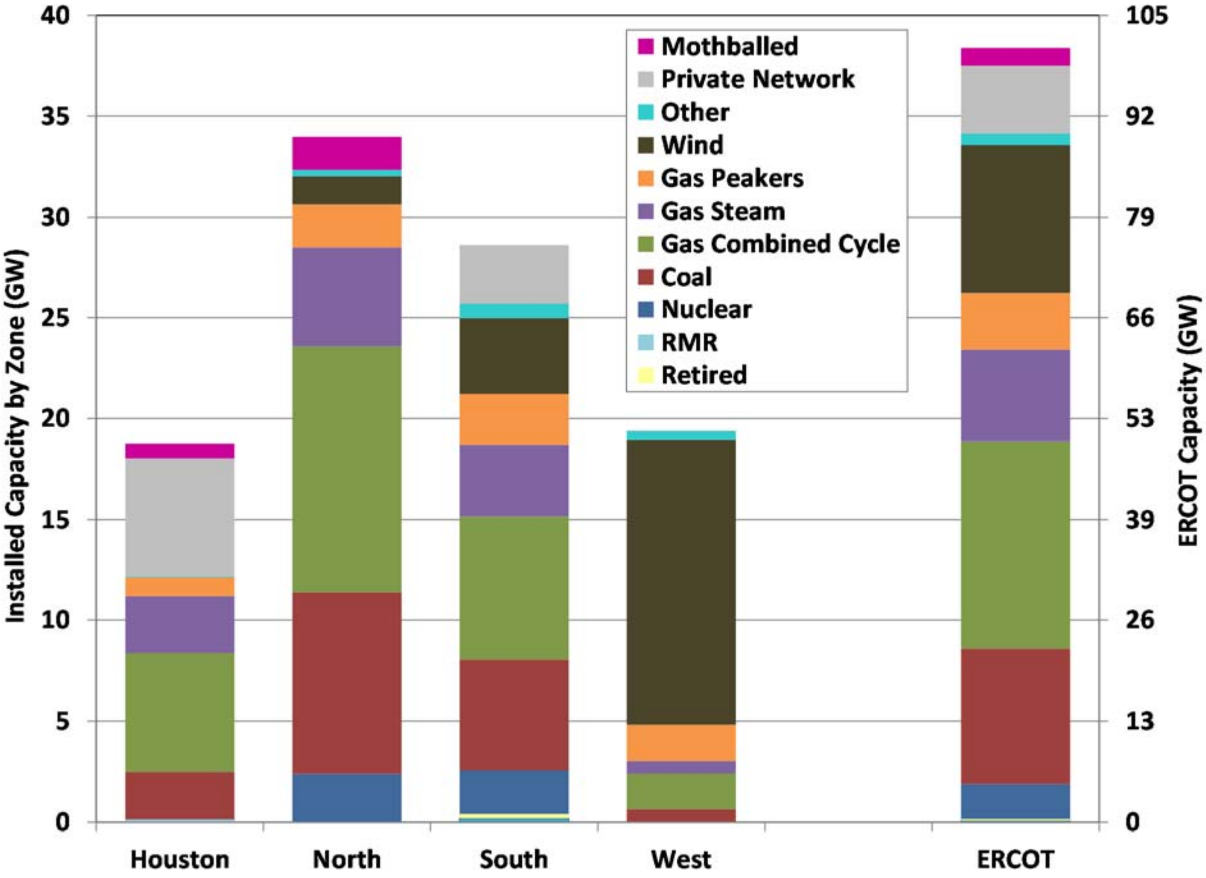


B. Generation Capacity in ERCOT

The generation mix in ERCOT is evaluated in this subsection. The distribution of capacity among the four ERCOT geographic zones is similar to the distribution of demand with the exception of the large amount of wind capacity in the West. The North zone accounts for approximately 33 percent of capacity, the South zone 29 percent, the Houston zone 19 percent, and the West zone 19 percent. Excluding mothballed resources and including only the fraction of wind capacity available to reliably meet peak demand,[23] the North zone accounts for approximately 37 percent of capacity, the South zone 32 percent, the Houston zone 22 percent, and the West zone 9 percent. Figure 59 shows the installed generating capacity by type in each zone.

23 The percentages of installed capacity to serve peak demand assume wind availability of 14 percent for non-coastal wind and 58 percent for coastal wind.

Figure 59: Installed Capacity by Technology for Each Zone



Approximately 5.5 GW of new generation resources came online in 2016, but it only provided roughly 2 GW of net effective capacity. The overwhelming majority of new capacity was from wind generation. The 4.1 GW of newly installed wind capacity provides approximately 645 MW of capacity at summer peak. The remaining 1.4 GW of new capacity consisted of 370 MW of solar resources, 10 MW of storage resources, and approximately 1 GW of new natural gas combined-cycle units. Although still a small portion of the newly installed capacity, the installed solar megawatts in 2016 were more than three times the amount added in the prior year.

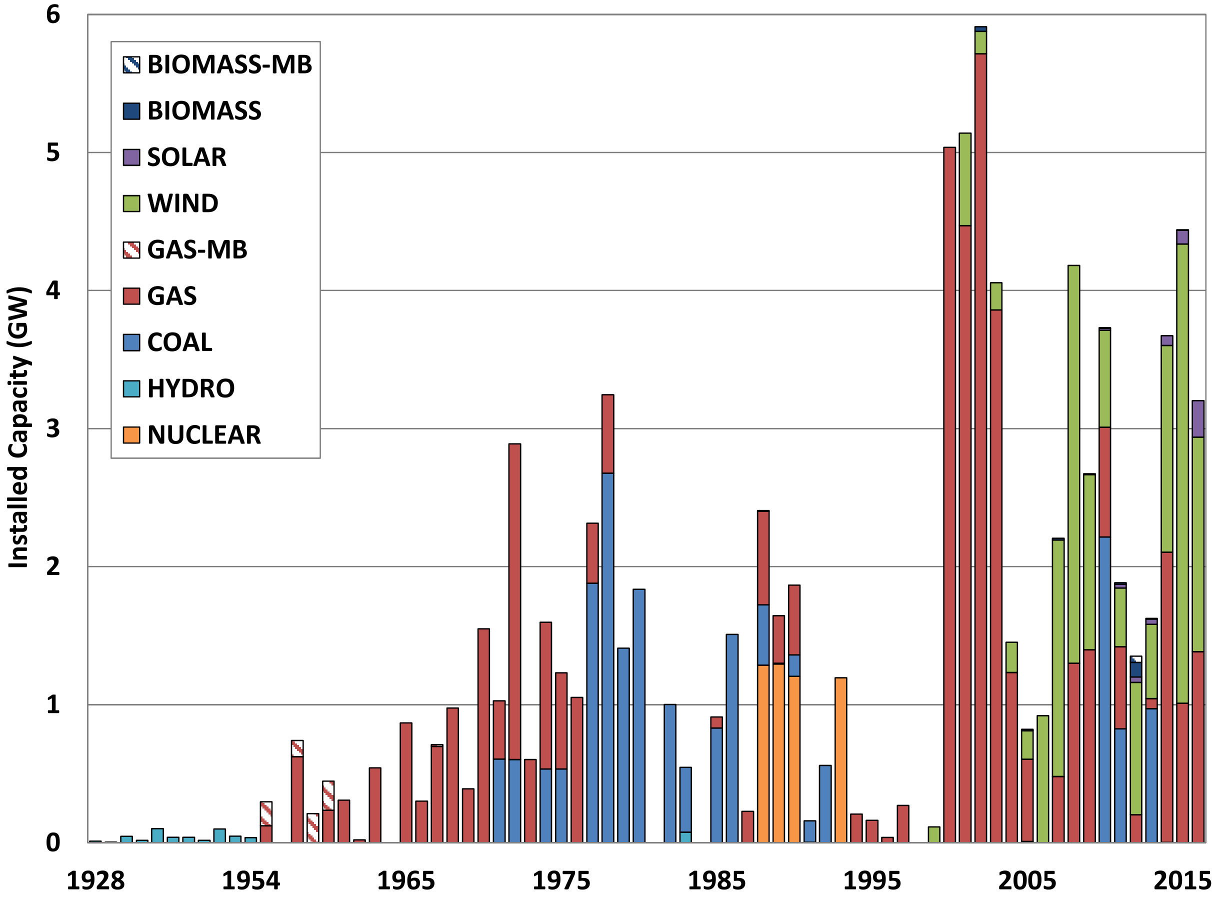
Considering these additions and retirements in 2016, natural gas generation decreased slightly from 48 percent of total ERCOT installed capacity in 2015 to 45 percent in 2016. The share of total installed capacity for coal generation also decreased slightly from 20 percent in 2015 to 17 percent in 2016.

Figure 60 shows the age of generation resources in ERCOT that were operational in the December 2016 Capacity, Demand, and Reserves Report.[24] The bulk of the coal fleet in ERCOT

[24] ERCOT Capacity, Demand, and Reserves Report (Dec. 2016), available at http://www.ercot.com/gridinfo/resource.

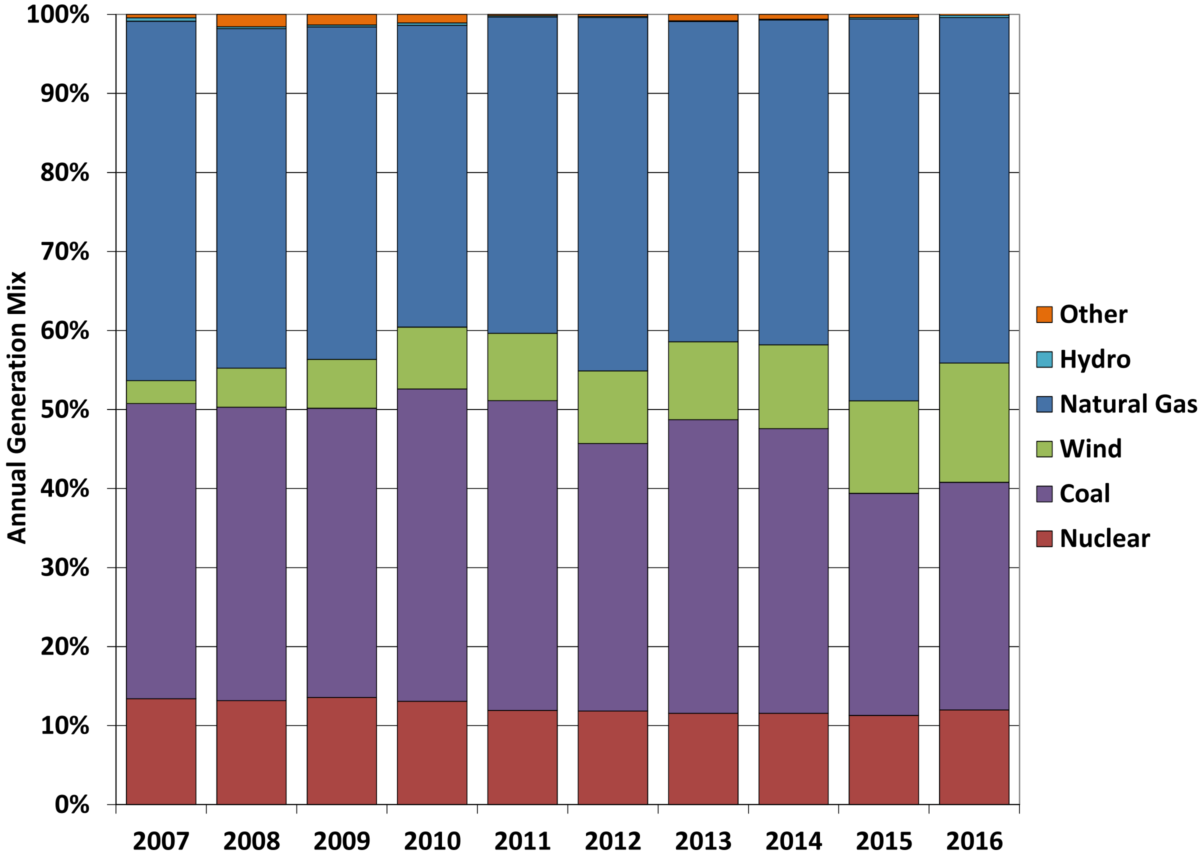
was built before 1990 and is approaching the end of useful life for this vintage of coal-fired power plants. When the ERCOT market was deregulated, there was a large increase in the construction of combined-cycle gas units. A few new coal units were added around 2010. As the figure demonstrates, wind capacity has been the dominant technology for newly installed capacity since 2006.

Figure 60: Vintage of ERCOT Installed Capacity



The shifting contribution of coal and wind generation is evident in Figure 61, which shows the percent of annual generation from each fuel type for the years 2007 through 2016.

Figure 61: Annual Generation Mix

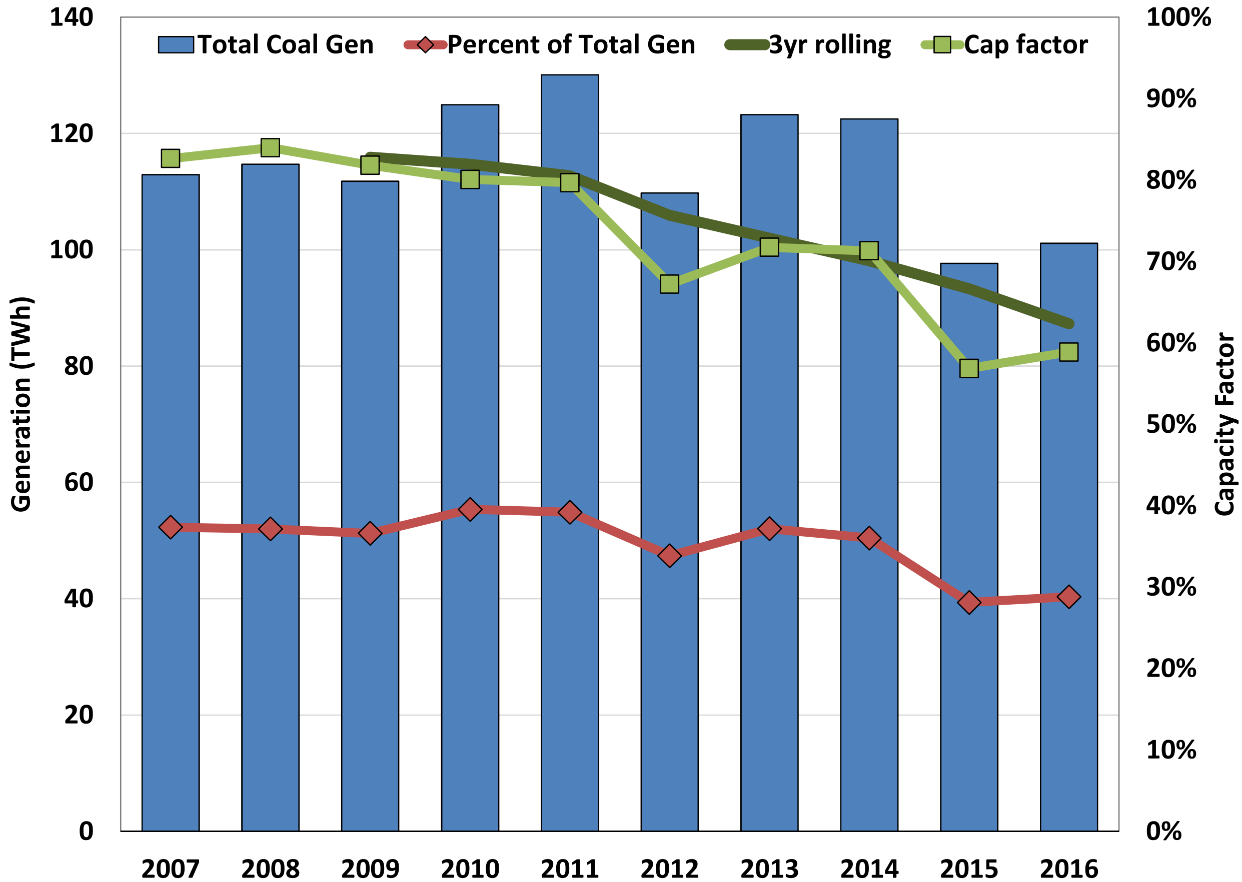


The generation share from wind has increased every year, reaching 15 percent of the annual generation requirement in 2016, up from 3 percent in 2007 and 12 percent in 2015. While the percent of generation from coal had declined significantly between 2014 and 2015, its share increased slightly to 29 percent in 2016. Natural gas declined from its high point in 2015 at 48 percent to 44 percent in 2016.

While coal/lignite and nuclear plants operate primarily as base load units in ERCOT, it is the reliance on natural gas resources that drives the high correlation between real-time energy prices and the price of natural gas fuel. There are approximately 24 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.

Figure 62 shows the total coal generation, percent of total generation by coal, and the capacity factor for coal in years 2007 through 2016. The chart includes the annual capacity factor as well as the three-year rolling average capacity factor. While there was a slight increase in the coal capacity factor between 2015 and 2016, the three-year rolling average demonstrates the long-term decline in the coal capacity factor in ERCOT.

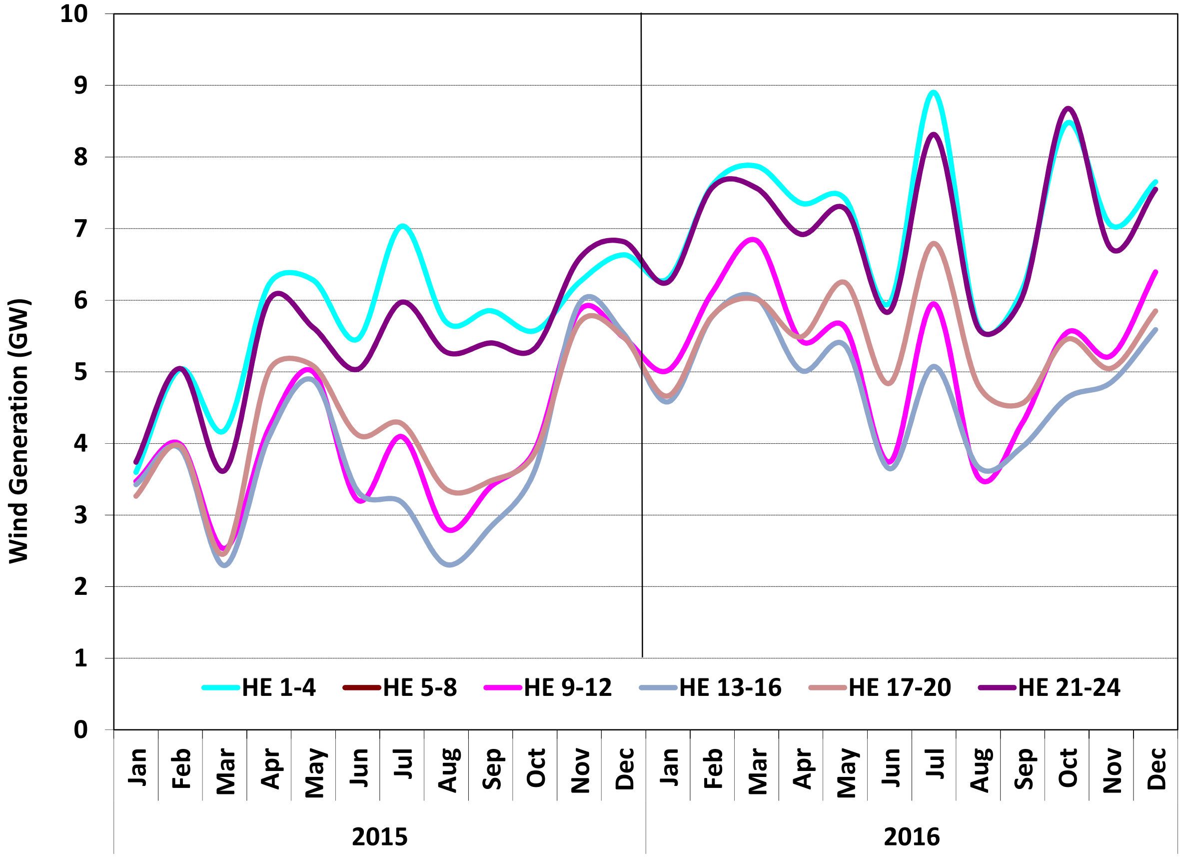
Figure 62: Historic Coal Generation and Capacity Factor



The amount of wind generation installed in ERCOT was approximately 19 GW by the end of 2016. Although the large majority of wind generation is located in the West zone, more than 3 GW of wind generation has been located in the South zone. Additionally, a private transmission line that went into service in late 2010 allows another nearly 1 GW of West zone wind to be delivered directly to the South zone. In 2007, wind generation in ERCOT was located in 14 counties; by 2016, there were more than 50 counties with wind generators serving ERCOT.

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 63 shows average wind production for each month in 2015 and 2016, with the average production in each month divided into four-hour blocks. Though the lowest wind output generally occurs during summer afternoons, there has been such a large amount of wind generation added in ERCOT that the average wind output during summer peak period now averages in excess of 4 GW. This may be a small fraction of the total installed capacity but is now a non-trivial portion of generation supply, even at its lowest outputs.

Figure 63: Average Wind Production

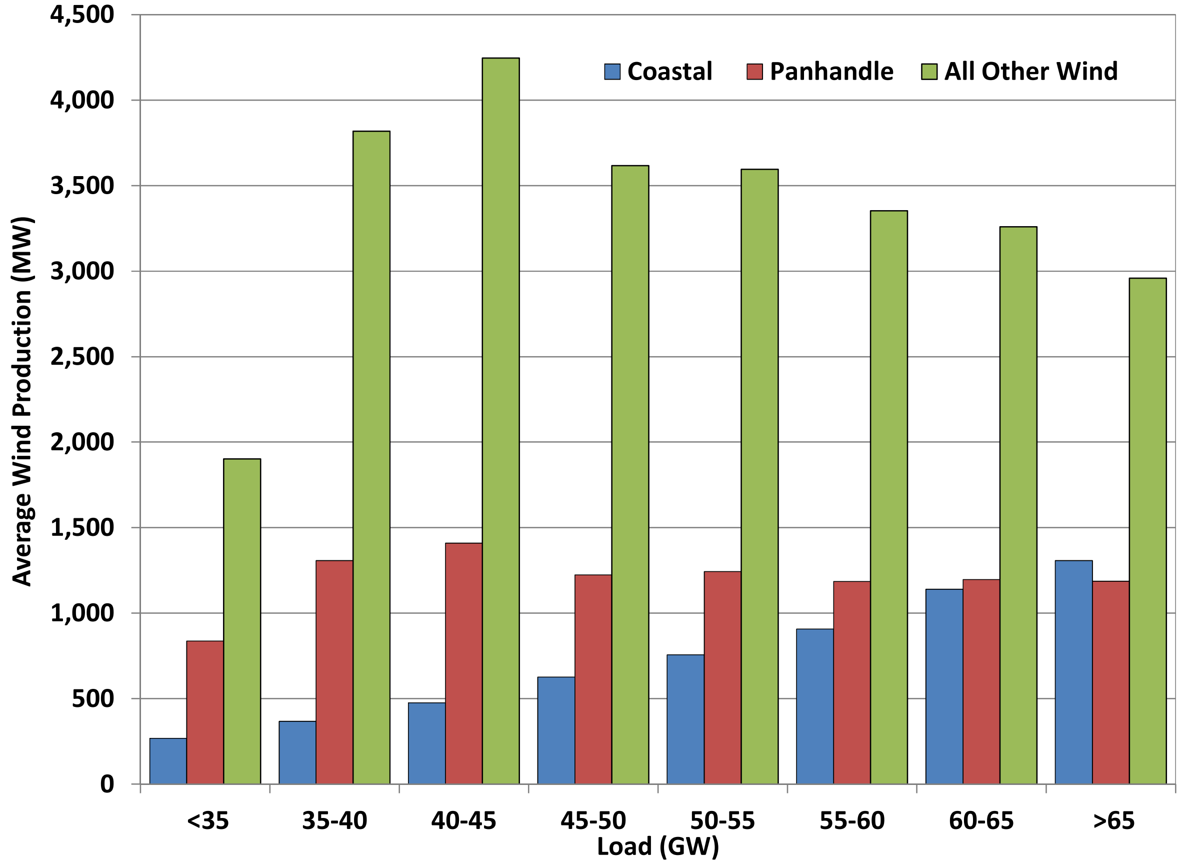


ERCOT continued to set new records for peak wind output in 2016. On December 25, wind output exceeded 16 GW, setting the record for maximum output and serving nearly 47 percent of the total load.

Examining wind generation in total masks the different wind profiles that exist for locations across ERCOT. The attraction to sites along the Gulf Coast of Texas is due to the higher correlation of the wind resource in that location with electricity demand. More recently, the Texas Panhandle has attracted wind developer interest due to its abundant wind resources. The differences in output for wind units located in the coastal area of the South zone and those located elsewhere in ERCOT are compared below.

Figure 64 below presents data for the summer months of June through August, comparing the average output for wind generators located in the coastal region, the Panhandle and all other areas in ERCOT across various load levels.

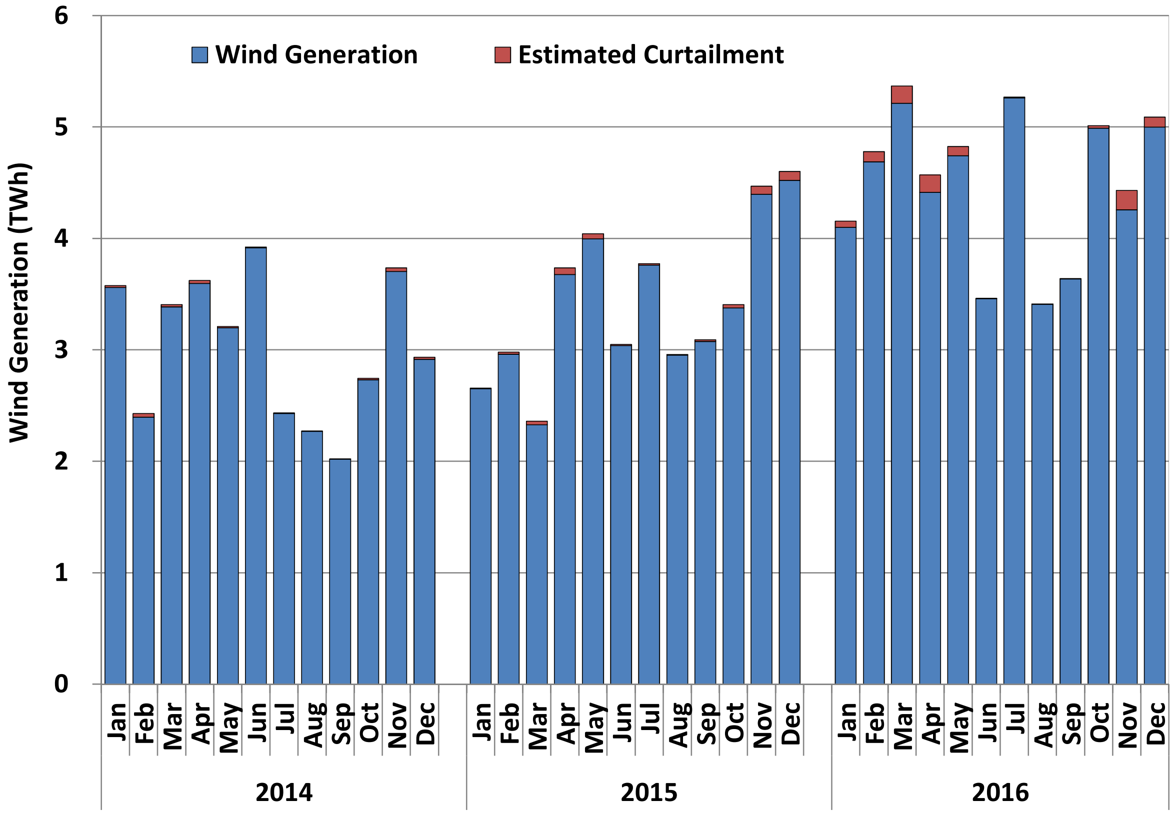
Figure 64: Summer Wind Production vs. Load



The typical profile for wind units not located along the coast or in the panhandle is negatively correlated with peak electricity demand. However, output from wind generators located in the coastal area of the South zone is much more highly correlated with peak electricity demand. Panhandle wind shows a more stable output across the load levels.

Figure 65 shows the wind production and estimated curtailment quantities for each month of 2013 through 2016.

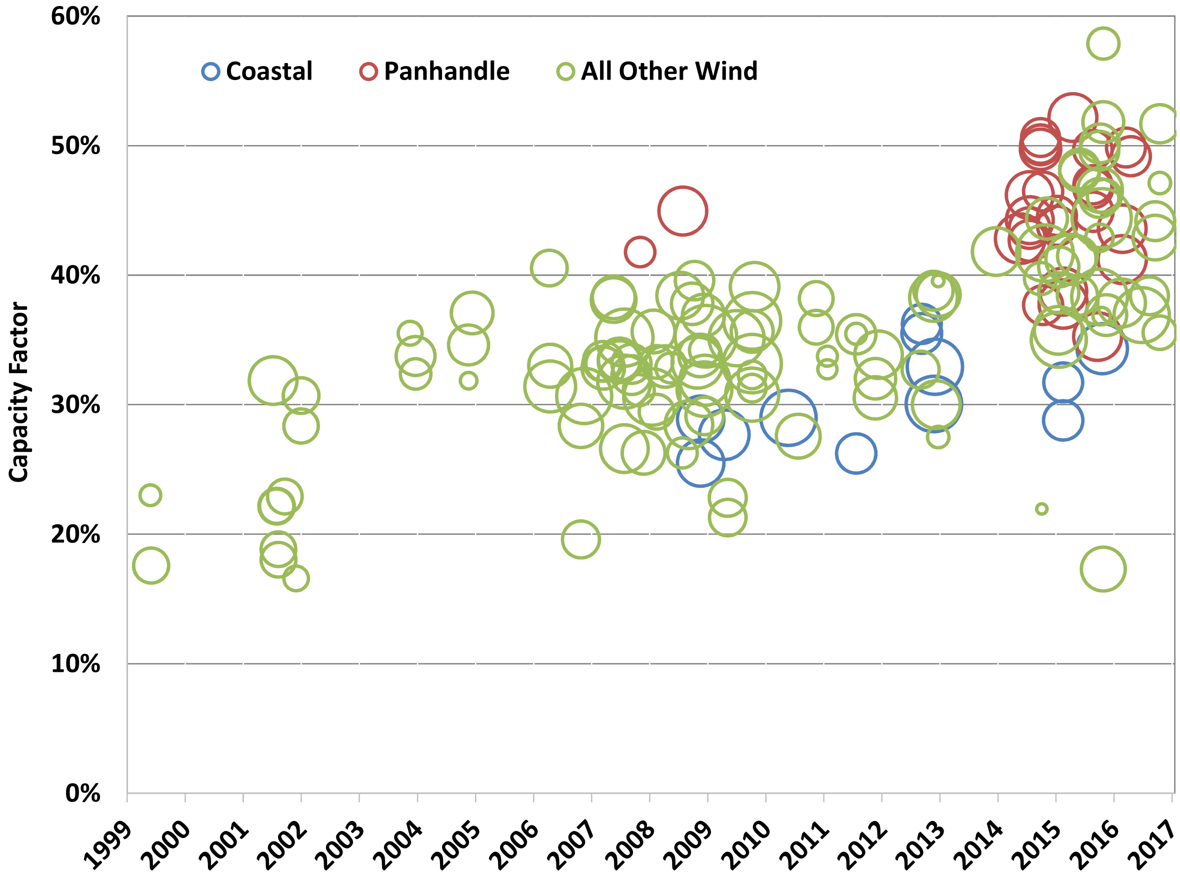
Figure 65: Wind Production and Curtailment



This figure reveals that the total production from wind resources continued to increase, while the quantity of curtailments also increased. The volume of wind actually produced in 2016 was estimated at 98 percent of the total available wind, compared with 99 percent in 2015 and 99.5 percent in 2014. As a comparison, in 2009, the year with the most wind curtailment, the amount of wind delivered was only 83 percent.

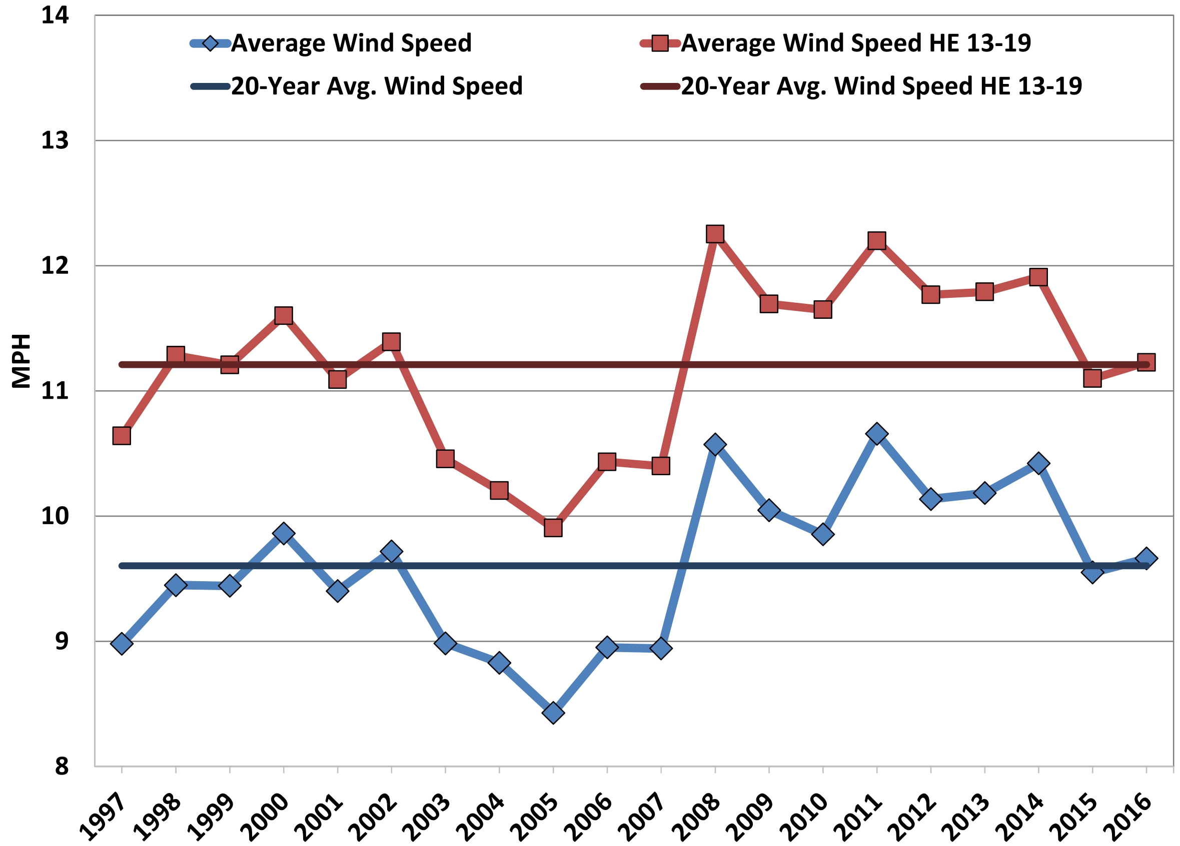
Figure 66 shows the capacity factor for wind generators based on the year installed. Wind generation units located along the coast and in the panhandle are depicted with different colors because of the different wind profiles for these regions. Coastal wind generally has a lower annual capacity factor, but as previously described their output is generally more coincident with summer peak loads. Completion of CREZ transmission lines has enabled more wind units to locate in the windier Panhandle area. The figure also shows a trend toward greater capacity factors for newer units.

Figure 66: Wind Generator Capacity Factor by Year Installed



The next figure shows average wind speeds in ERCOT, weighted by the current installed wind generation locations. Figure 67 provides a picture of the wind supply in 2016, averaged across the year and the average during peak hours, compared to the previous 20 years. The wind supply in 2016 was similar to the average over the past 20 years for all hours and for the peak hours ending 13-19. With 2016 being an average wind supply year, if the existing fleet of wind generation had existed in prior years, total wind production could have been much greater. Notably, one of the years with higher than average wind speeds was 2011.

Figure 67: Historic Average Wind Speed



Increasing wind output also has important implications for the net load served by non-wind resources. Net load is the system load minus wind production. Figure 68 shows the net load duration curves for the years 2007, 2011, and 2016.

Figure 68 shows the reduction of remaining energy available for non-wind units to serve during most hours of the year, even after factoring in several years of load growth. The impact of wind on the highest net load values is much smaller.

Figure 68: Net Load Duration Curves

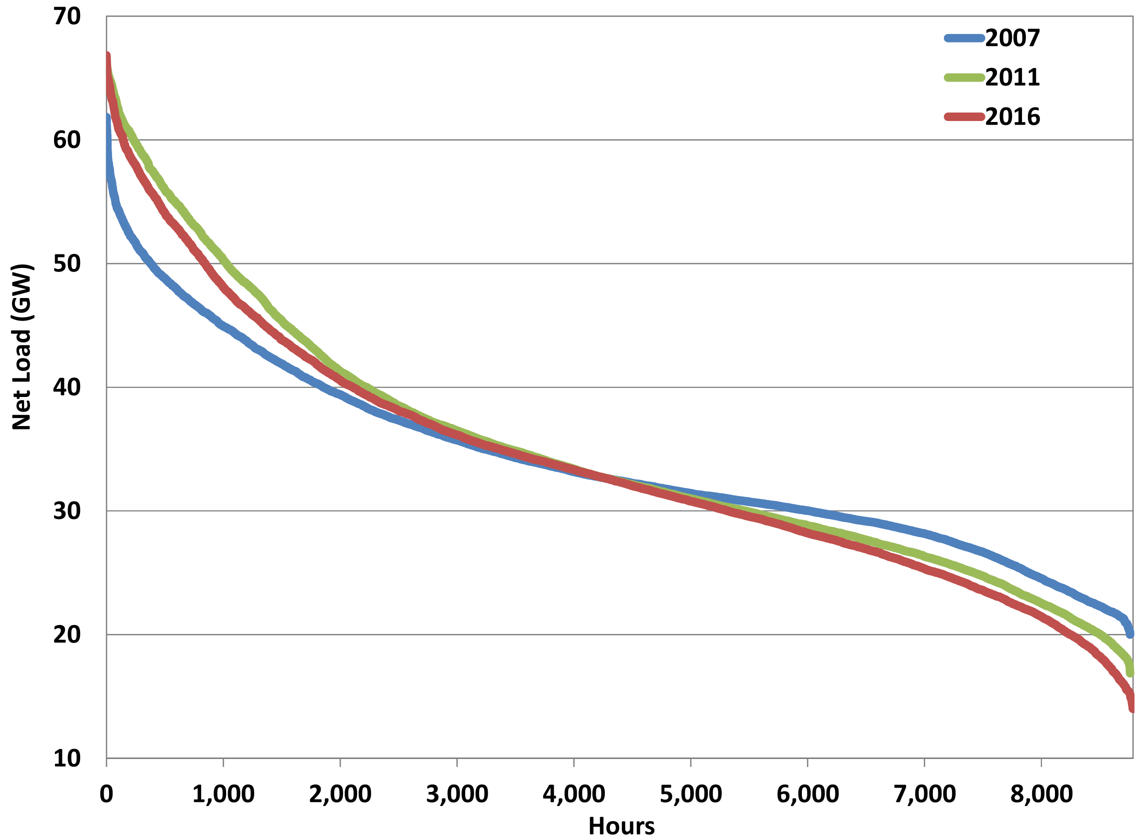
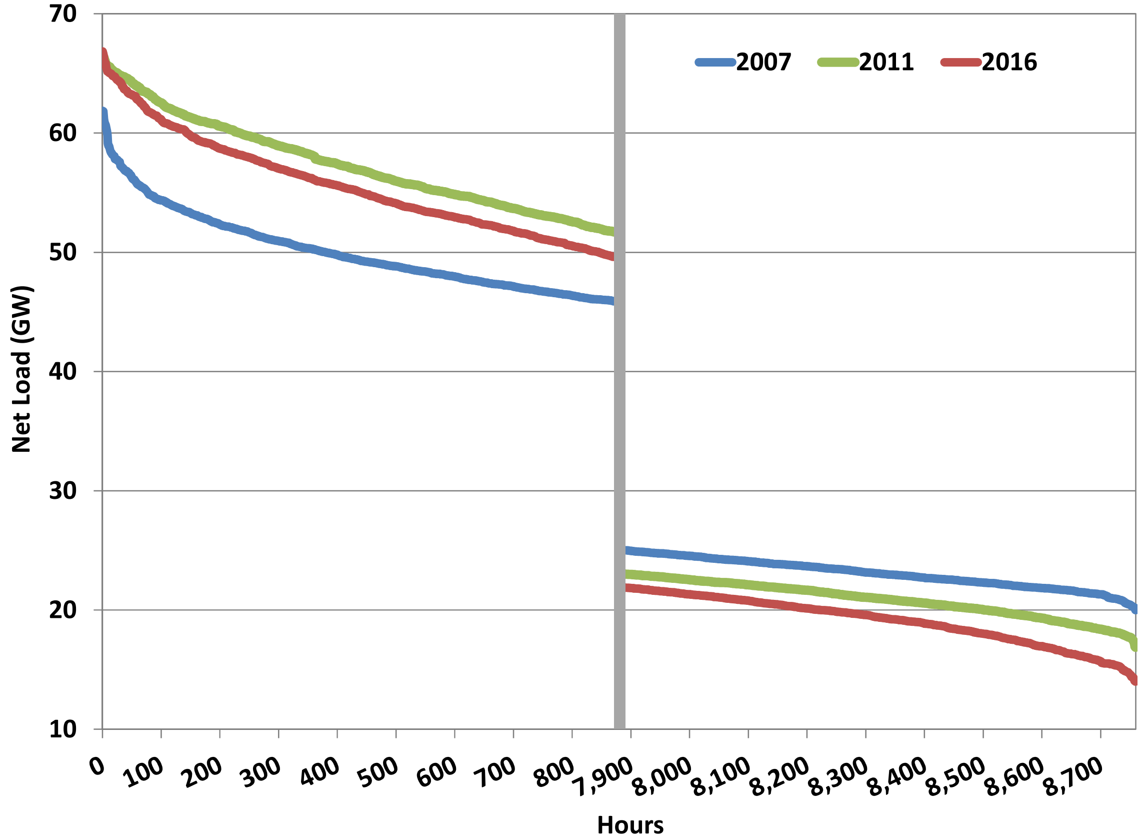


Figure 69 shows net load in the highest and lowest hours. Even with the increased development activity in the coastal area of the South zone, 73 percent of the wind resources in the ERCOT region are located in West Texas. The wind profiles in this area are such that most of the wind production occurs during off-peak hours or other times of low system demand. This profile results in only modest reductions of the net load relative to the actual load during the highest demand hours, but much larger reductions in the net load in the other hours of the year. Wind generation erodes the total load available to be served by base load coal units, while doing very little to reduce the amount of capacity necessary to reliably serve peak load.

In the hours with the highest net load (left side of the figure above), the difference between peak net load and the 95th percentile of net load has averaged 12.3 GW the past three years. This means that 12.3 GW of non-wind capacity is needed to serve load less than 440 hours per year.

Figure 69: Top and Bottom Ten Percent of Net Load

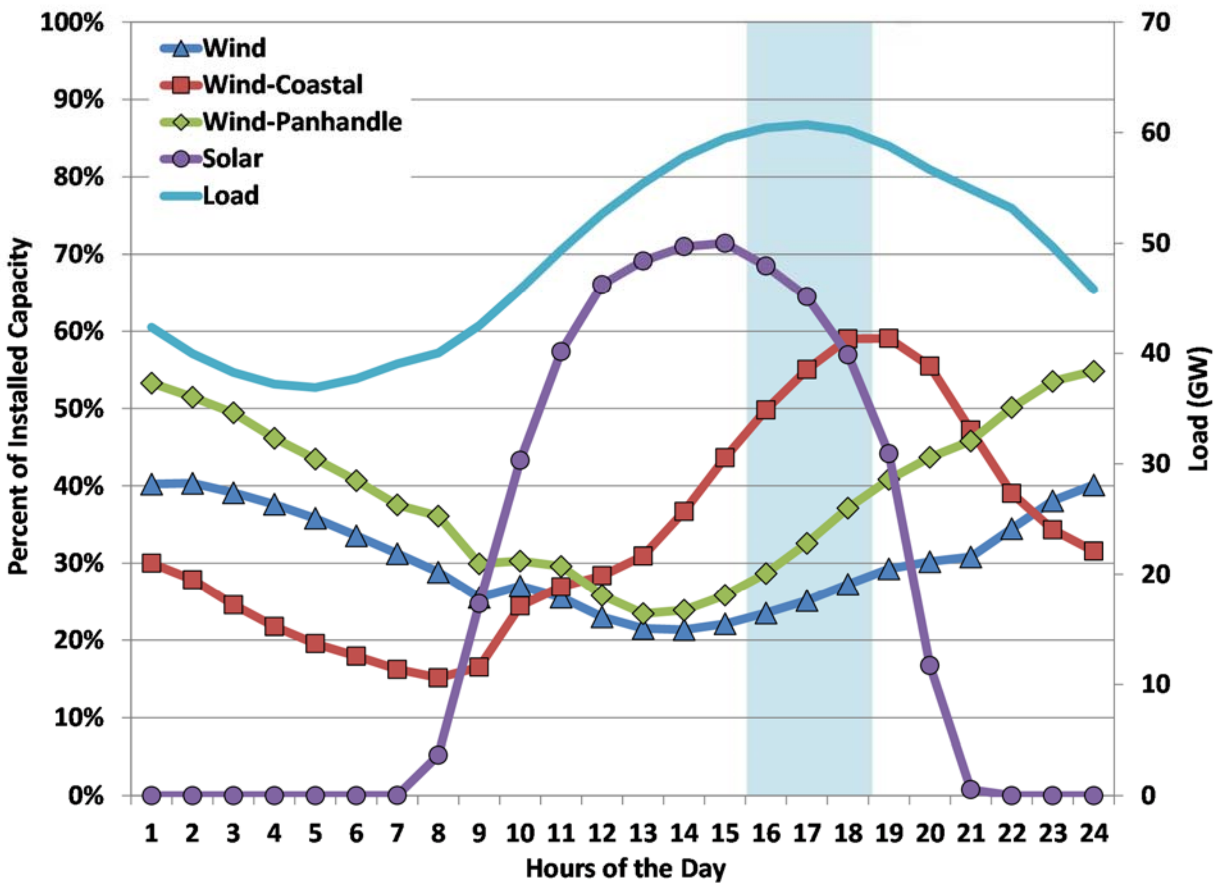


In the hours with the lowest net load (right side of the figure), the minimum net load has dropped from approximately 20 GW in 2007 to below 15.4 GW in 2016, even with the sizable growth in annual load that has occurred. This continues to put operational pressure on the 24 GW of nuclear and coal generation 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 satisfy ERCOT's reliability requirements, the non-wind fleet can expect to operate for fewer hours as wind penetration increases. This outlook further reinforces the importance of efficient energy pricing during peak demand conditions and other times of system stress, particularly in the context of the ERCOT energy-only market design.

The growing numbers of solar generation facilities in ERCOT have an expected generation profile highly correlated with peak summer loads. Figure 70 compares average summertime (June through August) hourly loads with observed output from solar and wind resources. Generation output is expressed as a ratio of actual output divided by installed capacity.

Figure 70: Summer Renewable Production



This figure shows that while the total installed capacity of solar generation is much smaller than that of wind generation, its production as a percentage of installed capacity is the highest in the early afternoon, around 70 percent, and producing more than 60 percent of its installed capacity during peak load hours.

The contrast between coastal wind and all other wind is also clearly displayed in Figure 70. Coastal wind produced over 50 percent of its installed capacity during summer peak hours. Output from Panhandle wind exceeded 30 percent, while output from all other wind (primarily West zone) was less than 30 percent during summer peak hours.

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. A second way that loads may participate is through ERCOT-dispatched reliability programs, including Emergency Response Service and legislatively-mandated demand response programs administered by transmission providers. Additionally, loads may self-dispatch by adjusting consumption in response to energy prices or by reducing consumption during specific hours to lower transmission charges.

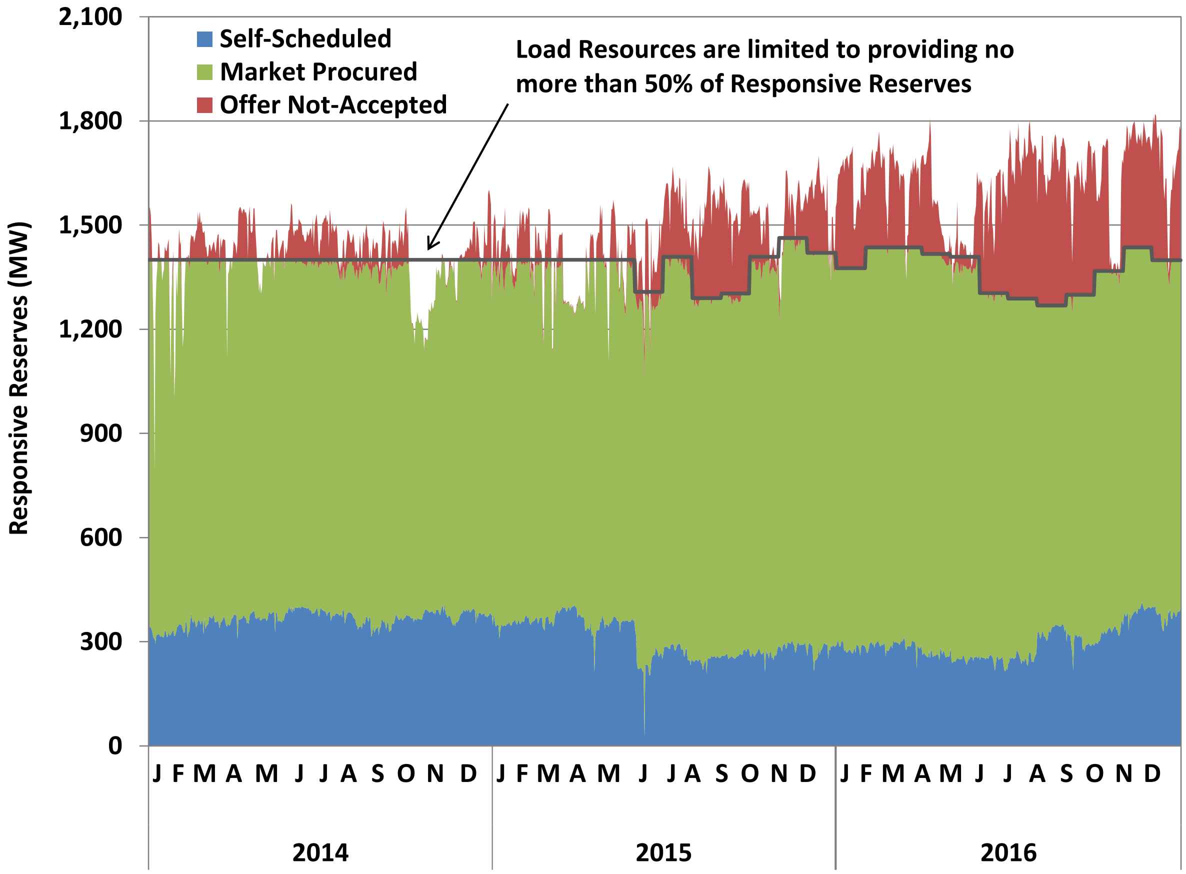
Reserve Markets

ERCOT allows qualified load resources to offer responsive reserves into the day-ahead ancillary services markets. Those providing responsive reserves have high set under-frequency relay equipment. This equipment enables the load to be automatically tripped when the frequency falls below 59.7 Hz, which will typically occur only a few times each year. As of December 2016, approximately 3,616 MW of qualified Load Resources were providing RRS, an increase of approximately 200 MW during 2016.

On June 1, 2015, ERCOT began procuring a variable amount of RRS based on season and time of day. The total amount of RRS varied between 2,300 to 3,000 MW. In 2016, the first full year with variable RRS procurement, the quantity of megawatts offered but not accepted by load resources increased. During 2016, there were no system-wide manual deployments of load resources providing RRS. There was, however, one automatic deployment of 927 MW of frequency responsive load on May 1, 2016.

Figure 71 below shows the average amount of responsive reserves provided from load resources on a daily basis for the past three years.

Figure 71: Daily Average of Responsive Reserves Provided by Load Resources



In 2016, load resources were limited to providing a maximum of 50 percent of responsive reserves. The quantity of offers submitted by load resources exceeds the 50 percent limit most of the time. The exception is when real-time prices are expected to be high. Since load resources provide capacity by reducing consumption, they have to be consuming energy to be eligible to provide the service. During periods of expected high prices the price paid for the energy can exceed the value received from providing responsive reserves. Reduced offer quantities observed during the spring and fall months may reflect the lack of availability of load resources due to annual maintenance at some of the larger load resource facilities.

ERCOT Protocols permit load resources to provide non-spinning reserves and regulation services, but for a variety of reasons there has been minimal participation by load resources.

Reliability Programs

There are two main reliability programs in which demand can participate in ERCOT – Emergency Response Service (ERS) and transmission provider load management programs. The ERS program is defined by a PUCT Rule enacted in March 2012 setting a program budget of

$50 million.[25] The program was modified from a pay as bid auction to a clearing price auction in 2014, providing a clearer incentive to load to submit offers based on the costs to curtail, including opportunity cost. In 2016, the procurement for ERS shifted from four time periods per contract term to six time periods per contract term. The additional time periods were created to separate the higher risk times of early morning and early evening from the overnight and weekend hours. The time and capacity-weighted average price paid for ERS over the contract periods from February 2016 through January 2017 was $6.86 per MWh, significantly higher than the average price of $3.91 per MWh paid for non-spinning reserves in 2016. ERS was not deployed in 2016.

Beyond ERS there are slightly less than 200 MW of load participating in load management programs administered by transmission providers.[26] Energy efficiency and peak load reduction programs are required under state law and PUCT rule and most commonly take the form of load management, where participants allow electricity to selected appliances (typically air conditioners) to be curtailed. These programs administered by transmission providers may be deployed by ERCOT during a Level 2 Energy Emergency Alert (EEA).

Self-dispatch

In addition to active participation in the ERCOT market and ERCOT-dispatched reliability programs; loads in ERCOT can observe system conditions and reduce consumption accordingly. This response comes in two main forms. The first is by participating in programs administered by competitive retailers and/or third parties to provide shared benefits of load reduction with end-use customers. The second is through actions taken to avoid the allocation of transmission costs. Of these two methods, the more significant impacts are related to actions taken to avoid the allocation of transmission costs.

For decades, transmission costs have been allocated on the basis of load contribution to the highest 15-minute system demand during each of the four months from June through September. This allocation mechanism is routinely referred to as four coincident peak, or 4CP. By reducing demand during peak periods, load entities seek to reduce their share of transmission charges. Over the last three years, transmission costs have risen by more than 60 percent, thus significantly increasing an already substantial incentive to reduce load during probable peak

---

25 *See* 16 TEX. ADMIN. CODE § 25.507.

26 *See* ERCOT 2016 Annual Report of Demand Response in the ERCOT Region (Mar. 2017) at 6, available at http://www.ercot.com/services/programs/load.

intervals in the summer.[27] ERCOT estimates that 835-1,491 MW of load were actively pursuing reduction during the 4CP intervals in 2016, an increase from the estimated response in 2015.[28]

Load curtailment to avoid transmission charges may be resulting in price distortion during peak demand periods since the response is targeting peak demand rather than responding to wholesale prices. This was readily apparent in 2016 as there were significant load curtailments corresponding to peak load days in June, July and September when real-time prices on those days were in the range of $25 to $40 per MWh.

Two recent changes in the ERCOT market have made advances in appropriately pricing actions taken by load in the real-time energy market. First, the initial phase of "Loads in SCED" was implemented in 2014, allowing controllable loads that can respond to 5-minute dispatch instructions to specify the price at which they no longer wish to consume. Although an important first step, there are currently no loads qualified to participate in SCED. Second, the reliability adder, discussed in more detail in Section I: Review of Real-Time Market Outcomes, performs a second pricing run of SCED to account for the amount of load deployed, including ERS.

27 Transmission Cost of Service (TCOS) in 2013 was $2 billion and for 2016 it was $3.2 billion. See PUCT Docket No. 40946, Commission Staff's Application to Set 2013 Wholesale Transmission Service Charges for the Electric Reliability Council of Texas, Final Order (Mar. 28, 2013) and PUCT Docket No. 45382, Commission Staff's Application to Set 2016 Wholesale Transmission Service Charges for the Electric Reliability Council of Texas, Final Order (Mar. 25, 2016).

28 *See* ERCOT, *2016 Annual Report of Demand Response in the ERCOT Region* (Mar. 2017) at 8, available at http://www.ercot.com/services/programs/load.

V. Reliability 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 the real-time market and inefficiently high energy prices; while over-commitment can result in excessive start-up costs, uplift charges, and inefficiently low energy prices.

The ERCOT market does not include a mandatory centralized unit commitment process. The decision to start-up or shut-down a generator is made by the market participant. ERCOT's day-ahead market outcomes help to inform these decisions, but ERCOT's day-ahead market is only financially binding. That is, when a generator's offer to sell is selected (cleared) in the day-ahead market there is no corresponding requirement to actually start that unit. The generator will be financially responsible for providing the amount of capacity and energy cleared in the day-ahead market whether or not the unit operates. This decentralized commitment depends on clear price signals to ensure an efficient combination of units are online and available for dispatch. ERCOT, in its role as reliability coordinator, has the responsibility to commit units it deems necessary to ensure the reliable operation of the grid. There can be gaps between what individual resources, in aggregate, view as economic commitment and what ERCOT views as necessary to ensure the reliability of the region. In the event of these gaps, ERCOT uses its discretion to commit additional units to ensure reliability.

This section describes the evolution of rules and procedures regarding reliability unit commitments (RUC), the outcomes of RUC commitments, and the price mitigation that occurs during RUC and local congestion. The section concludes with a discussion of the reliability must run procurement by ERCOT in 2016.

A. History of RUC-Related Protocol Changes

The RUC process has undergone several modifications since the nodal market began. The following changes were implemented in an effort to improve the commitment process and market outcomes associated with RUC. In March 2012, an offer floor was put in place for energy above the Low-Sustained Limit (LSL) for units committed through RUC.[29] Initially, the RUC offer floor was set at the system-wide offer cap. The RUC offer floor was subsequently adjusted to $1,000 per MWh[30] and then to the current offer floor of $1,500 per MWh.[31]

[29] NPRR435, Requirements for Energy Offer Curves in the Real Time SCED for Generation Resources Committed in RUC, implemented on March 1, 2012.

[30] NPRR568, Real-Time Reserve Price Adder Based on Operating Reserve Demand Curve, implemented on June 1, 2014.

[31] NPRR626, Reliability Deployment Price Adder, partially-implemented to update the RUC offer floor on October 1, 2014.

Resources committed through the RUC process receive a make-whole payment and forfeit market revenues through a "clawback" provision. Beginning on January 7, 2014, resources committed through the RUC process could forfeit the make-whole payments and waive the clawback charges, effectively self-committing and accepting the market risks associated with that decision.[32] This buyback or "opt-out" mechanism for RUC requires a resource to update its Current Operating Plan (COP) before the close of the adjustment period for the first hour of a RUC commitment.[33]

On June 25, 2015, ERCOT automated the RUC offer floor of $1,500 per MWh and implemented the Real-Time On-Line Reliability Deployment Adder (reliability adder).[34] Since that date, when a resource properly telemeters a status indicating it has been RUC committed, ERCOT systems automatically set the energy offer floor at $1,500 per MWh. The reliability adder, as discussed more in Section I: Review of Real-Time Market Outcomes, captures the impact of reliability deployments such as RUC on energy prices.

To provide even greater flexibility to resource owners, the RUC process will soon be modified to permit the ability to opt-out of RUC instructions given after the close of the adjustment period. NPRR744 modifies the opt-out trigger to real-time telemetry status rather than the COP submittal. This NPRR is expected to be implemented mid-year 2017.

During 2016, approximately 40 percent of RUC instructions were given after the close of the adjustment period, thereby foreclosing the opportunity for resources to self-commit the units and shoulder the market risk. The late RUC commitments, however, demonstrate ERCOT exercising restraint in waiting as long as possible for the market to respond before committing resources through RUC.

B. RUC Outcomes

ERCOT continually assesses the adequacy of market participants' resource commitment decisions using a reliability unit commitment (RUC) process that executes both on a day-ahead and hour-ahead basis. Additional resources may be determined to be needed for two reasons – to satisfy the total forecasted demand, or to make a specific generator available resolve a transmission constraint. The constraint may be either a thermal limit or a voltage concern.

---

32 NPRR416, Creation of the RUC Resource Buyback Provision (formerly "Removal of the RUC Clawback Charge for Resources Other than RMR Units"), as modified by NPRR575, Clarification of the RUC Resource Buy-Back Provision for Ancillary Services.

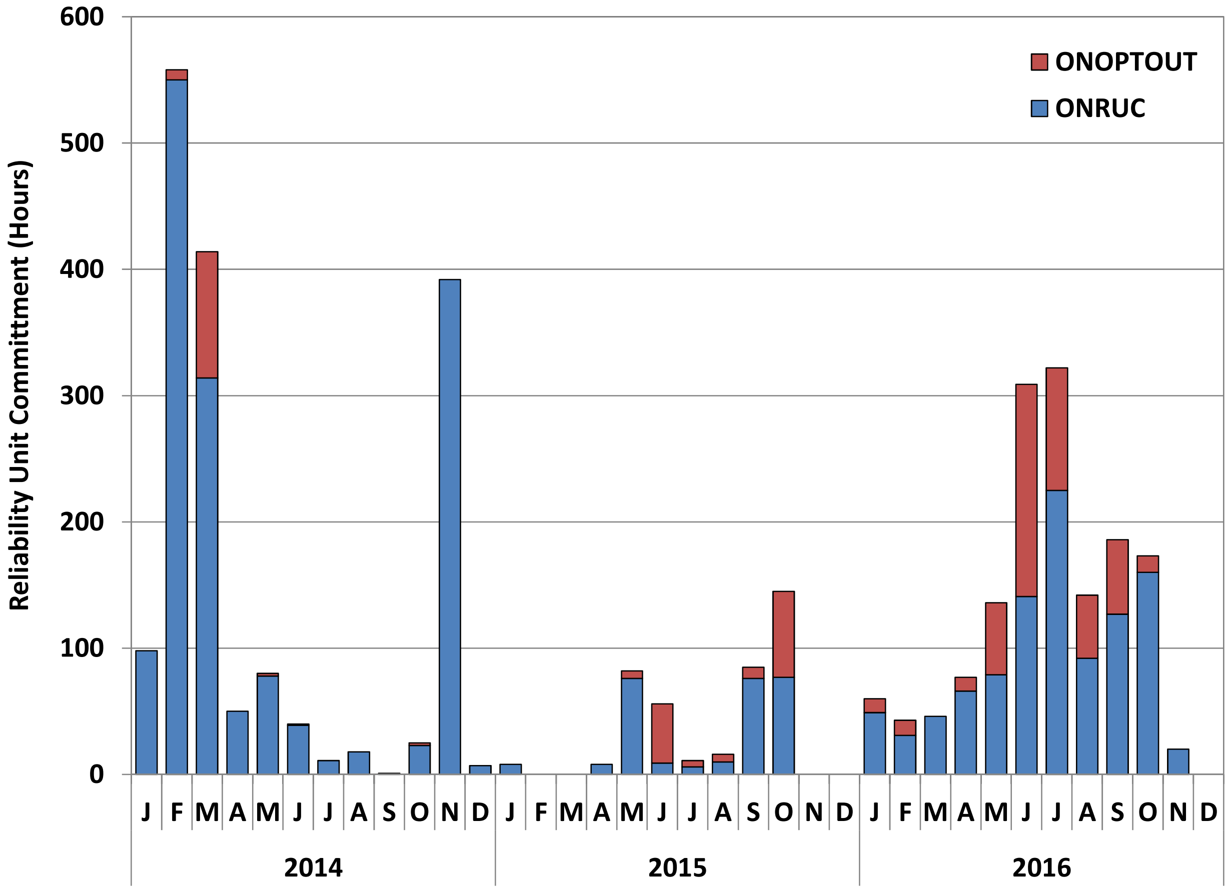
33 Note that the process for electing to opt-out of a RUC will be based on real-time telemetry when NPRR744, RUC Trigger for the Reliability Deployment Price Adder and Alignment with RUC Settlement, goes into effect in mid-2017.

34 *See* NPRR626, Reliability Deployment Price Adder (Formerly "ORDC Price Reversal Mitigation Enhancements").

A unit that receives a RUC instruction is guaranteed payment of its start-up and minimum energy costs (RUC make-whole payment). However, if the energy payments received by a unit operating under a RUC instruction exceed that unit's costs, payment to that unit is reduced (RUC clawback charge). Beginning in January 2014, a unit receiving a RUC instructions had the choice to "opt out," meaning it would forgo all RUC make-whole payments in return for not being subject to RUC clawback charges.

Figure 72 shows how frequently these reliability unit commitments have occurred over the past three years, measured in unit-hours.

Figure 72: Frequency of Reliability Unit Commitments



RUC commitments in 2016 were more frequent than in recent years. Although the total unit-hours were similar to the unit-hours in 2014, they were much more consistent in 2016. Almost twelve percent of hours in 2016 had at least one unit receiving a reliability unit commitment instruction. The reliability commitments in 2016 were primarily made to manage transmission constraints (98 percent of unit-hours), most of which were made to manage persistent congestion in the Houston area and in the Rio Grande Valley. The RUC activity in 2014 was concentrated during cold weather events in February and March and in response to transmission outages in

March and November. In 2015, RUC commitments were most frequent in the fall due to congestion in Dallas and the Rio Grande Valley.

During 2016, QSE telemetry of RUC status served as the trigger for calculating a reliability adder. There were 740 hours in which units were settled as RUC in 2016 and less than 500 cumulative hours of pricing intervals with non-zero reliability adders that occurred coincident with a settled RUC hour.

Table 9 provides the units most frequently called upon for RUC. Also provided are the hours of RUC instruction, the number of hours in which the unit opted-out, and the average low-sustained limit (LSL) for the unit. In 2016, units receiving RUC instructions successfully opted-out of 31.5 percent of unit-hours. The units highlighted in gray on Table 9 are units that were also on the most-frequent RUC commitment list in 2015.

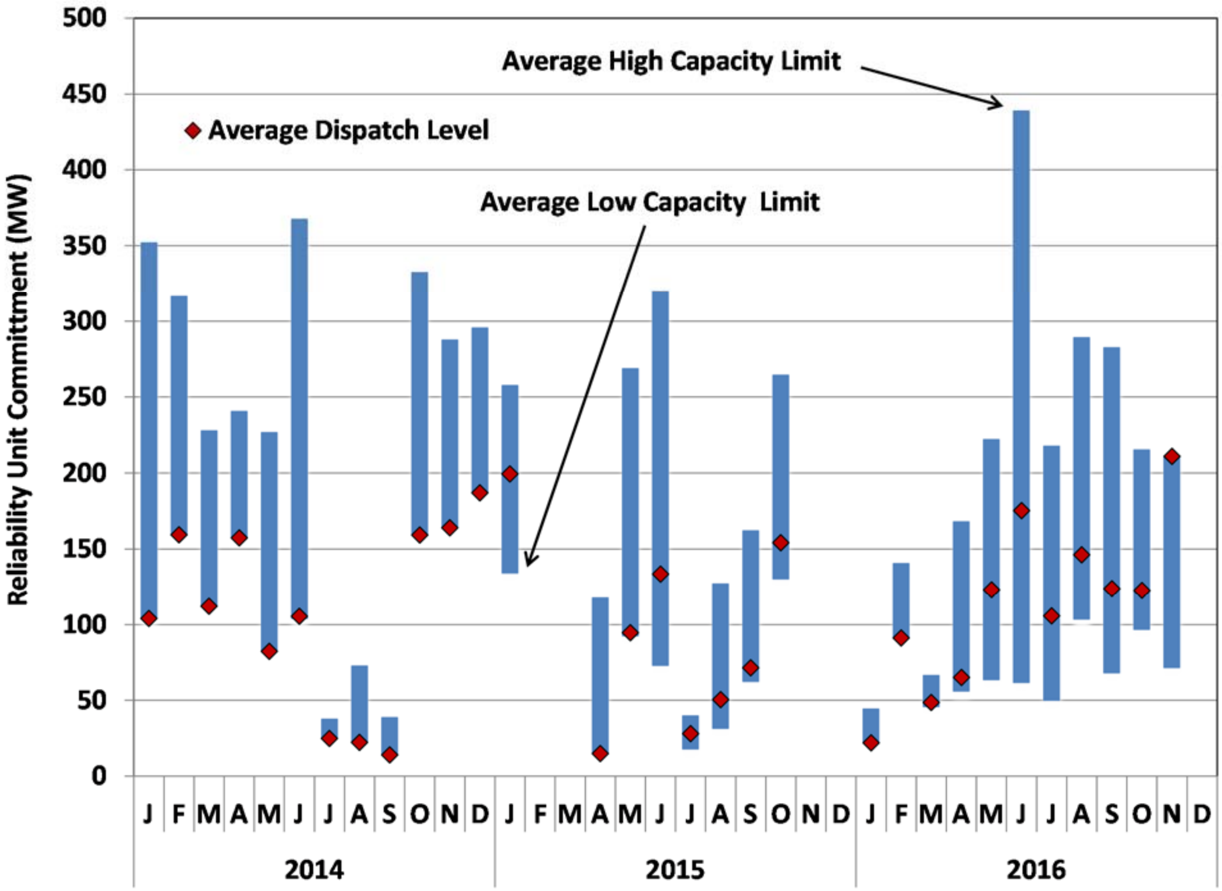
Table 9: Most Frequent Reliability Unit Commitments

Resource	Location	Unit RUC Hours	Unit OPTOUT Hours	Average LSL during RUC Hours
Silas Ray CC1	Valley	165	43	40
WA Parish G4	Houston	46	83	102
Silas Ray 10	Valley	83	28	21
WA Parish G2	Houston	53	34	29
Barney Davis G1	Corpus Christi	8	66	55
Spencer 5	Denton	54	13	17
Cedar Bayou G2	Houston	57	9	168
WA Parish G1	Houston	47	19	28
WA Parish G3	Houston	27	32	65
Cedar Bayou G1	Houston	-	51	-
Barney Davis CC1	Corpus Christi	43	3	238
North Edinburg CC1	Valley	32	8	222
Laredo G5	Laredo	35	-	35
Mountain Creek Unit 7	Dallas	33	-	15
Nueces Bay CC1	Corpus Christi	24	8	173

There were 1514 unit-hours with RUC instructions in 2016, compared with 411 unit-hours with RUC instructions during 2015. The majority of the RUC commitments were to resolve localized thermal transmission constraints (98 percent), and of those the majority were to units located in the Houston area (33 percent) and in the Rio Grande Valley (24 percent). There were 33 unit-hour commitments (2 percent) for system-wide capacity requirements. There were no commitments for voltage in 2016. Comparing 2016 to 2015 shows the same percent of RUC commitments for system-wide capacity at 2 percent; however, the total hours for system-wide capacity were significantly less in 2015 at only 8 unit-hours.

The next analysis compares the average dispatched output of the reliability-committed units, including those that opted-out, with the operational limits of the units. Figure 73 shows that the quantity of reliability unit commitment generation increased in 2016 compared to the prior two years. This figure shows that the average quantity dispatched for May through October 2016 exceeded 100 MW, and in November exceeded 200 MW.

Figure 73: Reliability Unit Commitment Capacity

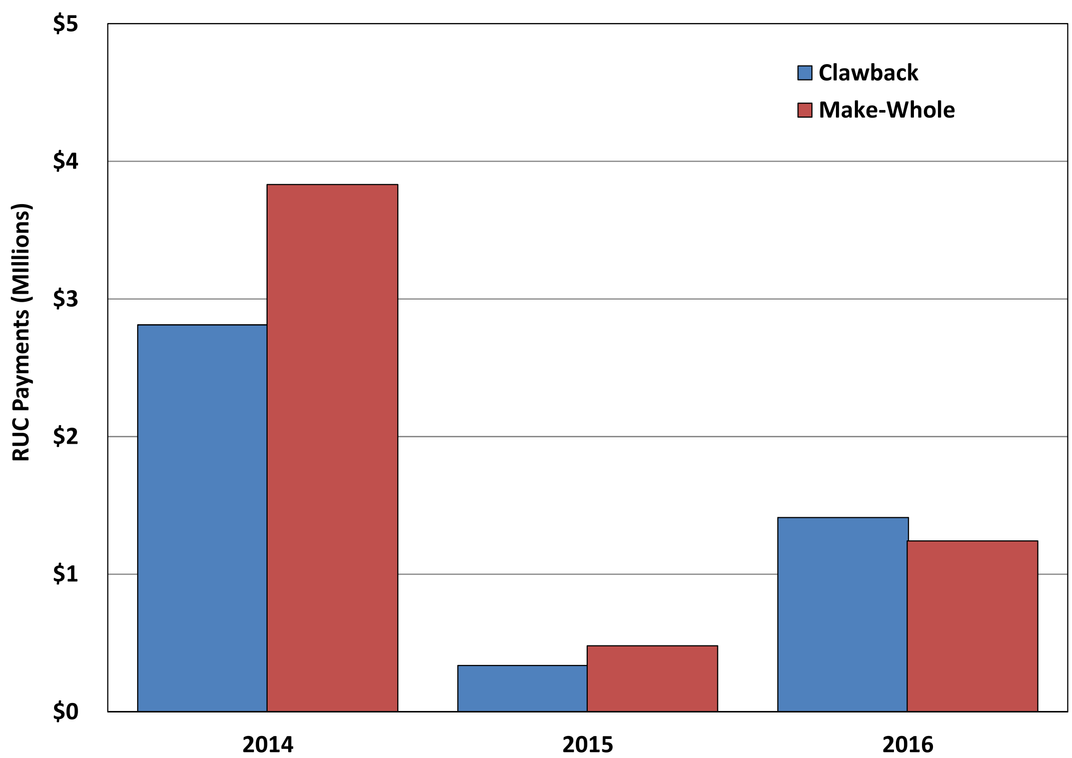


Units committed for RUC in 2016 showed a significant increase in the dispatch level compared to prior years. In twelve percent of intervals with RUC-committed resources, one or more resources were dispatched above their Low Dispatchable Limit (LDL), whereas in prior years, resources receiving RUC commitments were infrequently dispatched above LDL. Nonetheless, the higher dispatch levels in 2016 were rarely dispatched at the $1,500 per MWh offer floor because the commitments to address localized congestion were frequently mitigated.

When a unit is committed for RUC, the unit will receive a make-whole payment if the real-time revenues are less than the costs incurred to commit the unit. These costs can be based on generic values or unit-specific verifiable costs. Approximately 50 percent of resources in ERCOT have unit-specific verifiable costs. Of the 61 different resources that received a RUC instruction in 2016, 53 resources had approved unit-specific verifiable costs for start-up costs and minimum load costs. Those 53 resources represent 93 percent of total RUC-instructed megawatt-hours in 2016.

Figure 74 displays the total amount of make-whole payments and clawback charges attributable to reliability unit commitments annually for 2014-2016. Units that are RUC committed are guaranteed to be paid start-up and minimum energy costs. To the extent that the real-time energy market does not provide sufficient revenue to cover these costs, RUC-committed resources will receive a make-whole payment. There are two sources of funding for RUC make-whole payments. The first is from QSEs that do not provide enough capacity to meet their obligations. If there are remaining RUC make-whole funds required after contributions from any capacity short QSEs, any remaining RUC make-whole funding will be uplifted to all QSEs on a load-ratio share.

Figure 74: RUC Make-Whole and Clawback

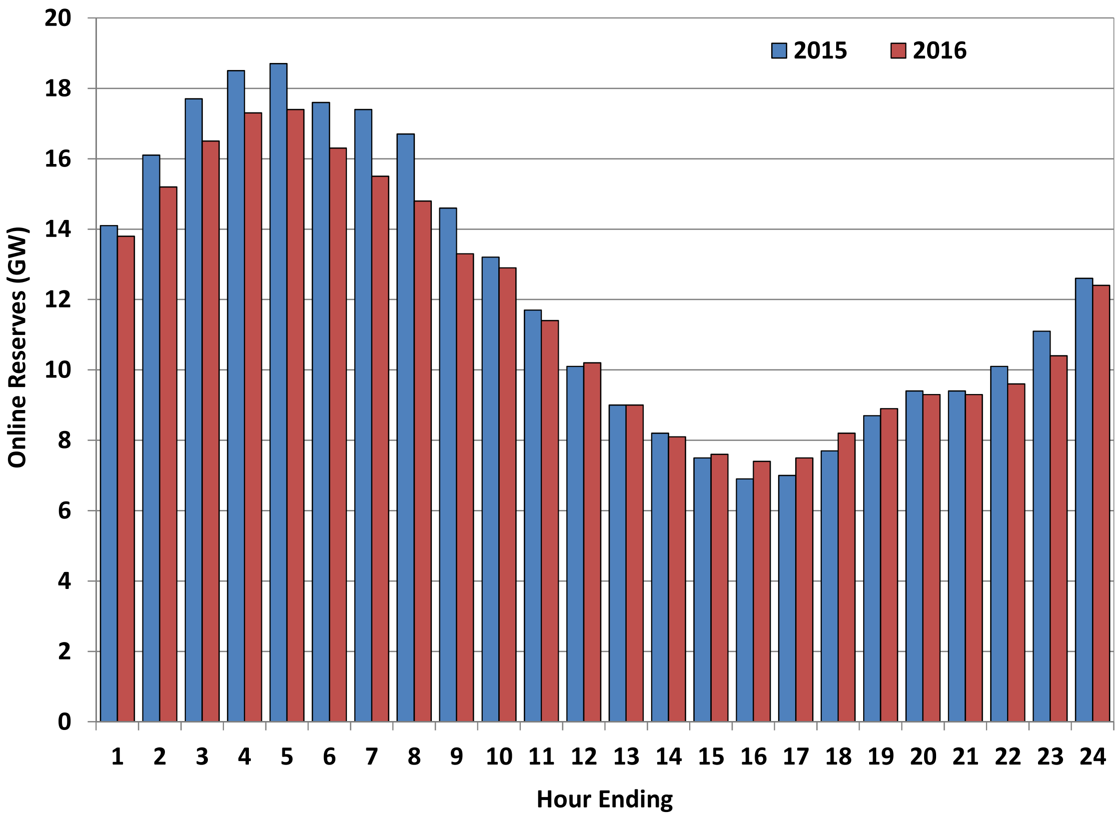


If real-time revenues received by a RUC committed resource exceed the operating costs incurred by the unit, then excess revenues are clawed-back and returned to QSEs representing load. During 2016, the make-whole and clawback amounts were nearly equal, with only slightly higher clawback charges. The source of funds for all RUC make-whole payments in 2016 were from QSEs that were capacity short. There was no general uplift to loads for RUC make-whole payments in 2016. The magnitude of both the clawback and make-whole amounts are very small in the scheme of the overall ERCOT real-time energy market.

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 the real-time market and inefficiently high energy prices; while over-commitment can result in excessive start-up costs, uplift charges, and inefficiently low energy prices.

The following figure compares the amount of on-line reserves, by hour, for the summer months of June through August in 2016 and 2015. The amount of on-line reserves is equal to the amount of capacity committed in excess of expected demand. Figure 75 displays available online reserves by operating hour and shows the expected pattern of declining reserves as system load increases during peak demand hours. In 2016, the average online reserves were greater than in 2015 for hours ending 12 through 19; in all other hours, the average online reserves were less than 2015.

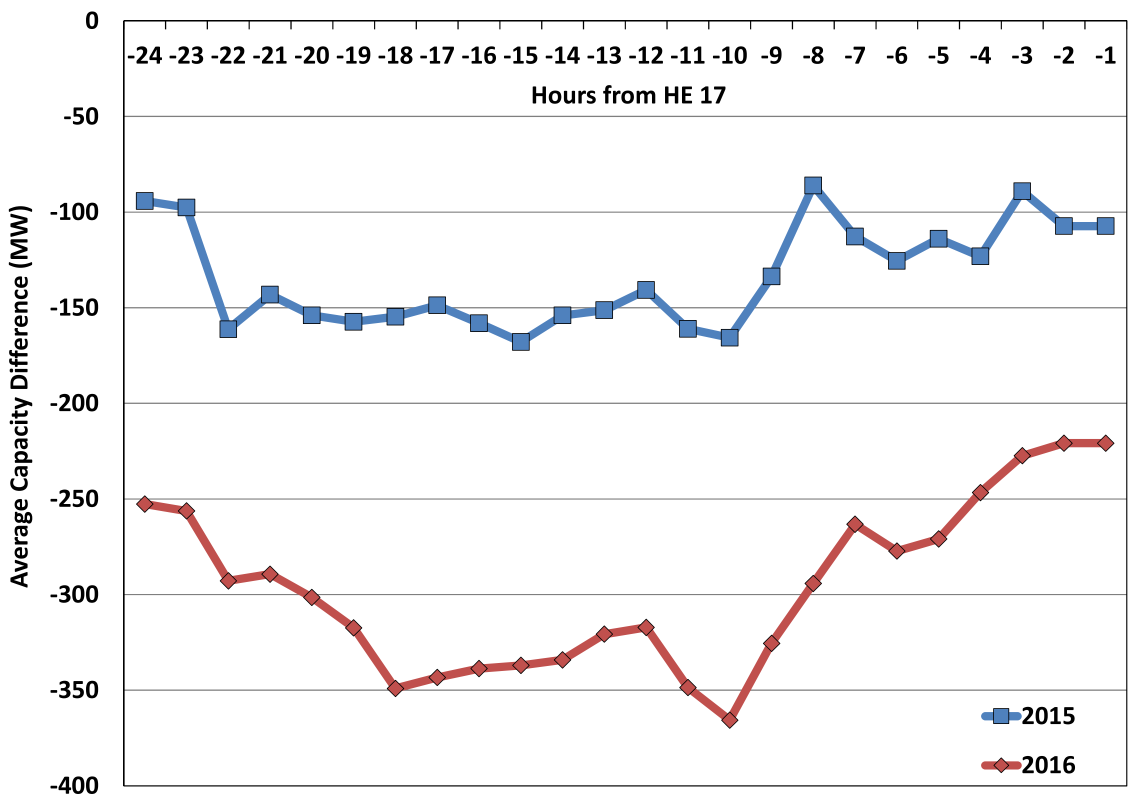
Figure 75: Average On-line Summer Reserves



The reduction in reserves during off-peak hours of the summer 2016 indicates that resource owners chose not to run units overnight. However, despite higher load levels during peak hours in 2016, average on-line summer reserves levels during peak hours were greater than in 2015. Lower energy prices are expected during periods of higher reserves.

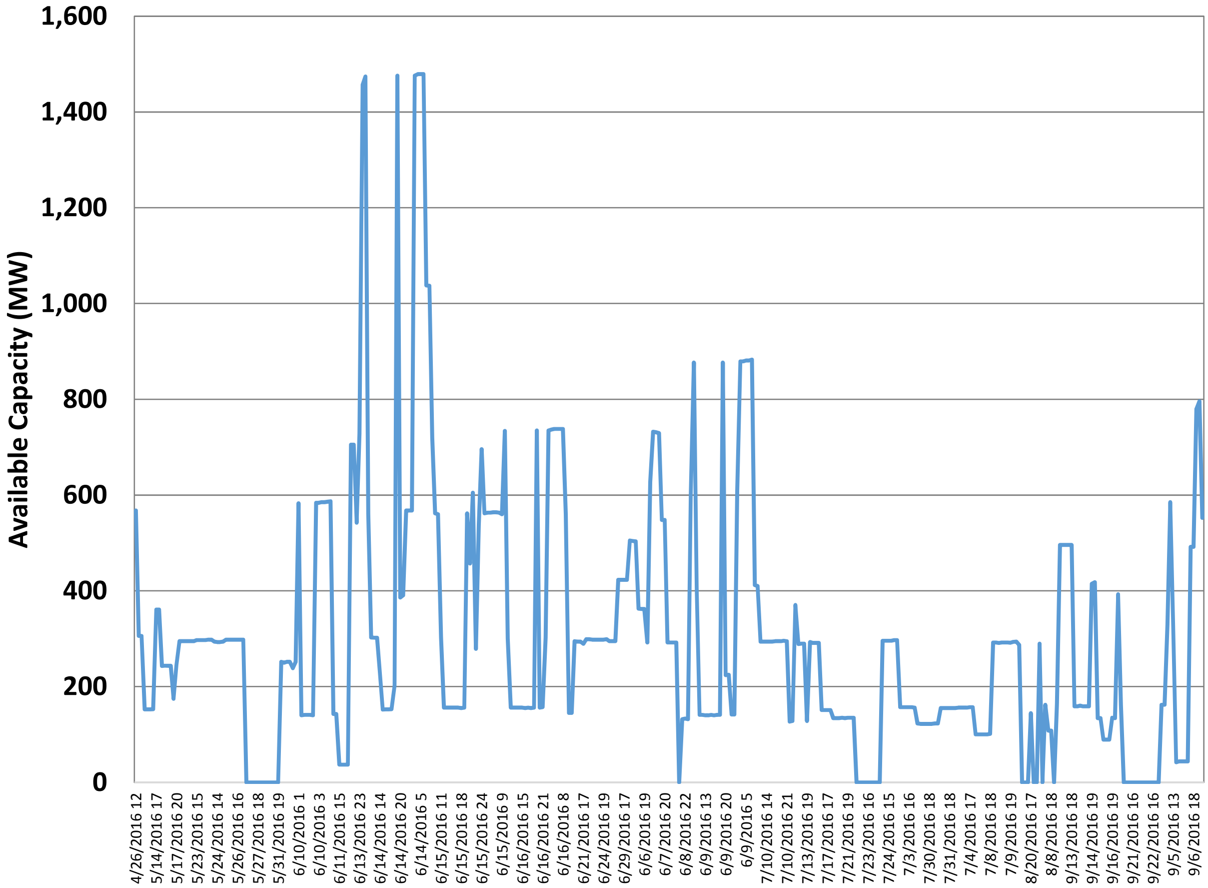
For a different look at self-commitment during the summer of 2016, Figure 76 shows the average difference between the actual online unit capacity in the peak hour and the amount of capacity committed for the peak hour by the online units for each of the 24 hours leading up to the close of the adjustment period. This data is for hour ending 17, averaged over the months of July and August for 2015 and 2016. As can be seen from this chart, the amount of capacity committed in advance of the operating hour was less in 2016 than 2015. In 2015 about 100 MW of capacity, on average, was committed in the last hour before real time. In 2016, the amount increased to over 200 MW, with even larger deficiencies seen in the last hours leading up to real time. From an ERCOT operator perspective, the self-commitment by market participants appears deficient and may be a potential contributor to the increased RUC activity in 2016.

Figure 76: Capacity Commitment Timing – July and August Hour 17



The last analysis of RUC activity in 2016 quantifies the amount of incremental combined-cycle capacity currently unavailable for RUC. Combined-cycle generators are comprised of multiple individual units, gas turbines and steam turbines that may be operated in various combinations. These different combinations, or configurations, have different operating characteristics and costs reflected in ERCOT systems. A common type of combined-cycle unit in ERCOT is comprised of two gas turbines and one steam turbine. When the resource operates in a configuration with only one gas turbine and the steam turbine, ERCOT's RUC software does not recognize the additional capacity available from the second gas turbine. This inability of the RUC software to evaluate changes to combined-cycle configurations may lead to situations where other, potentially more costly units receive RUC instructions to come online. A preliminary analysis was performed to quantify the amount of additional capacity available from combined-cycle units that had self-committed in a configuration less than the unit's largest capacity configuration. Figure 77 below displays the additional combined cycle megawatts located in Houston that could have been made available to RUC during the hours that at least one unit in Houston received a RUC instruction. These values exclude any incremental capacity from private use network resources.

Figure 77: Potential for Combined Cycle Capacity Available to RUC in Houston



The changes required to the RUC process to account for larger configurations of combined-cycle resources would be complex, including changes to the RUC engine and settlement systems. In addition, market participants would be required to provide significantly more detailed information on combined-cycle configurations. Given the relatively low overall cost to the market for RUC make-whole payments, implementing such a change may not be cost effective. However, the data indicates a sizable amount of incremental capacity is available.

C. Mitigation

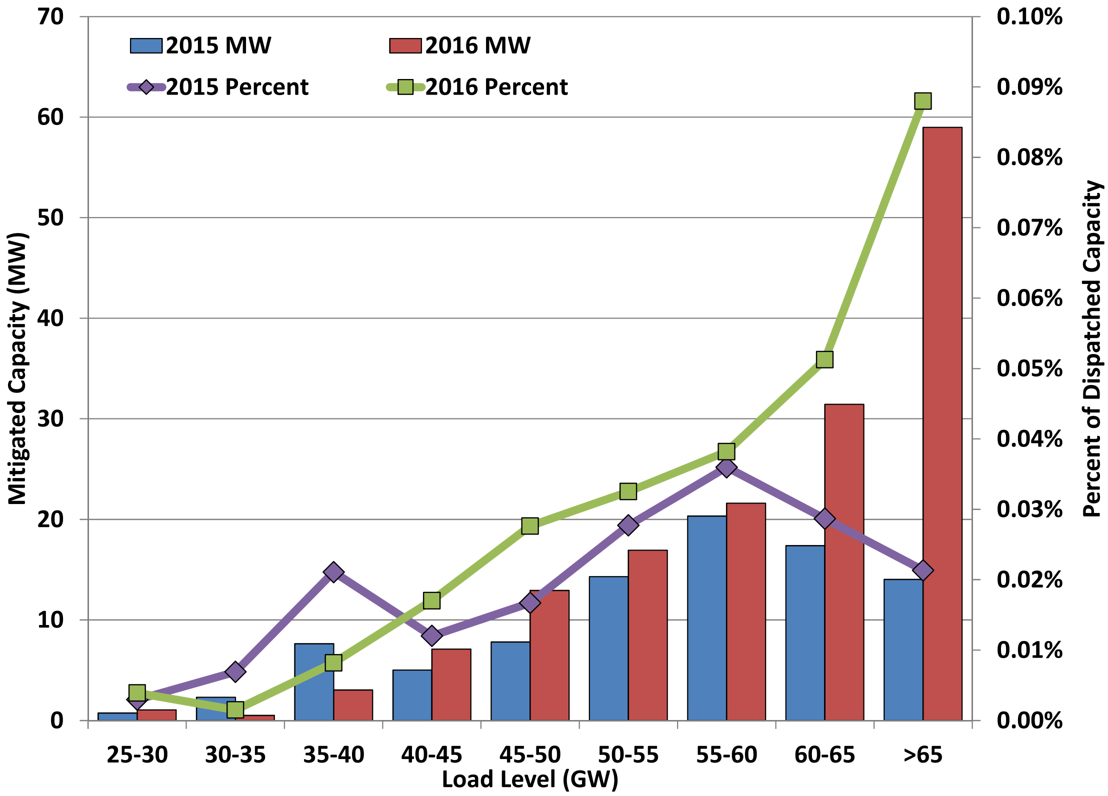
In situations where competitive forces are not sufficient, it can be necessary to mitigate prices to a level that approximates competitive outcomes. ERCOT's real-time market includes a mechanism to mitigate prices for resources that are required to resolve a transmission constraint. Mitigation applies whether the unit is self-committed or RUC committed. Units are typically RUC committed to resolve transmission constraints and as such they are typically required to resolve a transmission constraint, and therefore mitigated. As shown previously in Figure 73, it was more common for RUC-committed units to be dispatched above their low operating limits in 2016. This higher dispatch was due to the RUC-committed units being dispatched based on their mitigated price, not the RUC offer floor of $1,500 per MWh.

ERCOT's dispatch software includes an automatic, two-step price mitigation process. In the first step, the dispatch software calculates output levels (Base Points) and associated locational marginal prices using the participants' offer curves and considers only the 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.

This approach is intended to limit the ability of a generator to raise prices in the event of a transmission constraint that requires its output to resolve. In this subsection the quantity of mitigated capacity in 2016 is analyzed. Although executing all the time, the automatic price mitigation aspect of the two-step dispatch process only has the potential to have an effect when a non-competitive transmission constraint is active. With the introduction of an impact test in 2013 to determine whether units are relieving or contributing to a transmission constraint, only the relieving units are now subject to mitigation. This change has significantly reduced the amount of capacity subject to mitigation.

The analysis shown in Figure 78 computes the percent of capacity, on average, that is actually mitigated during each dispatch interval. The results are provided by load level.

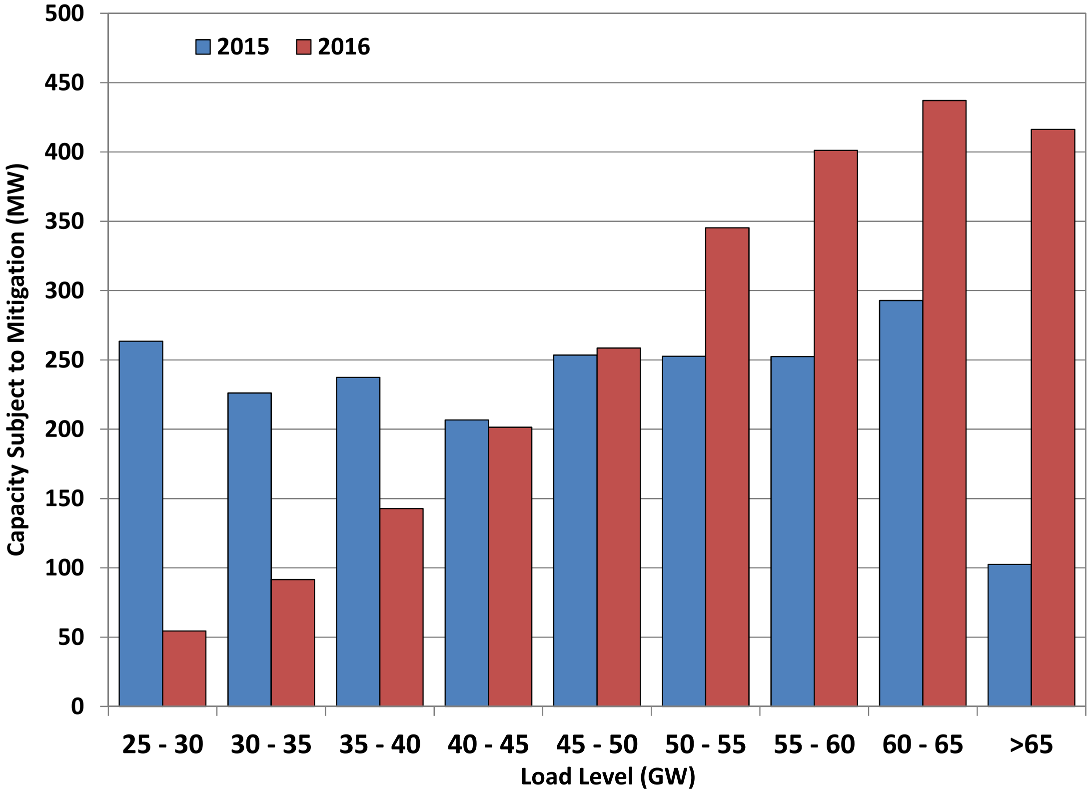
Figure 78: Mitigated Capacity by Load Level



The level of mitigation in 2016 was higher, particularly at higher load levels, than in 2015. The average amount of mitigated capacity was less than 20 MW for all load levels in 2015, but averaged almost 60 MW at loads greater than 65 GW in 2016. The greater frequency of congestion that occurred in 2016, as described in Section III: Transmission Congestion and Congestion Revenue Rights, supports the higher mitigation levels experienced in 2016.

In the previous figure, only the amount of capacity that could be dispatched within one interval was counted as mitigated. The next analysis computes the total capacity subject to mitigation, by comparing a generator's mitigated and unmitigated (as submitted) offer curves and determining the point at which they diverge. The difference between the total unit capacity and the capacity at the point the curves diverge is calculated for all units and aggregated by load level. The results are shown in Figure 79.

Figure 79: Capacity Subject to Mitigation



The amount of capacity subject to mitigation in 2016 was higher than 2015, especially at higher load levels. In 2015 and 2014, the largest amount of capacity subject to mitigation did not exceeded 300 MW. It is important to note that this measure includes all capacity above the point at which a unit's offers become mitigated, without regard for whether that capacity was actually required to serve load.

D. Reliability Must Run

Five units provided notice of the intent to suspend operations with a suspension date in 2016, amounting to approximately 1,100 MW of capacity retired or mothballed during the year. For the first time since 2011 ERCOT determined that there was a reliability need that warranted putting a unit under a reliability must run (RMR) contract. Greens Bayou 5 is a 371 MW natural gas steam unit built in 1973 and located in Houston. The RMR agreement was effective June 2, 2016 for a term of 25 months and a budgeted cost of $58.1 million, plus the opportunity for up to 10% more as an availability incentive. ERCOT initially determined that Greens Bayou 5 was needed for transmission system stability in the Houston region during the summers of 2016 and 2017 until the Houston Import Project transmission upgrade was completed. Following changes to the RMR study parameters[35] and the earlier than expected completion of new generation in Houston, the contract with the unit was cancelled effective May 29, 2017.

Prior to Greens Bayou 5, the last time units in the ERCOT market were under RMR agreements was in 2011 – a year of extreme heat and drought. That year, ERCOT required four units that had previously been allowed to enter mothball status to return to service under RMR contracts for the peak summer demand. The protocols were changed shortly thereafter to require that any energy from RMR units be offered at the system-wide offer cap.[36] Pricing out of market energy at the system-wide offer cap ensures that energy from RMR units is dispatched last.

The Greens Bayou 5 RMR presented a different pricing issue, since it was procured to resolve a transmission constraint. The Houston import constraint is frequently a non-competitive constraint, and hence, the price of energy from the RMR unit would be mitigated. Given the unit's significant helping impact on the constraint and the relatively low mitigated price, it was likely that if the unit was committed it would be dispatched before other similarly-priced or even lower-priced units in the Houston area. NPRR784 was proposed to address mitigated offer caps for RMR units, but market participants could not reach consensus on this approach and the protocol change request was not approved. Thus, any future RMR units could still be dispatched at a mitigated price that is not reflective of the reliability value of the resource.

The Greens Bayou 5 RMR drew significant scrutiny from market participants on the RMR process. In addition to NPRR784, there were other Protocol changes put in place as a result of the RMR contract. The ERCOT evaluation criteria for potential RMR units was adjusted to require that RMR units have a material impact on the expected transmission overload in order to be procured under an RMR contract.[37] A material impact was defined to mean more than a two percent helping shift factor and more than a five percent unloading factor on the transmission facility that is overloaded. This Protocol change facilitated ERCOT's re-evaluation of the RMR contract for Greens Bayou 5 and ultimately resulted in early termination of the contract. Other protocol changes clarified the ERCOT commitment process for RMR units,[38] updated the contracting and reimbursement process for RMR units,[39] and created a mechanism for clawback of capital contributions from an RMR unit if the unit returns to the market.[40]

---

35 *See* NPRR788, RMR Study Modifications.

36 *See* NPRR442, Energy Offer Curve Requirement for Generation Resources Providing Reliability Must-Run Service.

37 NPRR788, RMR Study Modifications.

38 NPRR793, Clarification to RMR RUC Commitment and Other RMR Cleanups.

39 *Id.*

40 NPRR795, Provisions for Refunds of Capital Contributions Made in Connection with an RMR Agreement.

VI. 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 system demands and reliability needs. This section begins with an evaluation of these economic signals by estimating the "net revenue" resources received from ERCOT real-time and ancillary services markets and providing comparisons to other markets. Next, the effectiveness of the Scarcity Pricing Mechanism is reviewed. The current estimate of planning reserve margins for ERCOT and other regions are presented, followed by a description of the factors necessary to ensure resource adequacy in an energy-only market design.

A. Net Revenue Analysis

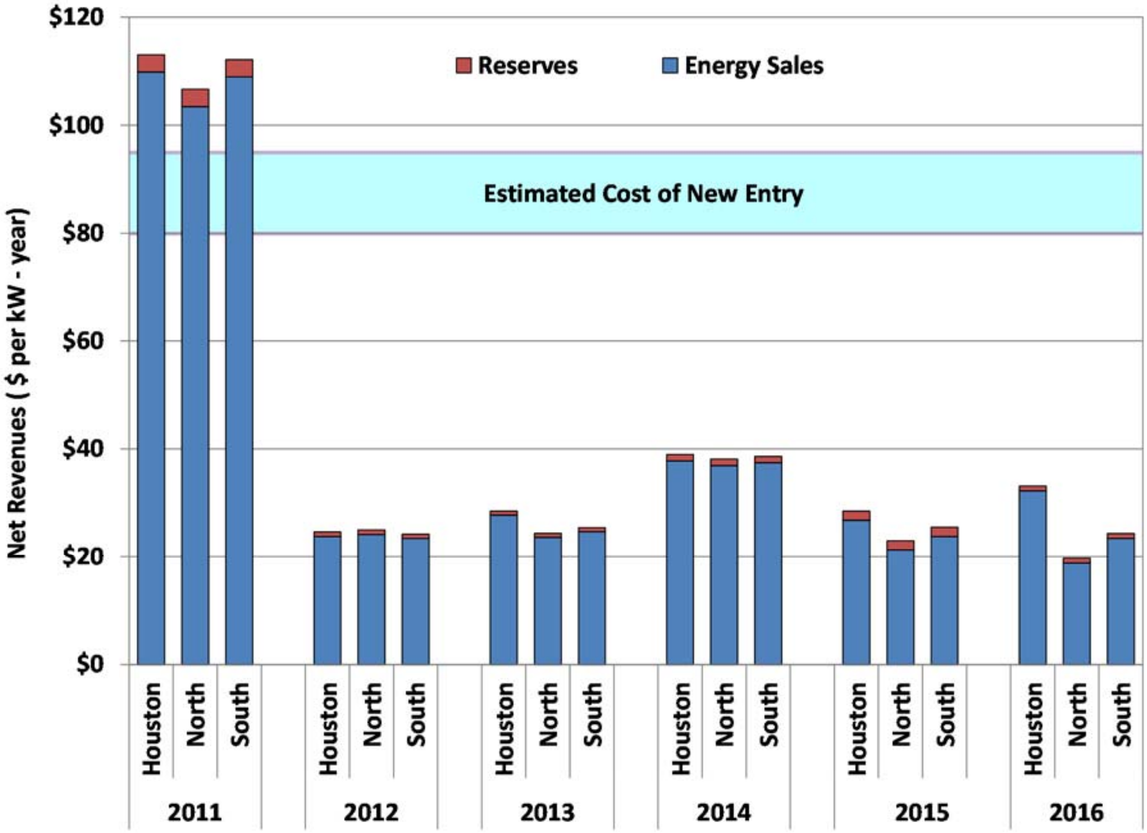
Net revenue is calculated by determining the total revenue that could have been earned by a generating unit less its variable production costs. Put another way, it is the revenue in excess of short-run operating costs that is available to recover a unit's fixed and capital costs, including a return on the investment. In ERCOT's energy-only market, the net revenues from the real-time energy and ancillary services markets alone provide the economic signals that inform suppliers' decisions to invest in new generation or retire existing generation. To the extent that revenues are available through the day-ahead market or other forward bilateral contract markets, these revenues are ultimately derived from the expected real-time energy and ancillary service prices. 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 bilateral energy prices over time and thus are appropriate to use for this evaluation. It is important to note that this net revenue calculation is a look back at the estimated contribution based on actual market outcomes. Suppliers will typically base investment decisions on expectations of future electricity prices. Although expectations of future prices should be informed by history, they will also factor in the likelihood of shortage pricing conditions that could be very different than what actually occurred.

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 facilitates comparisons between geographic zones, but will mask what could be very high values for a specific generator location. This analysis does not consider any payments for potential reliability unit commitment actions. The analysis necessitates reliance on simplifying assumptions that can lead to over-estimates of the profitability of operating in the wholesale market. Start-up costs and minimum running times are not accounted for in the net revenue analysis. Ramping restrictions, which can prevent generators from profiting during brief price spikes, are also excluded. But despite these limitations, the net revenue analysis provides a useful summary of signals for investment in the wholesale market.

For purposes of this analysis, the following assumptions were used for natural gas units: heat rates of 7 MMBtu per MWh for a combined cycle unit, 10.5 MMBtu per MWh for a combustion turbine, and $4 per MWh in variable operating and maintenance costs. A total outage rate (planned and forced) of 10 percent was assumed for each technology. 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 (combined cycle units only) in all other hours.

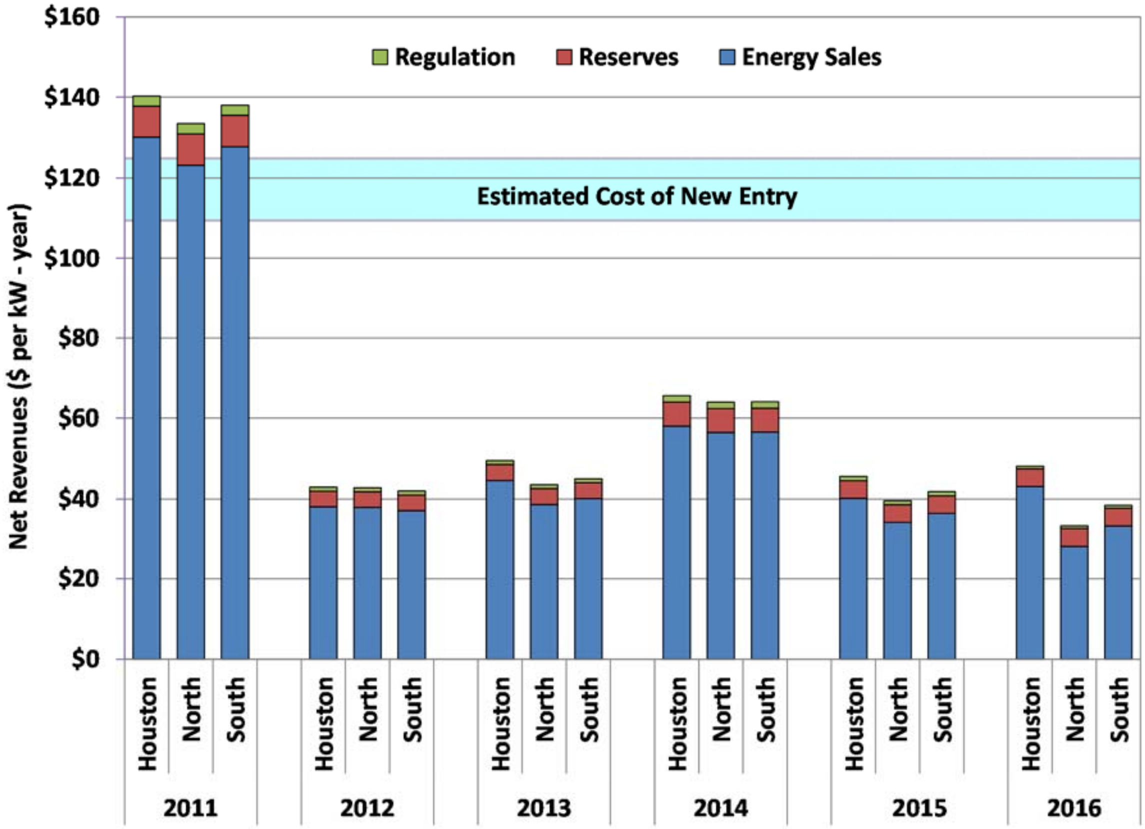
The next two figures provide an historical perspective of the net revenues available to support investment in a new natural gas combustion turbine (Figure 80) and combined cycle generation (Figure 81), selected to represent the marginal new supply that may enter when new resources are needed. The figure also shows the estimated "cost of new entry," which represents the revenues needed to break even on the investment.

Figure 80: Combustion Turbine Net Revenues



Based on 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 $95 per kW-year. The net revenue in 2016 for a new gas turbine was calculated to be approximately $20 to 33 per kW-year, depending on the zone, which are well below the estimated cost of new gas turbine generation.

Figure 81: Combined Cycle Net Revenues



For a new combined cycle gas unit, the estimate of net revenue requirement is approximately $110 to $125 per kW-year. The net revenue in 2016 for a new combined cycle unit was calculated to be approximately $33 to 48 per kW-year, depending on the zone. These values are well below the estimated cost of new combined cycle generation.

These results are consistent with the current surplus capacity, which contributed to infrequent shortages in 2015 and 2016. In an energy only market, shortages play a key role in delivering the net revenues an investor would need to recover its investment. Such shortages will tend to be clustered in years with unusually high load and/or poor generator availability. Hence, these results alone do not raise substantial concern regarding design or operation of ERCOTs ORDC mechanism for pricing shortages.

Table 10 displays the calculated output-weighted price by generation type.

Table 10: Settlement Point Price by Fuel Type

Generation Type	Output-Weighted Price
Combined-cycle > 90 MW	$24.59
Combined-cycle <= 90 MW	$27.74
Coal and lignite	$23.98
Diesel	$45.60
Gas steam non-reheat	$53.53
Gas steam reheat boiler	$44.60
Gas steam supercritical boiler	$35.12
Hydro	$22.04
Nuclear	$21.46
PhotoVoltaic Generation Resources	$31.95
Power Storage	$22.75
Renew	$28.21
Simple-cycle > 90 MW	$23.91
Simple-cycle <= 90 MW	$39.68
Wind	$16.18

Given the very low energy prices during 2016 in non-shortage hours, the economic viability of existing coal and nuclear units was evaluated. Non-shortage prices, which have been substantially affected by the prevailing natural gas prices, determine the vast majority of net revenues received by these base load units. As previously described, the load-weighted ERCOT-wide average energy price in 2016 was $24.62 per MWh. The generation-weighted average price for the four nuclear units in ERCOT - approximately 5 GW of capacity - was only $21.46 per MWh in 2016, down from $24.56 per MWh in 2015. According to the Nuclear Energy Institute (NEI), total operating costs for all nuclear units across the U.S. averaged $27.17 per MWh in 2016, a slight decrease from the reported costs for 2015.[41] Assuming that operating costs in ERCOT are similar to the U.S. average, it is likely that these units were not profitable in 2016 based on the fuel and operating and maintenance costs alone. To the extent nuclear units in ERCOT had any associated capital costs, it is likely those costs were not recovered. Compared to other regions with larger amounts of nuclear generation, the four nuclear units in ERCOT are relatively new and owned by four entities with sizable load obligations. Although not profitable on a stand-alone basis, the nuclear units have substantial

41 NEI Whitepaper, "Nuclear Costs in Context," April 2017, available at https://www.nei.org/www.nei.org/files/fe/fed92b11-8ea6-40df-bb0c-29018864a668.pdf.

option value for the owners because they ensure that their cost of serving their load will not rise substantially if natural gas prices increase. Nonetheless, the economic pressure on these units does potentially raise a resource adequacy issue that will need to be monitored.

The generation-weighted price of all coal and lignite units in ERCOT during 2016 was $23.98 per MWh. Although specific unit costs may vary, index prices for Powder River Basin coal delivered to ERCOT were approximately $2.50 per MMBtu in 2016, a decrease from approximately $2.60 per MMBtu in 2015. For the past two years delivered coal costs in ERCOT have been about $0.03 to $0.05 per MMBtu higher than natural gas prices at the Houston Ship Channel. Given that the coal units generally have higher heat rates and more expensive non-fuel operations and maintenance costs, it follows that they have been losing market share to natural gas. As with nuclear units, it appears that coal units were likely not profitable in ERCOT during 2016. With the bulk of the coal fleet in ERCOT being more than 30 years old, the retirement or suspended operation of some of these units could cause ERCOT's capacity margin to fall to unreliable levels more quickly than anticipated. While both nuclear and coal are feeling the pressure of an increased reliance on lower-priced natural gas units, coal units appear to be at greater risk of retirement than the nuclear units in ERCOT. This may be due to their relative age and inefficiency.

These results indicate that during 2016 the ERCOT markets would not have provided sufficient revenues to support profitable investment in any of the types of generation technology evaluated, which may seem inconsistent with the fact that new generation continues to be added in the ERCOT market. This can be explained by a number of factors.

First, resource investments are driven primarily by forward price expectations. Historical net revenue analyses do not provide a view of the future pricing expectations that will spur new investment. Suppliers will develop their own view of future expected revenue and given the level to which prices will rise under shortage conditions, small differences in expectations about the frequency of shortage pricing can greatly influence revenue expectations.

Second, this analysis does not account for bilateral contracts. The only revenues considered in the net revenue calculation are those that came directly from the ERCOT real-time energy and ancillary services markets in a specific year. Some developers may have bilateral contracts for unit output that would provide more revenue than the ERCOT market did in 2016. Given the level to which prices will rise under shortage conditions, buyers may enter bilateral contracts to hedge against high shortage pricing.

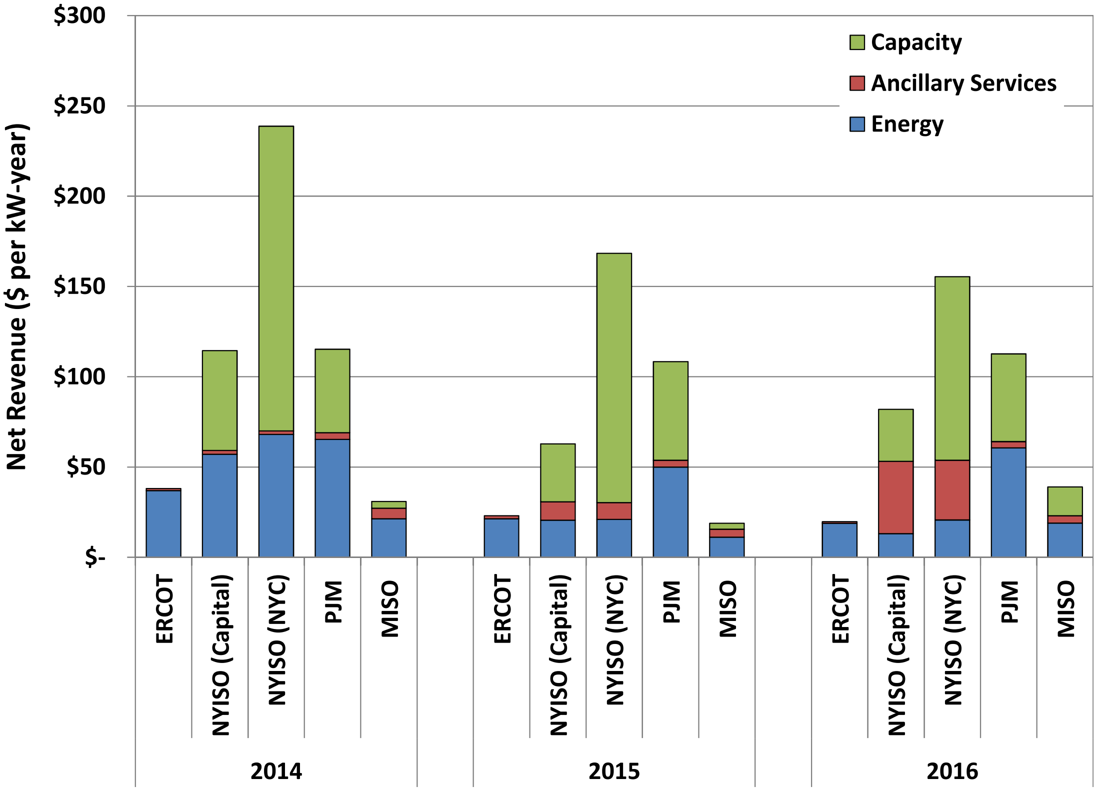
Third, net revenues in any one year may be higher or lower than an investor would require over the long term. In 2016, shortages were much less frequent than would be expected over the long term. Shortage revenues play a pivotal role in motivating investment in an energy-only market like ERCOT. Hence, in some years shortage pricing will be frequent and net revenues may substantially exceed the cost of entry, while in most others it will be less frequent and net revenue will be less than the cost of entry.

Finally, the costs of new entry used in this report are generic and reflective of the costs of a new unit on an undeveloped greenfield site. They have been reduced somewhat to reflect the lower costs of construction in Texas. However, companies may have opportunities to build generation at much lower cost than these estimates; either by having access to lower cost equipment, or by adding the new unit to an existing site, or some combination of both. Financing structures and costs can vary greatly between suppliers and may be improved to be lower than the generic financing costs assumed in the net revenue analysis.

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

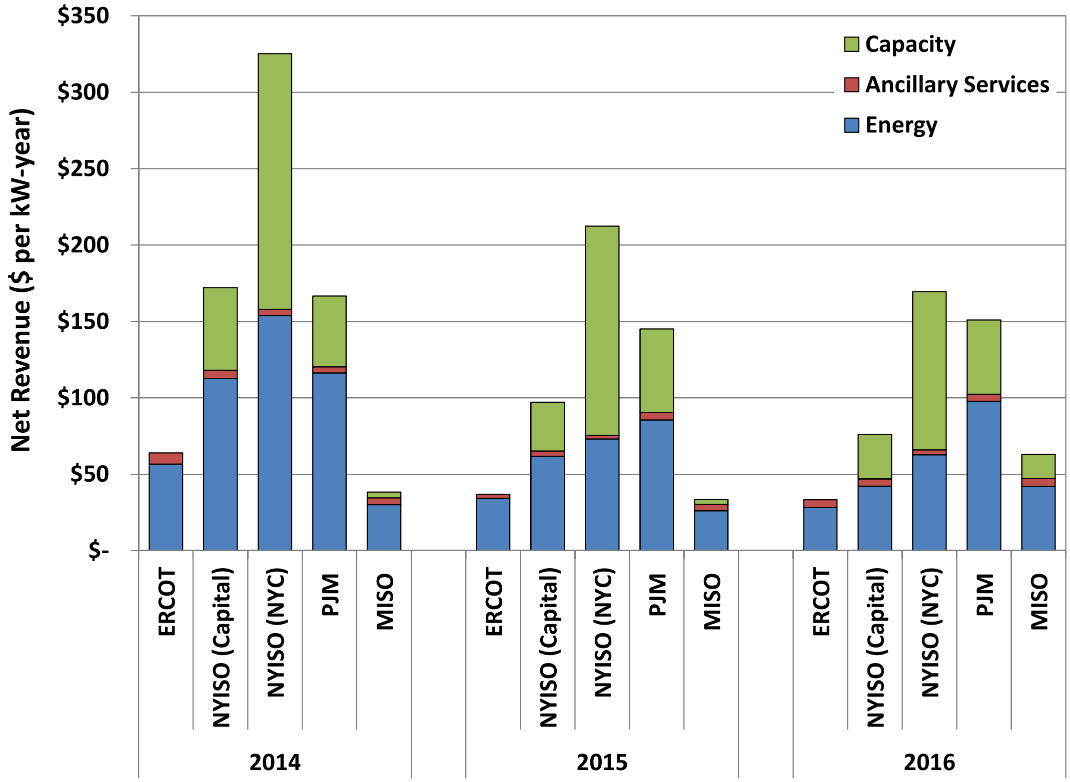
The next two figures compare estimates of net revenue for these two types of natural gas generators for the ERCOT North zone, PJM, two locations within the New York ISO, and the Midcontinent ISO. Figure 82 provides a comparison of net revenues for a combustion turbine and Figure 83 provides the same comparison for a combined cycle unit.

Figure 82: Combustion Turbine Net Revenue Comparison Between Markets



The figures include 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. Most of the locations shown are central locations, but there are load pockets within each market where net revenue and the cost of new entry may be higher. The NYC zone of the New York ISO is an example of much higher value in a load pocket. Thus, even if new investment is not generally profitable in a market, it may be economic in certain areas.

Figure 83: Combined Cycle Net Revenue Comparison Between Markets



Both figures indicate that across all markets, with the exception of New York ISO (Capital) for combustion turbine, net revenues decreased substantially in 2016 because of low natural gas prices across the country and sufficient installed reserves, typically a result of flat or no load growth. With the exception of MISO, capacity revenues provide a meaningful portion of the net revenues for new resources. In ERCOT, these revenues will be provided through its shortage pricing, which is evaluated in the next section.

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. In accordance with the IMM's charge to conduct an annual review,[42] this subsection assesses the Scarcity Pricing Mechanism (SPM) in 2016 under ERCOT's energy-only market structure.

Revisions to 16 TEX. ADMIN. CODE § 25.505 were adopted in 2012 that specified a series of increases to the ERCOT system-wide offer cap. The last step went into effect on June 1, 2015, increasing the system-wide offer cap to $9,000 per MWh. As shown in Figure 20 on page 20, there have been very brief periods when energy prices rose to the cap since the system-wide offer

[42] *See* 16 TEX. ADMIN. CODE § 25.505(g)(6)(D).

cap was increased to greater than $3,000 per MWh. There have been no instances of prices rising above $5,000 per MWh.

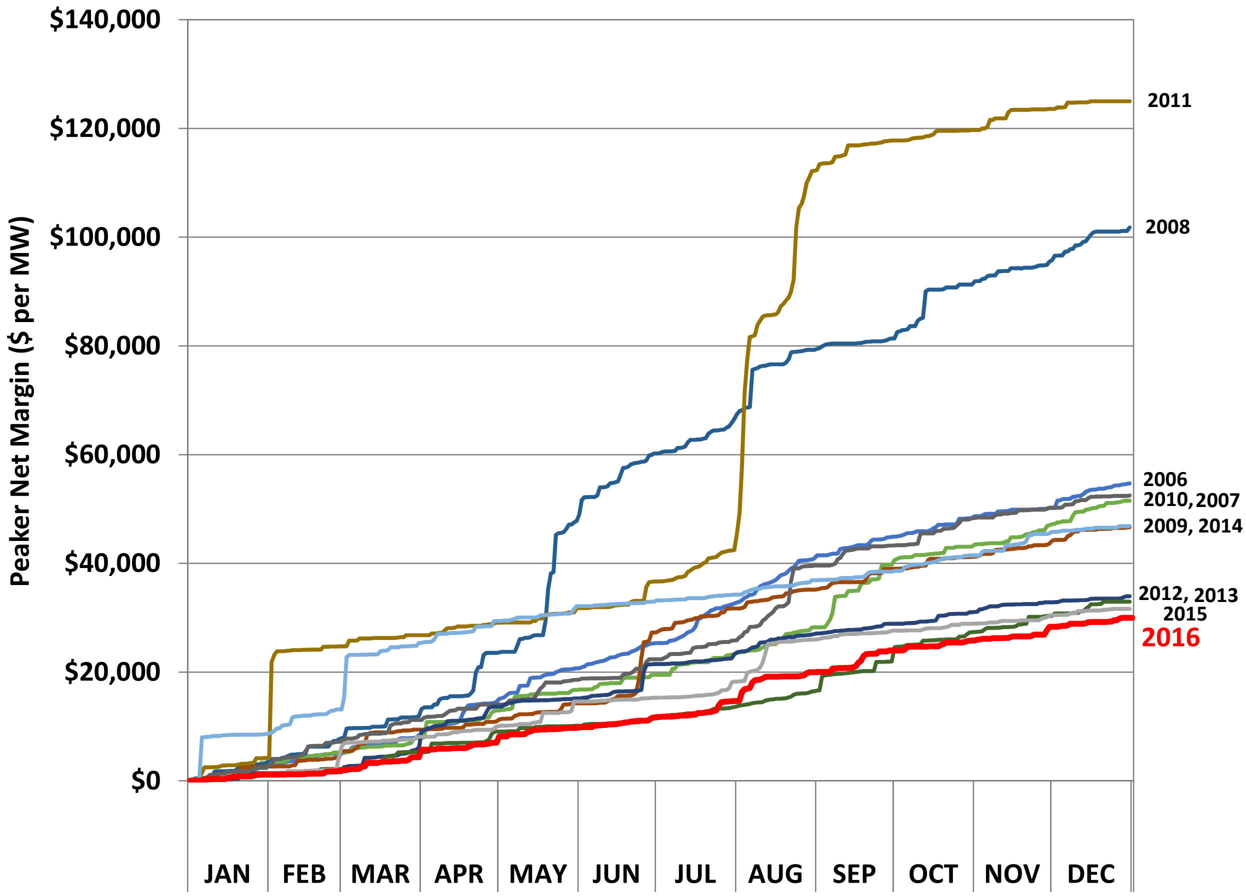
The SPM includes a provision termed the Peaker Net Margin (PNM) that is designed to provide a fail-safe pricing measure, which if exceeded would cause the system-wide offer cap to be reduced. If the PNM for a year reaches a cumulative total of $315,000 per MW, the system-wide offer cap is then reduced to the higher of $2,000 per MWh or 50 times the daily natural gas price index.[43] PNM also serves as a simplified measure of the annual net revenue of a hypothetical peaking unit.[44]

Figure 84 shows the cumulative PNM results for each year from 2006 through 2016 and shows that PNM in 2016 was the lowest it has been since it became effective in 2006. Considering the purpose for which the PNM was initially defined, that is to provide a "circuit breaker" trigger for lowering the system-wide offer cap, it has not approached levels that would dictate a needed reduction in the system wide offer cap.

43 The threshold established in the initial Rule was $300,000 per MW-year. For 2014 and each subsequent year, ERCOT shall set the PNM threshold at three times the cost of new entry of new generation plants. The current threshold is based on the analysis prepared by Brattle dated June 1, 2012, and will remain in place until there is a change identified in the cost of new entry of new generation plants.

44 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.

Figure 84: Peaker Net Margin



As with net revenues, the PNM is expected to be less than the cost of new entry in most years. Concerns with the SPM under the zonal market design were addressed in every State of the Market Report produced during that period.[45] The implementation of the nodal market design, which included a power balance penalty curve, created the opportunity for real-time energy prices to systematically reflect the value of reduced reliability imposed under shortage conditions, regardless of submitted offers.

In 2013, the PUCT took another step toward improve resource adequacy signals, by directing ERCOT to implement the Operating Reserve Demand Curve (ORDC). As discussed in Section I: Review of Real-Time Market Outcomes, ORDC is a shortage pricing mechanism that reflects the loss of load probability at varying levels of operating reserves multiplied by the value of lost load. In the short time it has been in effect ORDC has had a small impact on real-time prices.

---

45 The zonal market design was not the problem per se, rather its reliance on high-priced offers to set high prices during periods of shortage was of concern.

In October, 2015 the PUCT signaled its interest in reviewing ORDC "in order to examine how it has functioned and whether there is a need for minor adjustments to improve its efficiency."[46] Given the short time period with ORDC in effect, it is difficult to evaluate whether adjustments are warranted. As previously discussed, shortages are generally clustered in periods when weather-dependent load is unusually high and/or generation availability is poor; neither of which was the case in 2015 or 2016. The PUCT has taken comment from stakeholders, but to date the PUCT has not directed modification of the reserve adder component of ORDC.[47]

The fact that responsive and regulating reserves are forced to be maintained (held behind the High Ancillary Service Limit (HASL)) under the current market design will continue to be problematic, regardless of the ORDC parameters that are selected. Jointly optimizing all products would improve the utilization of ERCOT resources, ensure that shortage pricing only occurs when the system is actually short after fully utilizing its resources, and establish prices for each product that efficiently reflect its reliability value without the use of administrative caps and adders. Hence, the IMM continues to recommend that ERCOT make the investment necessary to achieve the full benefits of real-time co-optimization across all resources.

C. Planning Reserve Margin

The prior subsection discusses and evaluates the economic signals produced by the ERCOT markets to facilitate efficient decisions by suppliers to maintain an adequate base of resources. This subsection summarizes and discusses the current level of capacity in ERCOT, as well as the long-term need for capacity in ERCOT. The figure below shows ERCOT's current projection of planning reserve margins.

46 PUCT Docket No. 40000, *Commission Proceeding to Ensure Resource Adequacy in Texas*, Memorandum from Commissioner Kenneth W. Anderson, Jr. (Oct. 7, 2015).

47 See PUCT Docket No. 45572, Review of the Parameters of the Operating Reserve Demand Curve.

Figure 85: Projected Planning Reserve Margins

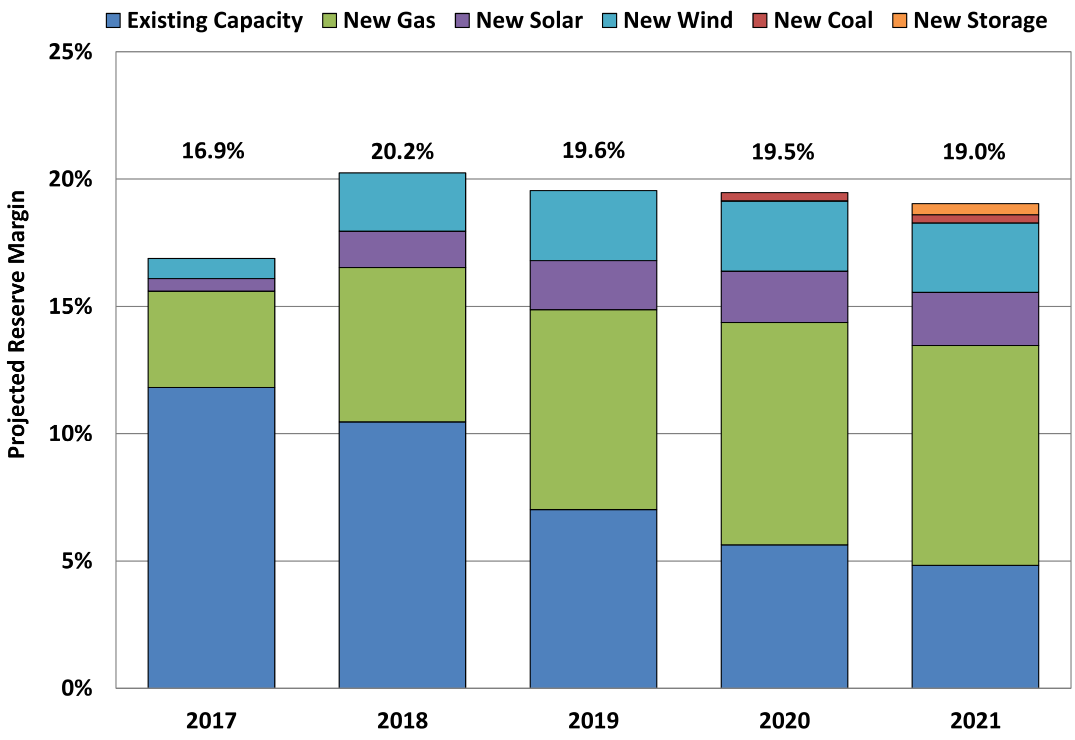


Figure 85 above indicates that the region will have a 16.9 percent reserve margin heading into the summer of 2017. While these projections are slightly lower than those developed last year, the current outlook is very different than it was in 2013, when planning reserve margins were expected to be below the then-existing target level of 13.75 percent for the foreseeable future.[48]

This current projection of planning reserve margins combined with relatively infrequent shortage pricing may raise doubts regarding the likelihood of announced generation coming on line as planned. Given the projections of continued low prices, investors of some of the new generation included in the Report on the Capacity, Demand and Reserves in the ERCOT Region (CDR) may choose to delay or even cancel their project. Additionally, the profitability analysis of existing base load resources casts doubt on the assumption embedded in the CDR that all existing generation will continue to operate. Hence, it is likely that the planning reserve margins will be lower than forecasted in the figure above.

[48] The target planning reserve margin of 13.75 percent was approved by the ERCOT Board of Directors in November 2010, based on a 1 in 10 loss of load expectation (LOLE). The PUCT recently directed ERCOT to evaluate planning reserve margins based on an assessment of the Economically Optimal Reserve Margin (EORM) and the Market Equilibrium Reserve Margin (MERM). *See* PUCT Project No. 42303, ERCOT Letter to Commissioners (Oct. 24, 2016).

With expectations for future natural gas prices to remain relatively low, the pressure on the ability of coal units in the ERCOT market to economically operate is not expected to subside any time soon. These challenging fuel market economics exist regardless of the future of environmental regulations that could require additional capital investment for existing coal units.

The retirement of uneconomic generation should not in any way be viewed as failure to provide resource adequacy. Having the right pricing signals to encourage sufficient and efficient generation signals is the goal. Most of the coal units facing the greatest price and environmental pressure have been operating for more than thirty-five years. Similar to the forces that have led to the retirement of less efficient natural gas fueled steam units, the retirement of older, less efficient coal units is an expected market outcome.

D. Ensuring Resource Adequacy

One of the primary goals of an efficient and effective electricity market is to ensure that, over the long term, there is an adequate supply of resources to meet customer demand plus any required installed or planning reserves. In a region like ERCOT, where customer requirements for electricity have been and are expected to continue to increase, even with growing demand response efforts, maintaining adequate supply requires capacity additions. To incent these additions the market design must provide revenues such that the marginal resource receives revenues sufficient to make that resource economic. In this context, "economic" includes both a return of and on capital investment.

Generators earn revenues from three sources: energy prices during non-shortage, energy prices during shortage and capacity payments. The capacity payments generators receive in ERCOT are related to the provision of ancillary services. Ancillary service payments are a small contributor, approximately $5 per kW-year. Setting ancillary service payments aside, generator revenue in ERCOT is overwhelmingly derived from energy prices under both shortage and non-shortage conditions.

Expectations for energy pricing under non-shortage conditions are the same regardless of whether payments for capacity exist. In ERCOT, with no capacity payments available, the amount a generator may receive from energy pricing under shortage conditions must be large enough to provide the necessary incentives for new capacity additions. This will occur when energy prices are allowed to rise substantially during times when the available supply is insufficient to simultaneously meet both energy and minimum operating reserve requirements.

Ideally, energy and reserve prices during shortages should reflect the diminished system reliability under these conditions, which is equal to the increased probability of "losing" load times the value of the lost load. 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 exclusively on these relatively infrequent occurrences of high prices to provide the appropriate price signal for demand response and new investment, when required. In this way, energy-only markets can provide price signals that will sustain a portfolio of resources to be used in real-time to satisfy the needs of the system. However, this portfolio may not include enough capacity to meet a specified target for planning reserves.

Faced with reduced levels of generation development activity coupled with increasing loads that resulted in falling planning reserve margins, in 2012 and 2013 the PUCT devoted considerable effort deliberating issues related to resource adequacy. In September 2013 the PUCT Commissioners directed ERCOT to move forward with implementing ORDC, a mechanism designed to ensure effective shortage pricing when operating reserve levels decrease. Over the long term, a co-optimized energy and operating reserve market will provide more accurate shortage pricing. Planning reserves should continue to be monitored to determine whether shortage pricing alone is leading to the desired level of planning reserves.

VII. Analysis of Competitive Performance

In this section, market power is evaluated from two perspectives – structural (does market power exist) and behavioral (have attempts been made to exercise it). Market structure is examined by using a pivotal supplier analysis that indicates the frequency with which a supplier was pivotal at higher load levels. This section also includes a summary of the Voluntary Mitigation Plans in effect during 2016. Market participant conduct is evaluated by reviewing measures of physical and economic withholding. These withholding patterns are further examined relative to the level of demand and the size of each supplier's portfolio. Based on these analyses, we find the overall performance of the ERCOT wholesale market to be competitive in 2016.

A. Structural Market Power Indicators

The market structure is analyzed by using the Residual Demand Index (RDI). 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.[49] 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 if 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 could raise prices significantly by withholding resources.

Figure 86 shows the ramp-constrained RDI relative to load for all hours in 2016. The trend line indicates a strong positive relationship between load and the RDI. The 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. To the extent that the resources scheduled by the largest QSEs are not controlled by or provide revenue to the QSE, the RDIs will tend to be slightly overstated.

49 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.

Figure 86: Residual Demand Index

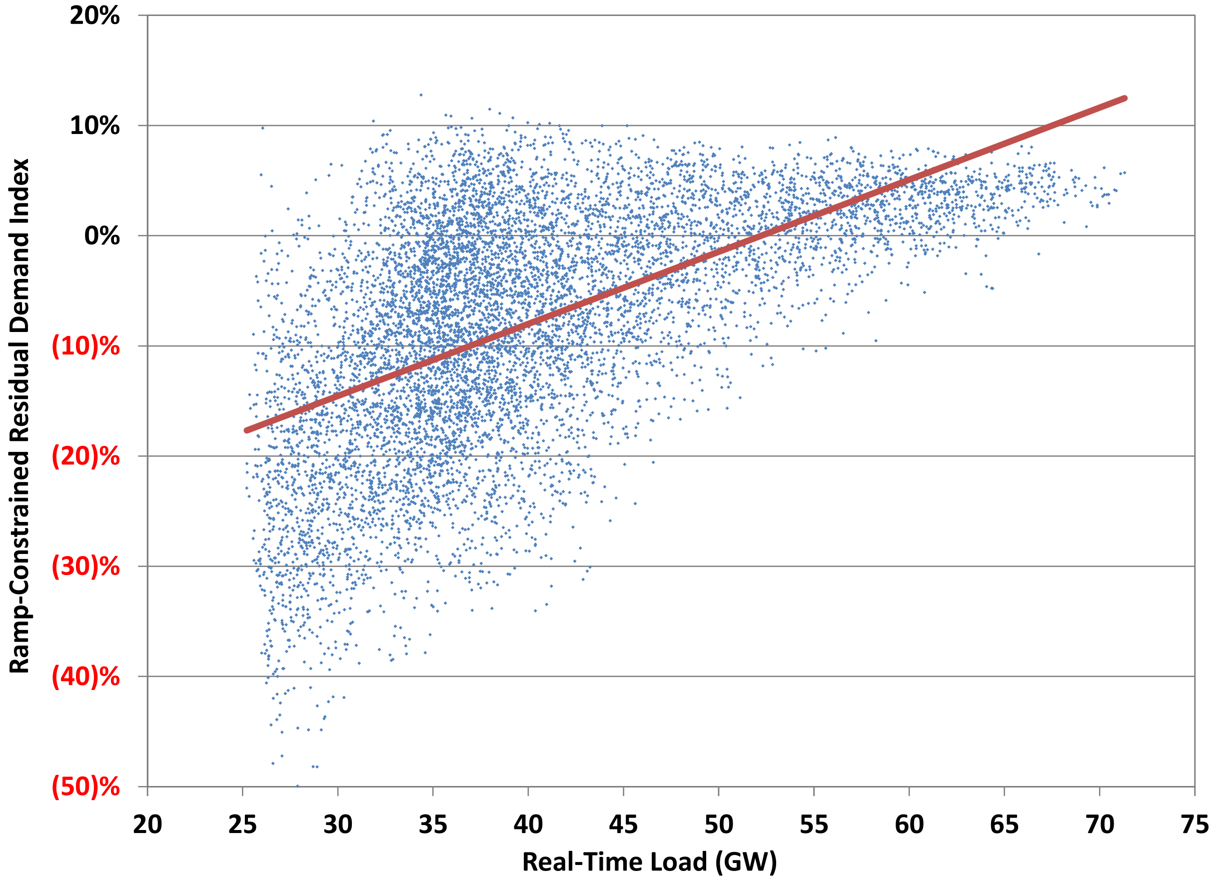
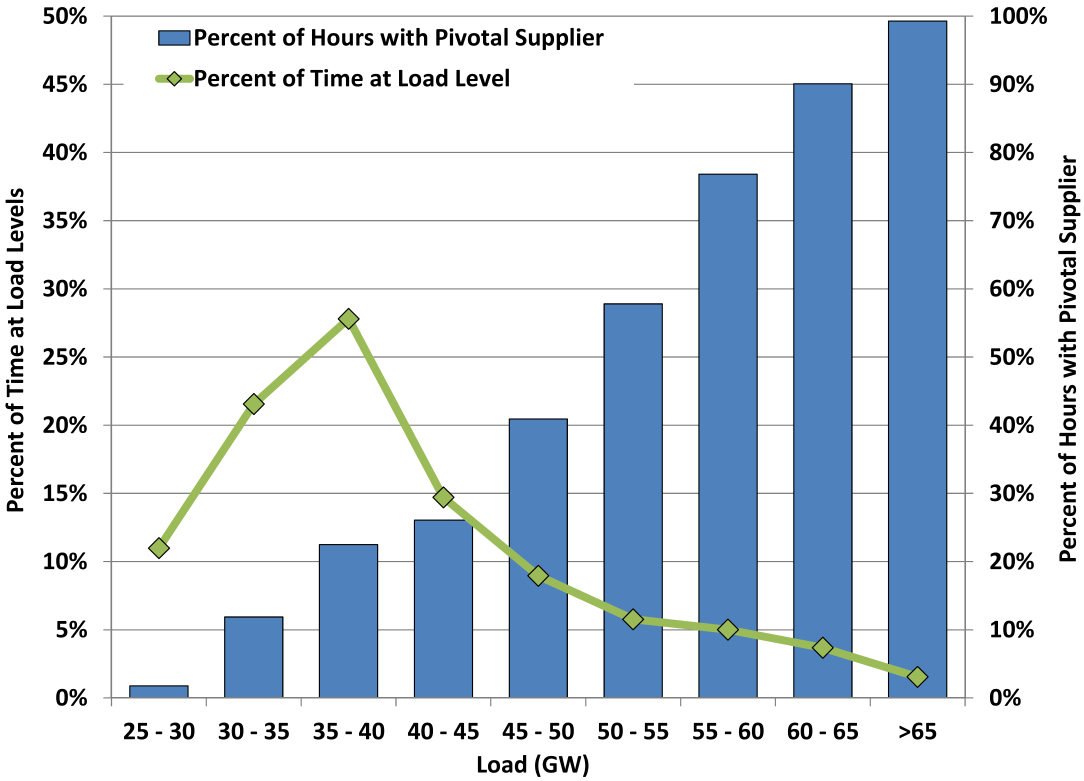


Figure 87 below summarizes the results of the RDI analysis by displaying the percent of time at each load level there was a pivotal supplier. The figure also displays the percent of time each load level occurs.

Figure 87: Pivotal Supplier Frequency by Load Level



At loads greater than 65 GW there was a pivotal supplier 99 percent of the time. This is expected because at high load levels, larger suppliers are more likely to be pivotal because other suppliers' resources are more fully utilized serving the load. The frequency of relatively high loads increased in 2016. This led to an increase in the pivotal supplier frequency to 28.5 percent of all hours in 2016, up from 26 and 23 percent of all hours in 2015 and 2014, respectively. This indicates that market power continues to be a potential concern in ERCOT and underscores the need for effective mitigation measures to address it.

Inferences regarding market power cannot be made solely from pivotal supplier data. Bilateral and other financial contract obligations can affect a supplier's potential market power. For example, a small supplier selling energy only in the real-time energy market may have a much greater incentive to exercise market power than a large supplier with substantial long-term sales contracts. The RDI measure shown in the previous figures do 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.

It should be noted that the analysis above evaluates the structure of the entire ERCOT market. In general, local market power in narrower areas that can become isolated by transmission constraints raise more substantial competitive concerns. As more fully discussed in Section V, Reliability Commitments, this local market power is addressed through: (a) structural tests that determine "non-competitive" constraints that can create local market power; and (b) the application of limits on offer prices in these areas.

Voluntary Mitigation Plans

Voluntary Mitigation Plans (VMPs) existed for three market participants in 2016. Generation owners are motivated to enter into VMPs because adherence to a plan approved by the PUCT constitutes an absolute defense against an allegation of market power abuse through economic withholding with respect to behaviors addressed by the plan. This increased regulatory certainty afforded to a generation owner regarding its energy offers in the ERCOT real-time market must be balanced by appropriate protections against a potential abuse of market power in violation of PURA §39.157(a) and 16 TEX. ADMIN. CODE § 25.503(g)(7).

VMPs should promote competitive outcomes and prevent abuse of market power through economic withholding in the ERCOT real-time energy market. The same restrictions are not required in forward energy markets (e.g., the ERCOT day-ahead market) because the prices in forward energy markets are derived from the real-time energy prices. Because forward energy markets are voluntary and the market rules do not inhibit arbitrage between the forward energy markets and the real-time energy market, competitive outcomes in the real-time energy market serve to discipline the potential abuse of market power in the forward energy markets.

In 2016, there were three market participants with approved VMPs – NRG, Calpine, and Luminant. NRG's plan, initially approved in June 2012 and modified in May 2014,[50] allows the company to offer some of its capacity at prices up to the system-wide offer cap. Specifically, up to 12 percent of the difference between the high sustained limit and the low sustained limit – the dispatchable capacity – for each natural gas unit (5 percent for each coal/lignite unit) may be offered no higher than the higher of $500 per MWh or 50 times the natural gas price. Additionally, up to 3 percent of the dispatchable capacity for each natural gas unit may be offered no higher than the system-wide offer cap. The amount of capacity covered by these provisions is approximately 500 MW.

Calpine's VMP was approved in March of 2013.[51] Because its generation fleet consists entirely of natural-gas fueled combined cycle units, the details of the Calpine plan are somewhat different than NRG. Calpine may offer up to 10 percent of the dispatchable capacity of its portfolio at prices up to $500 per MWh. Additionally, Calpine may offer up to 5 percent of the dispatchable capacity of its portfolio at prices no higher than the system-wide offer cap. When approved, the

---

50 PUCT Docket No. 40488, Request for Approval of a Voluntary Mitigation Plan for NRG Companies Pursuant to PURA § 15.023(f) and P.U.C. Subst. R. 25.504(e), Order (Jul. 13, 2012); PUCT Docket No. 42611, Request for Approval of an Amended Voluntary Mitigation Plan for NRG Companies, Order (Jul. 11, 2014).

51 PUCT Docket No. 40545, Petition of Calpine Corporation for Approval of Voluntary Mitigation Plan, Order (Mar. 28, 2013).

amount of capacity covered by these provisions was approximately 500 MW. With recent additions to Calpine's generation fleet its current amount of offer flexibility has increased to approximately 700 MW.

Luminant received approval from the PUCT for a VMP in May 2015.[52] The Luminant plan is similar in many respects to the NRG plan. Under the VMP, Luminant is permitted to offer a maximum of 12 percent of the dispatchable capacity for its natural gas units (5 percent for coal/lignite units) at prices up to $500 per MWh and offer a maximum of 3 percent of the dispatchable capacity for natural gas units up to the system-wide offer cap. The amount of capacity covered by these provisions is slightly more than 500 MW. In addition, the plan contains a maximum offer for the approximately 1,000 MW of quick-start qualified combustion turbines owned by Luminant based on unit-specific verifiable costs and index prices for fuel and emissions.

Allowing small amounts of high-priced offers is intended to accommodate potential legitimate fluctuations in marginal cost that may exceed the base offer caps, such as operational risks, short-term fluctuations in fuel costs or availability, or other factors. However, all three VMPs contain a requirement that these offers, if offered in any hour of an operating day, must be offered in the same price and quantity pair for all hours of the operating day. This provision, along with the quantity limitations, significantly reduces the potential that the VMPs will allow market power to be exercised.

The final key element in the VMPs is the termination provisions. The approved VMPs may be terminated by the Executive Director of the PUCT with three business days' notice, subject to ratification by the Commission. PURA defines market power abuses as "practices by persons possessing market power that are unreasonably discriminatory or tend to unreasonably restrict, impair, or reduce the level of competition."[53] The exercise of market power may not rise to the level of an abuse of market power if it does not unreasonably impair competition, which would typically involve profitably raising prices significantly above the competitive level for a significant period of time. Thus, although the offer thresholds provided in the VMPs are designed to promote competitive market outcomes, the short termination provision provides additional assurance that any unintended consequences associated with the potential exercise of market power can be addressed in a timely manner rather than persisting and rising to the level of an abuse of market power.

52 PUCT Docket No. 44635, Request for Approval of a Voluntary Mitigation Plan for Luminant Companies Pursuant to PURA § 15.023(f) and P.U.C. Subst. R. 25.504(e), Order Approving VMP Settlement (May 22, 2015).

53 PURA § 39.157(a).

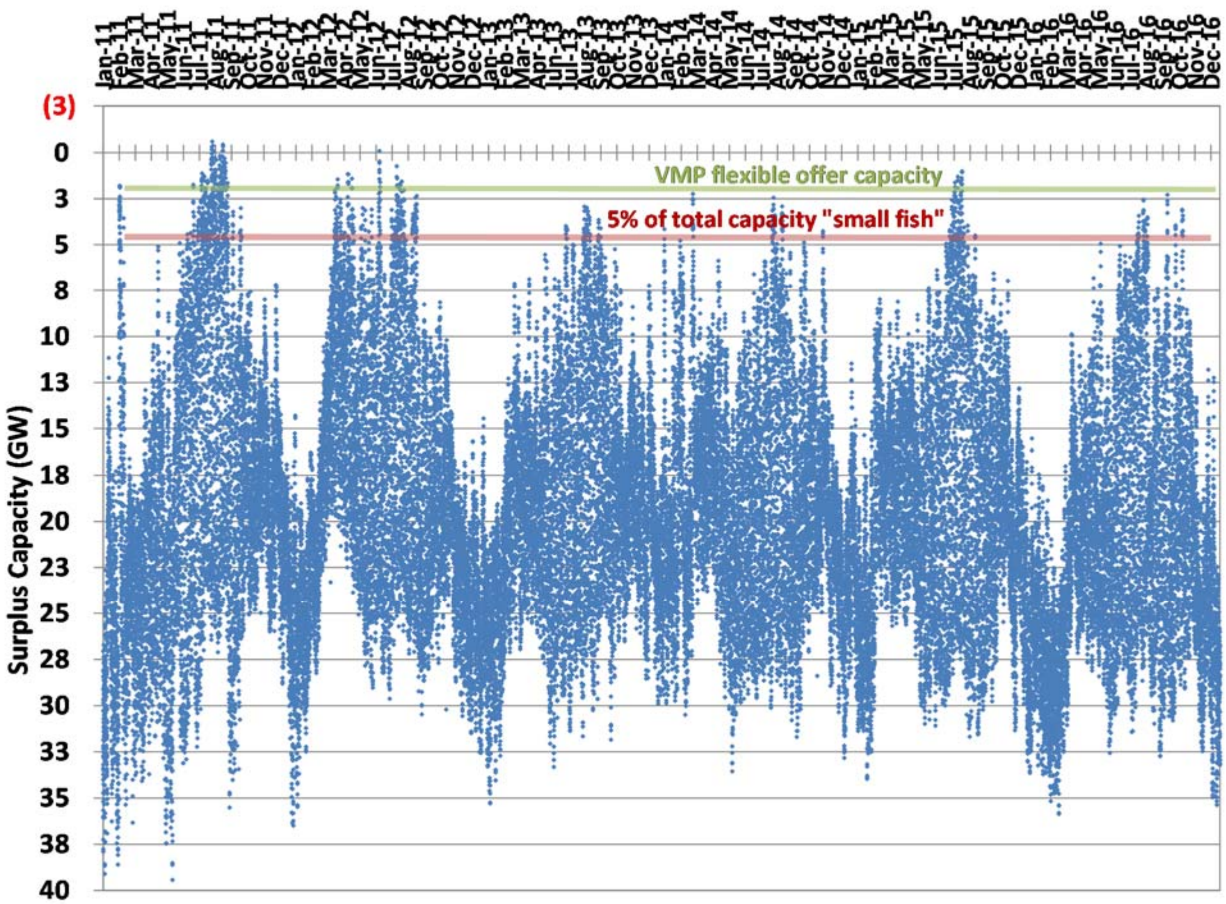
The amount of offer flexibility afforded by the VMPs is small when compared to the offer flexibility that small participants – those with less than 5 percent of total ERCOT capacity – are granted under 16 TEX. ADMIN. CODE § 25.504(c). Although 5 percent of total ERCOT capacity may seem like a small amount, the potential market impacts of a market participant whose size is just under the 5 percent threshold choosing to exercise flexibility and offering a significant portion of their fleet at very high prices can be large.

The figure below shows the amount of surplus capacity available in each hour of every day from 2011 to 2016. For this analysis, surplus capacity is defined as online generation plus any offline capacity that was available day ahead, plus DC Tie imports (minus exports), minus responsive reserves provided by generation, regulation up capacity, and load. Every hour of the past four years has had surplus capacity. Only during 2011 (12 hours) and for one hour in 2012 was ERCOT was unable to meet load and maintain all operating reserve obligations.

Currently, the 5 percent "small fish" threshold is roughly 4,000 MW, as indicated by the red line in Figure 88. There were 572 hours over the past six years with less than 4,000 MW of surplus capacity.[54] During these times a large "small fish" would have been pivotal and able to increase the market clearing price through its offer, potentially as high as the system-wide offer cap. In contrast, the combined amount of capacity afforded offer flexibility under the VMPs granted to NRG, Calpine, and Luminant totals less than 1,800 MW of capacity. This amount of capacity would have been pivotal for a total of 120 hours across the past six years, with none occurring in 2016.

[54] Surplus capacity was less than 4,000 MW for 296 hours in 2011, 154 hours in 2012, 15 hours in 2013, 26 hours in 2014, 56 hours in 2015, and 25 hours in 2016.

Figure 88: Surplus Capacity



B. Evaluation of Supplier Conduct

The previous subsection presented a structural analysis that supports inferences about potential market power. This subsection provides the results of evaluating actual participant conduct to assess whether market participants have attempted to exercise market power through physical or economic withholding. First, unit deratings and forced outages are examined to detect physical withholding. This is followed by an evaluation of the "output gap," used to detect economic withholding.

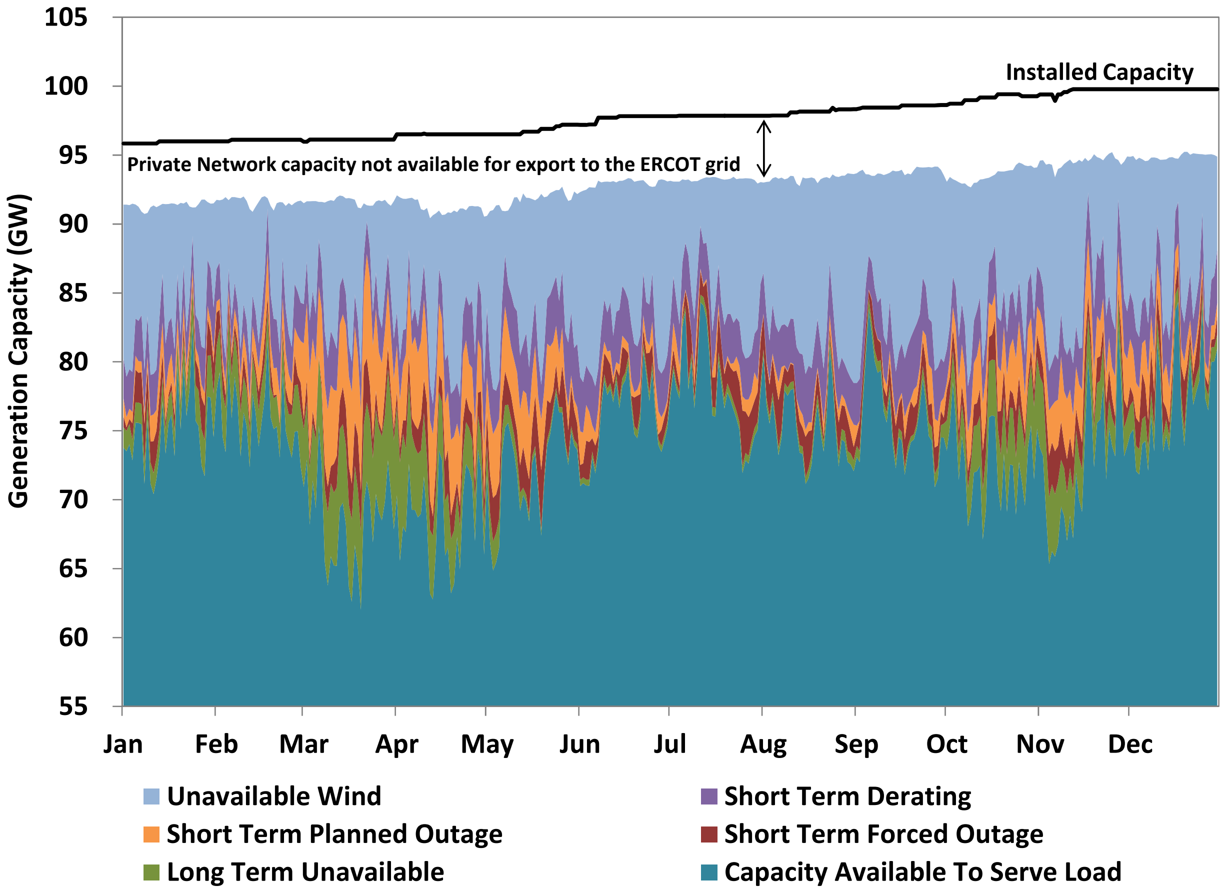
In a single-price auction like the real-time energy market, suppliers may attempt to exercise market power by withholding resources. The purpose of withholding is to cause more expensive resources to set higher market clearing prices, allowing the supplier to profit on its other sales in the real-time 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. This 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.

Generation Outages and Deratings

Some portion of installed capacity is commonly unavailable because of generator outages and deratings. Due to limitations in outage data, the outage type must be inferred. The outage type can be inferred by cross-referencing unit status information communicated to ERCOT with scheduled outages. If there is a corresponding scheduled outage, the unit is considered to be on a planned outage. If not, it is considered to be a forced outage. The derated capacity is defined as the difference between the summertime maximum capacity of a generating resource and its actual capability as communicated to ERCOT on a continuous basis. It is very common for generating capacity to be partially derated (e.g., by 5 to 10 percent) because the resource cannot achieve its installed capacity level due to technical or environmental factors (e.g., component equipment failures or ambient temperature conditions). Wind generators rarely produce at the installed capacity rating due to variations in available wind input. Because such a large portion of derated capacity is related to wind generation it is shown separately in the following evaluation of long-term and short-term deratings.

Figure 89 shows a breakdown of total installed capacity for ERCOT on a daily basis during 2016. This analysis includes all in-service and switchable capacity. From the total installed capacity the following are subtracted: (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 planned outages; (e) short-term forced outages; and (e) long-term outages and deratings greater than 30 days. What remains is the capacity available to serve load.

Figure 89: Reductions in Installed Capacity



Outages and deratings of non-wind generators fluctuated between 4 and 19 GW, as shown in Figure 89, while wind unavailability varied between 1 and 15 GW. Short-term planned outages were largest between March and April and smallest during the summer months, which is consistent with expectations. Short-term forced outages and deratings had no discernable seasonal pattern, occurring throughout the year.

The quantity of long-term (greater than 30 days) unavailable capacity, peaked in March at 7 GW, reduced to less than 1 GW during the summer months, and increased to 5 GW in November. This pattern reflects the choice by generation owners to schedule long duration outages during the spring and fall so as to ensure the units are available during the high load summer season when the units have a higher likelihood of operating.

The next analysis focuses specifically on short-term planned outages and forced outages and deratings of non-wind units because these classes of outages and deratings are the most likely to be used to physically withhold units in an attempt to raise prices. Figure 90 shows the average magnitude of the outages and deratings lasting less than 30 days for the year and for each month during 2016.

Figure 90: Short-Term Outages and Deratings

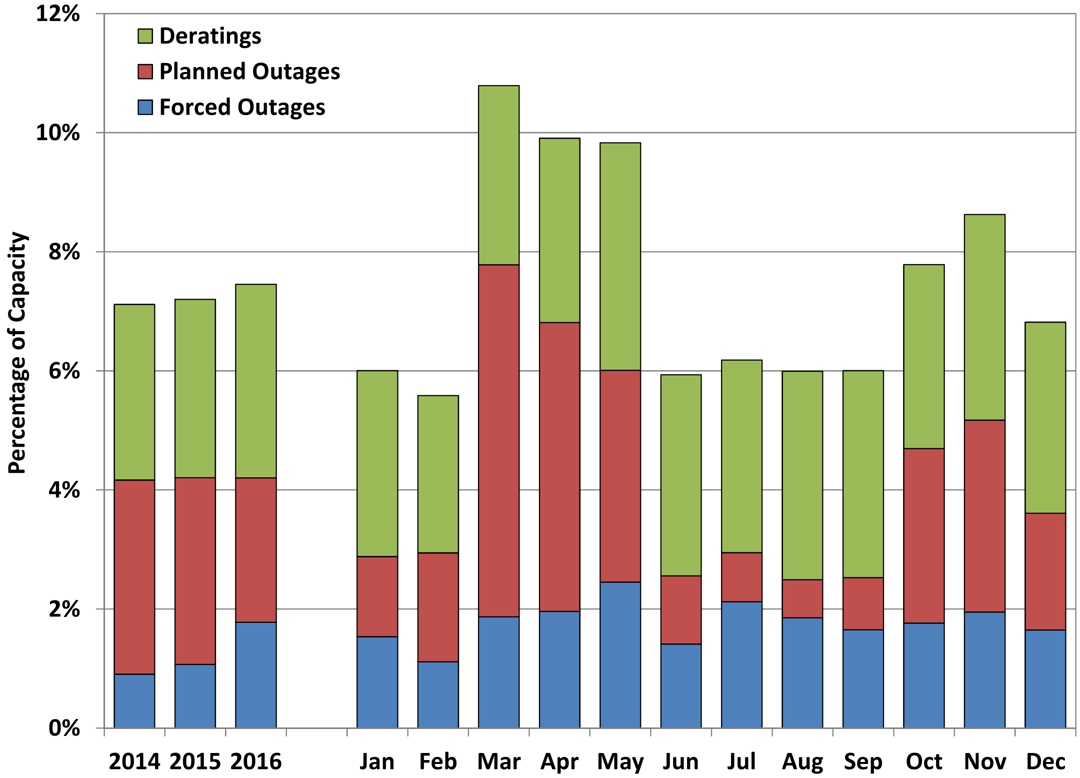


Figure 90 shows that total short-term deratings and outages were as large as 10.8 percent of installed capacity in March, and averaged around 6 percent during the summer. Most of this fluctuation was due to anticipated planned outages. The amount of capacity unavailable during 2016 averaged 7.5 percent of installed capacity. This is a slight increase from 7.2 percent experienced in 2015 and 7.1 percent experienced in 2014. Overall, the fact that outages and deratings are lowest during the summer when load is expected to be highest is consistent with expectations in a competitive market.

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 subsection is to differentiate justifiable deratings and outages from physical withholding. Physical withholding is tested for 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 86 and Figure 87 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 occurring, one 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 peak periods.

Figure 91 shows the average relationship of short-term deratings and forced outages as a percentage of total installed capacity to real-time load levels for large and small suppliers during summer months. 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.

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 five largest suppliers in ERCOT. The small supplier category includes the remaining suppliers.

Figure 91: Outages and Deratings by Load Level and Participant Size, June-August

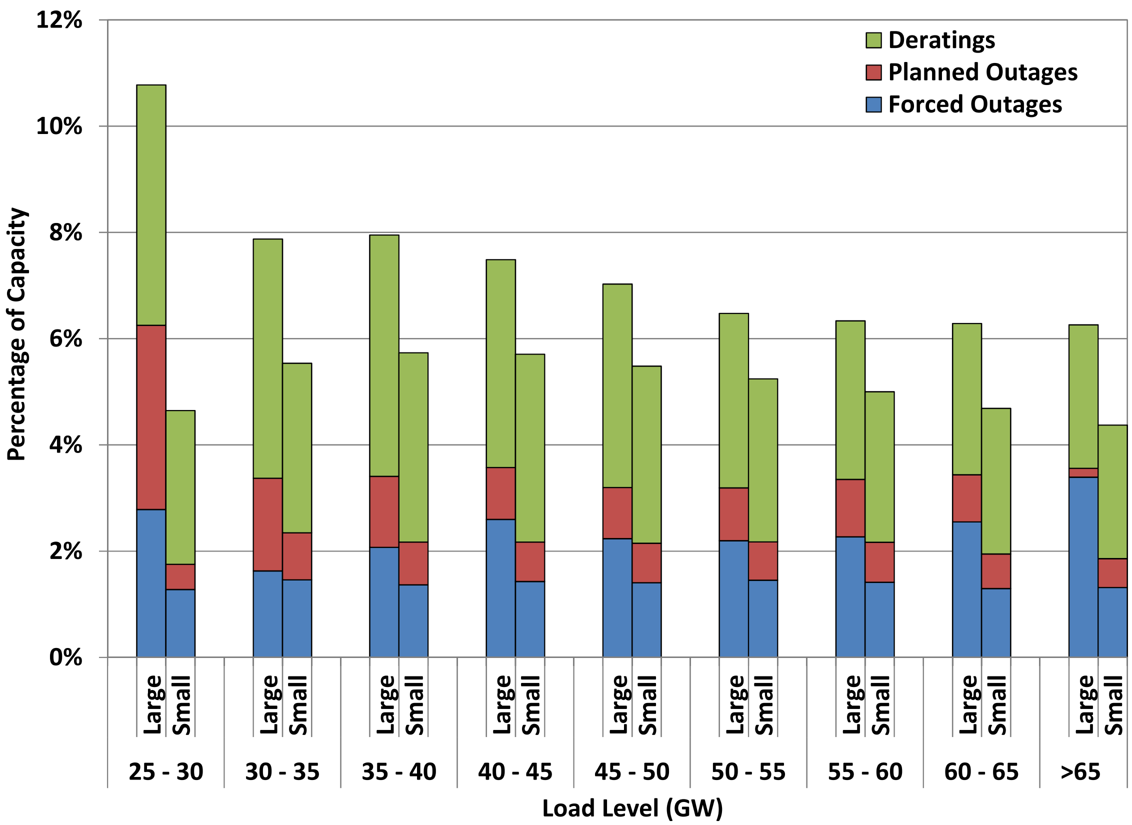


Figure 91 suggests that as demand for electricity increases, all market participants tend to make more capacity available to the market by scheduling planned outages during low load periods. Since small participants have less incentive to physically withhold capacity, the outage rates for small suppliers serves as a good benchmark for competitive behavior expected from the larger suppliers. For large suppliers, the percent of derated capacity declined at higher load levels, whereas for small providers the percent of derated capacity was fairly constant across all load levels. Although large providers had slightly higher forced outage rates than small providers, their level – 2.4 percent – does not raise potential competitive concerns.

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 the quantity of energy that is not being produced by online resources even though the output is economic to produce 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.

A resource is evaluated for inclusion in the output gap when it is committed and producing at less than full output. Energy not produced from a committed resource is included in the output gap if the real-time energy price exceeds that unit's mitigated offer cap by at least $30 per MWh.[55] The mitigated offer cap serves as a proxy for the marginal production cost of energy from that resource.

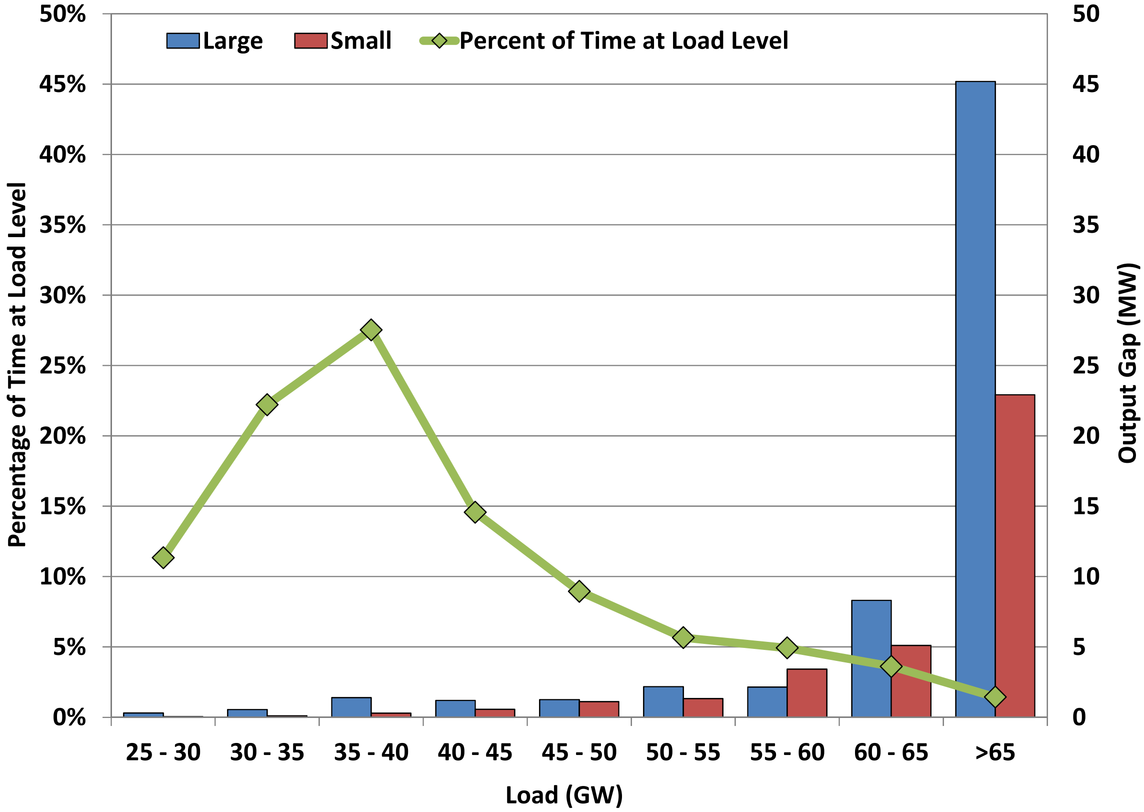
Before presenting the results of the output gap analysis, a description of ERCOT's two-step dispatch software is required. In the first step, the dispatch software calculates output levels (base points) and associated locational marginal prices using the participants' offer curves and only considering transmission constraints that have been deemed competitive. These "reference prices" at each generator location are compared with the generator's mitigated offer cap, and the higher of the two is used to formulate the offer curve for that generator during 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

55 Given the low energy prices during 2016, the output gap margin was reduced to $30 for purposes of this analysis. Prior to 2015, the State of the Market report used $50 for the output gap margin.

sent based on the first step. It is only used to screen whether a market participant is withholding in a manner that may influence the reference price.

Figure 92: Incremental Output Gap by Load Level and Participant Size – Step 1

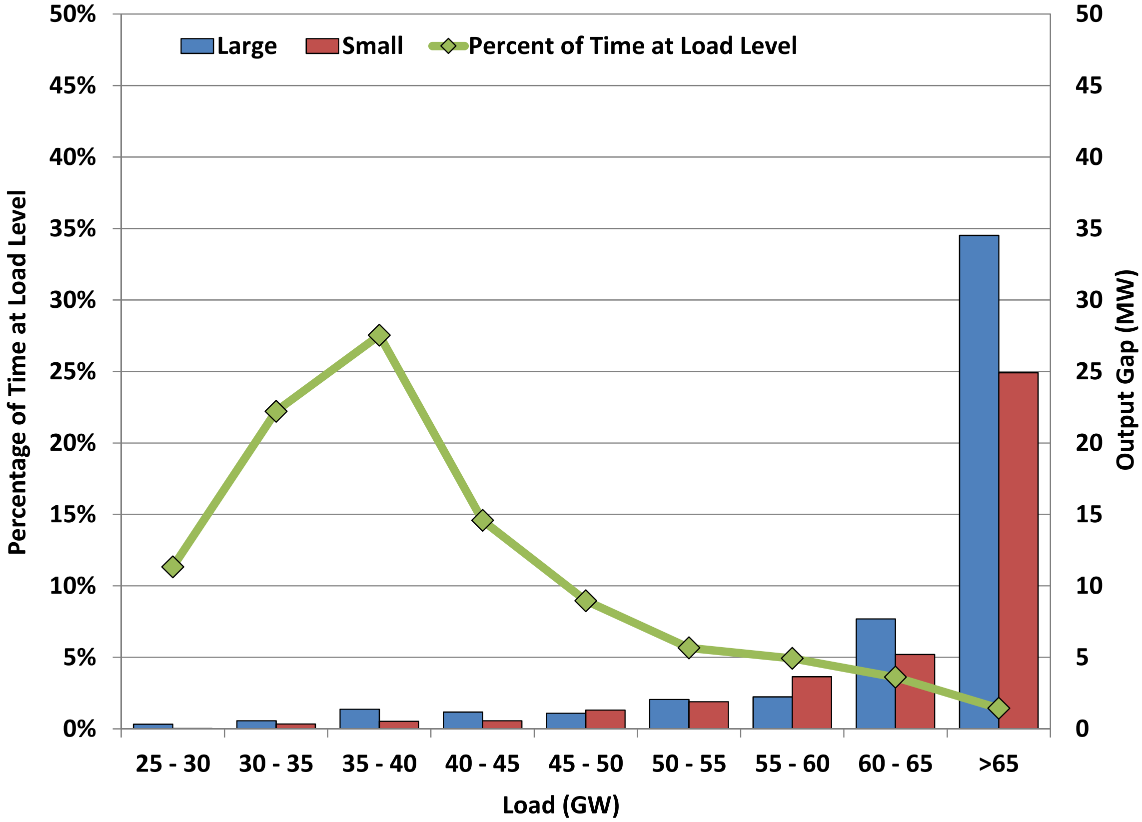


The results of the analysis shown in Figure 92 indicate that only very small amounts of capacity would be considered part of the first step output gap.

Figure 93 below shows the ultimate output gap levels, 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 being influenced by a market participant raising prices.

Similar to the previous analysis, Figure 93 also shows very small quantities of capacity that would be considered part of this output gap.

Figure 93: Incremental Output Gap by Load Level and Participant Size – Step 2



These results show that potential economic withholding levels were extremely low for the largest suppliers and small suppliers alike in 2016. Output gaps of the largest suppliers are routinely monitored individually and were found to be consistently low across all load levels. These results, together with our evaluation of the market outcomes presented in this report, allow us to conclude that the ERCOT market performed competitively in 2016.