# 2014 STATE OF THE MARKET REPORT FOR THE ERCOT WHOLESALE ELECTRICITY MARKETS

POTOMAC ECONOMICS, LTD.

Independent Market Monitor for the ERCOT Wholesale Market

July 2015

## TABLE OF CONTENTS

Execu	itive Summary	i
A.	Introduction	i
B.	Review of Real-Time Market Outcomes	ii
C.	Review of Day-Ahead Market Outcomes	vi
D.	Transmission and Congestion	x
E.	Demand and Supply	xii
F.	Resource Adequacy	xvi
G.	Analysis of Competitive Performance	xxi
H. I. R	Recommendations	xxv
A.	Real-Time Market Prices	1
B.	Real-Time Prices Adjusted for Fuel Price Changes	10
C.	Aggregated Offer Curves	14
D.	Prices at the System-Wide Offer Cap	15
E.	Real-Time Price Volatility	19
F.	Mitigation	21
G.	Revenue Sufficiency	23
II.	Review of Day-Ahead Market Outcomes	25
A.	Day-Ahead Market Prices	25
B.	Day-Ahead Market Volumes	28
C.	Point to Point Obligations	30
D.	Ancillary Services Market	33
III.	Transmission and Congestion	43
A.	Summary of Congestion	43
B.	Real-Time Constraints	45
C.	Day-Ahead Constraints	50
D.	Congestion Revenue Rights Market	51
IV.	Demand and Supply	61
A.	ERCOT Loads in 2014	61
B.	Generation Capacity in ERCOT	64
C.	Demand Response Capability	79
V.	Resource Adequacy	85
A.	Net Revenue Analysis	85
B.	Effectiveness of the Scarcity Pricing Mechanism	94
C.	Planning Reserve Margin	96
D.	Ensuring Resource Adequacy	99
E.	Operating Reserve Demand Curve Implementation	. 101
VI.	Analysis of Competitive Performance	103
А.	Structural Market Power Indicators	103
В.	Evaluation of Supplier Conduct	109

## LIST OF FIGURES

Figure 1: Average All-in Price for Electricity in ERCOT	2
Figure 2: Average Real-Time Energy Market Prices	3
Figure 3: Effective Real-Time Energy Market Prices	4
Figure 4: Comparison of All-in Prices Across Markets	5
Figure 5: ERCOT Price Duration Curve	6
Figure 6: ERCOT Price Duration Curve – Top 5% of Hours	6
Figure 7: Average Real-Time Energy Prices and Number of Price Spikes	7
Figure 8: Zonal Price Duration Curves	8
Figure 9: West Zone and ERCOT Price Duration Curves	9
Figure 10: Implied Marginal Heat Rate Duration Curve – All Hours	10
Figure 11: Implied Marginal Heat Rate Duration Curve –	11
Figure 12: Monthly Average Implied Heat Rates	12
Figure 13: Heat Rate and Load Relationship	13
Figure 14: Aggregated Generation Offer Stack - Annual	14
Figure 15: Aggregated Generation Offer Stack – Summer	15
Figure 16: Prices at the System-Wide Offer Cap	16
Figure 17: Load, Reserves and Prices in August	17
Figure 18: Real-Time Energy Price Volatility (May – August)	19
Figure 19: Monthly Price Variation	20
Figure 20: Mitigated Capacity by Load Level	22
Figure 21: Capacity Subject to Mitigation	23
Figure 22: Real-Time Congestion Rent and Payments	24
Figure 23: Convergence Between Forward and Real-Time Energy Prices	27
Figure 24: Day-Ahead and Real-Time Prices by Zone	28
Figure 25: Volume of Day-Ahead Market Activity by Month	29
Figure 26: Volume of Day-Ahead Market Activity by Hour	30
Figure 27: Point to Point Obligation Volume	31
Figure 28: Average Profitability of Point to Point Obligations	32
Figure 29: Point to Point Obligation Charges and Payments	33
Figure 30: Ancillary Service Capacity	34
Figure 31: Ancillary Service Prices	35
Figure 32: Ancillary Service Costs per MWh of Load	36
Figure 33: Responsive Reserve Providers	37
Figure 34: Non-Spin Reserve Providers	38
Figure 35: Frequency of SASM Clearing	39
Figure 36: Ancillary Service Quantities Procured in SASM	42
Figure 37: Frequency of Binding and Active Constraints	44
Figure 38: Real-Time Congestion Costs	45

Top Ten Real-Time Constraints	46
Most Frequent Real-Time Constraints	49
Top Ten Day-Ahead Congested Areas	50
Day-Ahead Congestion Costs by Zone	51
CRR Auction Revenue	52
CRR Auction Revenue and Payment Received	53
CRR Auction Revenue, Payments and Congestion Rent	54
CRR Shortfalls and Derations	55
West Hub to West Load Zone Price Spreads	57
North Hub to North Load Zone Price Spreads	58
South Hub to South Load Zone Price Spreads	59
Houston Hub to Houston Load Zone Price Spreads	60
Annual Load Statistics by Zone	62
Load Duration Curve – All Hours	63
Load Duration Curve – Top Five Percent of Hours	64
Installed Capacity by Technology for Each Zone	65
Installed Capacity by Type: 2007 to 2014	66
Annual Generation Mix	67
Average Wind Production	68
Summer Wind Production vs. Load	69
Summer Renewable Production	70
Wind Production and Curtailment	71
Net Load Duration Curves	72
Top and Bottom Ten Percent of Net Load	73
Average On-line Reserves	75
Frequency of Reliability Unit Commitments	76
Reliability Unit Commitment Capacity	78
Daily Average of Responsive Reserves Provided by Load Resources	80
Pricing During Load Deployments	83
Combustion Turbine Net Revenues	86
Combined Cycle Net Revenues	87
Estimated Net Revenue by Zone and Unit Type	88
Combustion Turbine Net Revenue Comparison between Markets	92
Combined Cycle Net Revenue Comparison Between Markets	93
Peaker Net Margin	95
Projected Reserve Margins	96
Peak Demand Forecast Comparison	97
Reserve Margins in Other Regions	98
Operating Reserve Demand Curve	101
Online Reserve Adder	102
Residual Demand Index	104
	Top Ten Real-Time Constraints. Most Frequent Real-Time Constraints. Top Ten Day-Ahead Congested Areas. Day-Ahead Congestion Costs by Zone CRR Auction Revenue and Payment Received CRR Auction Revenue and Payment Received CRR Auction Revenue, Payments and Congestion Rent. CRR Shortfalls and Derations. West Hub to West Load Zone Price Spreads North Hub to North Load Zone Price Spreads. North Hub to North Load Zone Price Spreads. South Hub to South Load Zone Price Spreads. Annual Load Statistics by Zone Load Duration Curve – All Hours. Load Duration Curve – Top Five Percent of Hours. Installed Capacity by Type: 2007 to 2014. Annual Generation Mix. Average Wind Production Summer Renewable Production. Summer Renewable Production. Wind Production and Curtailment. Net Load Duration Curves. Frequency of Reliability Unit Commitments. Reliability Unit Commitments. Reliability Unit Commitments. Reliability Unit Commitments. Reliability Unit Commitments. Combustion Turbine Net Revenues. Combustion Turbine Net Revenues. Pricing During Load Deployments. Combustion Turbine Net Revenues. Projected Reserve Margins. Peak Demand Forecast Comparison. Reserve Margins in Other Regions. Peak Demand Forecast Comparison. Reserve Margins in Other Regions. Operating Reserve Demand Curve. Online Reserve Adder . Residual Demand Index.

Figure 80:	Pivotal Supplier Frequency by Load Level	105
Figure 81:	Surplus Capacity	108
Figure 82:	Reductions in Installed Capability	111
Figure 83:	Short-Term Outages and Deratings	112
Figure 84:	Outages and Deratings by Load Level and Participant Size	114
Figure 85:	Incremental Output Gap by Load Level and Participant Size – Step 1	116
Figure 86:	Incremental Output Gap by Load Level and Participant Size – Step 2	117

### LIST OF TABLES

Table 1:	15-Minute Price Changes as a Percentage of Annual Average Prices	20
Table 2:	Ancillary Service Deficiency	40
Table 3:	Irresolvable Constraints	48

#### **EXECUTIVE SUMMARY**

#### A. Introduction

This report reviews and evaluates the outcomes of the ERCOT wholesale electricity markets in 2014, 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 procedures, 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).

Key findings and statistics from 2014 include the following:

- The ERCOT wholesale market performed competitively in 2014.
- The ERCOT-wide load-weighted average real-time energy price was \$40.64 per MWh in 2014, a 21 percent increase from \$33.71 per MWh in 2013. The increase was primarily driven by higher natural gas prices in 2014.
  - The average price for natural gas was 17 percent higher in 2014 than in 2013, increasing from \$3.70 per MMBtu in 2013 to \$4.32 per MMBtu in 2014. The highest prices occurred early in the year when unusually cold weather throughout the U.S. resulted in much higher and more volatile natural gas prices.
  - Loads in 2014 were slightly higher than 2013, and the frequency of shortage conditions increased. Total ERCOT load in 2014 was 2.5 percent higher than 2013, although the peak load decreased by 1.2 percent.
  - Prices at the system-wide offer cap were experienced in dispatch intervals which totaled 1.56 hours in 2014.
- The total congestion revenue generated by the ERCOT real-time market in 2014 was \$708 million, an increase of 52 percent from 2013. This increase was due the combination of higher gas prices, which generally increases the costs of re-dispatching generation to manage network flows, and more frequent congestion in the South and Houston zones.
  - The two most costly constraints were transformer limitations. They were the Heights TNP 138/69 kV autotransformers in the Houston area and the Lytton Springs 345/138 kV autotransformer in the Austin area. Although in different parts of the state and occurring at different times, both were the result of outages of other nearby transmission facilities.

- The Rio Grande Valley was the most congested area in 2014 because of transmission outages scheduled to accommodate the construction of transmission upgrades in the area. A large contribution to total cost occurred in October, when a combination of generation and transmission outages led to a significant congestion. Ultimately, on October 8th the situation in the Valley required that firm load be curtailed.
- Net revenues provided by the market during 2014 were less than the amount estimated to be needed to support new greenfield generation. The increased shortage pricing levels did not substantially increase net revenues in 2014 because shortages were less frequent than average over the long term. Nonetheless, reserve margins in ERCOT are expected to exceed the minimum target for the next several years.

#### B. Review of Real-Time Market Outcomes

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

The next figure summarizes changes in energy prices and other market costs by showing the allin price of electricity, which is a measure of the total cost of serving load in ERCOT. The ERCOT-wide price is the load-weighted average of the real-time market prices from all load zones. Ancillary services costs are estimated based on total system demand and prices in the ERCOT markets for regulation, responsive reserves, and non-spinning reserves. Uplift costs are assigned market-wide on a load-ratio share basis to pay for costs associated with reliability unit commitments and reliability must run contracts. Starting June 1, 2014, with the implementation of the Operating Reserve Demand Curve, the real-time energy price includes the online reserve adder. In the figure below this adder is shown separated out from the energy price.



The largest component of the all-in cost of wholesale electricity is the energy cost. ERCOT average real-time all-in prices were 21 percent higher in 2014 than in 2013. The ERCOT-wide load-weighted average price was \$40.64 per MWh in 2014 compared to \$33.71 per MWh in 2013. The Online Reserve adder was \$0.26 per MWh for the last half of the year.

The increase in real-time energy prices was correlated with much higher fuel prices in 2014. The high correlation of natural gas prices and energy prices shown in the figure is consistent with expectations in a well-functioning market. Fuel costs constitute most of the marginal production costs for generating resources in ERCOT and competitive markets provide incentives for suppliers to submit offers consistent with marginal costs. The average natural gas price in 2014 was \$4.32 per MMBtu, a 17 percent increase compared to \$3.70 per MMBtu in 2013. Gas prices were highest in the first quarter when unusually cold weather throughout the U.S. resulted in much higher and more volatile natural gas prices. Ancillary services are a small portion of the

all-in price of energy and increased from \$1.03 in 2013 to \$1.51 in 2014. Uplift costs continue to be minimal in ERCOT.

The average real-time all-in electricity prices by zone from 2011 through 2014 are shown below:

	Average Real-Time Electricity Price				
	(\$ per MWh)				
	2011	2012	2013	2014	
ERCOT	\$53.23	\$28.33	\$33.71	\$40.64	
Houston	\$52.40	\$27.04	\$33.63	\$39.60	
North	\$54.24	\$27.57	\$32.74	\$40.05	
South	\$54.32	\$27.86	\$33.88	\$41.52	
West	\$46.87	\$38.24	\$37.99	\$43.58	
Natural Gas					
(\$/MMBtu)	\$3.94	\$2.71	\$3.70	\$4.32	

To depict how real-time energy prices vary by hour in each zone, the next figure shows the hourly average price duration curve in 2014 for four ERCOT load zones. The Houston, North and South load zones had similar prices over the majority of hours.



#### **Zonal Price Duration Curves**

The price duration curve for the West zone is noticeably different than the other zones, with more hours when prices exceeded \$50 per MWh. The number of hours with prices less than \$0 per MWh was very similar for all zones in 2014. This is notable since for the past several years the West zone has consistently had much more frequent occurrences of negative prices than the other zones. Significant transmission additions have lowered the frequency of depressed West zone prices due to transmission congestion during times of high wind output. However, the trend of local transmission constraints during low wind and high load conditions has continued and causes West prices to be higher than the rest of ERCOT.

As discussed in Section IV. Demand and Supply, overall demand for electricity was slightly higher in 2014 than in 2013. There were also more occasions when the available supply of generation resources was insufficient to satisfy system demand while maintaining required levels of operating reserves and, thus, more frequent instances of shortage pricing. Significant shortages result in energy prices being set at the system-wide offer cap. The frequency of this shortage pricing is shown in the following figure.



Prices at the System-Wide Offer Cap

The figure above shows the aggregate amount of time where the real-time energy price was set at the system-wide offer cap, displayed by month. There were no instances in 2014 of energy prices rising to the cap after the system-wide offer cap was increased to \$7,000 per MWh on June 1. Prices during 2014 were at the system-wide offer cap for only 1.56 hours, an increase from 0.22 hours in 2013 and a slight increase from the 1.51 hours experienced in 2012. All years were much lower than the 28.44 hours at the cap experienced in 2011 and the average amount expected of the long term in an energy-only market.

These results are not surprising because shortage pricing is highly variable year-to-year. When temperatures lead to weather-dependent loads that are significantly higher than normal or supply is less available than normal, the frequency of shortages tend to increase exponentially. Hence, one should expect that shortages will be very infrequent in normal or mild years, such as in 2012 and 2013. The occasions when prices reached the system-wide offer cap in 2014 were during colder than typical winter weather. Although the shortages in 2011 seemed relatively severe, adequate long-term incentives in the ERCOT market require shortages in excess of the value exhibited in 2011 every few years.

#### C. Review of Day-Ahead Market Outcomes

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

Day-ahead market performance is primarily evaluated by the degree to which its outcomes converge with those of the real-time market because the real-time market reflects actual physical supply and demand for electricity. In a well-functioning market, participants should eliminate sustained price differences on a risk-adjusted basis by making day-ahead purchases or sales to arbitrage the price differences away over the long-term.



**Convergence Between Forward and Real-Time Energy Prices** 

The figure above shows the price convergence between the day-ahead and real-time market, summarized by month. The simple average of day-ahead prices in 2014 was \$40 per MWh, compared to the simple average of \$38 per MWh for real-time prices. The average absolute difference between day-ahead and real-time prices was \$12.87 per MWh in 2014; higher than in 2013 when average of the absolute difference was \$9.86 per MWh.

This day-ahead premium is consistent with expectations due to the much higher volatility of realtime 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 as a result, having to buy back energy at real-time prices. This explains why the highest premiums tend to occur during the months with the highest relative demand and highest prices. The overall day-ahead premium increased in 2014 compared to 2013, as a result of the much higher premiums in January through March. Although peak loads during the winter are somewhat lower than those in the summer, loads during the first months of 2014 set record highs for that time of the year. Day-ahead premiums in ERCOT remain higher than observed in other organized electricity markets. 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 2014, it should not be expected over time that every month will always produce a day-ahead premium as 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 January and March).

Summarized in the figure below is the volume of day-ahead market activity by month. It shows that day-ahead purchases are approximately 50 percent of real-time load.



Volume of Day-Ahead Market Activity by Month

This figure also shows the volume of Point-to Point (PTP) Obligations, which are financial instruments purchased in the day-ahead market. Although these instruments do not themselves involve the direct supply of energy, they do provide the ability to avoid the congestion costs associated with transferring the delivery of energy from one location to another. 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 2014 was almost 5 percent higher than in 2013.

Under the nodal market, ancillary service offers are co-optimized as part of the day-ahead market clearing. This means that market participants do not have to include expectations of forgone energy sales in ancillary services capacity offers. As a result of ancillary services clearing prices explicitly accounting for the value of energy in the day-ahead market, ancillary services prices are highly correlated with day-ahead energy prices and, by extension, with real-time energy prices.



The figure above presents the average clearing prices of capacity for the four ancillary services, along with day-ahead and real-time energy prices. With average energy prices varying between \$25 and \$60 per MWh, the prices of ancillary services remained fairly stable throughout the year. Considering these costs on a per MWh of ERCOT load, total ancillary services costs increased 47 percent to \$1.51 per MWh.

#### D. Transmission and Congestion

The total congestion revenue generated by the ERCOT real-time market in 2014 was \$708 million, an increase of 52 percent from 2013. This increase was due the combination of higher gas prices, which generally increases the costs of re-dispatching generation to manage network flows, and more frequent congestion in the South and Houston zones. The next figure provides a comparison of the amount of time transmission constraints were binding or active at various load levels in 2012 through 2014.



**Frequency of Binding and Active Constraints** 

Binding transmission constraints are those for which the dispatch levels of generating resources are being altered in order to maintain transmission flows at reliable levels. The costs associated with this re-dispatch are the system's congestion costs and are priced in its Locational Marginal Prices (LMPs). Active transmission constraints are those that did not require a re-dispatch of generation.

The frequency of binding transmission constraints decreased again in 2014. There was a binding constraint only 44 percent of time in 2014, down from 55 percent of the time in 2013 and 60 percent in 2012. The likelihood of binding constraints increases at higher load levels, which is consistent with the results in ERCOT shown in the figure. However, it is noteworthy that there were binding constraints less than 90 percent of time during the very highest load levels in 2014, much lower than in prior years.

Because the overall frequency of binding constraints decreased in 2014, their effect on LMPs also decreased in 2014. Congestion in the West zone remained about the same as it was in 2013. Completion of the CREZ transmission projects has eliminated the longstanding limitations on the export of power from the west. However, binding constraints that limit transfers of power into the West continue, particularly under high load and low wind conditions. Constraints associated with oil and gas activity in the Eagle Ford Shale area and limitations serving the lower Rio Grande Valley had a larger impact in 2014. The figure below displays the ten constraints that generated the most real-time congestion.

The two most costly constraints were related to transformer overloads. Specifically, they were the Heights TNP 138/69 kV autotransformers in the Houston area and the Lytton Springs 345/138 kV autotransformer in the Austin area. Both were also the result of outages of other nearby transmission facilities.

The Rio Grande Valley was the most congested area in 2014 as a result of constraints occurring when other transmission facilities in the area were taken out of service to accommodate construction of transmission upgrades in the area. A large contribution to total cost occurred in October when a combination of generation and transmission outages led to a significant congestion. Ultimately, on October 8th the situation in the Valley required that firm load be curtailed.



### **Top Ten Real-Time Constraints**

#### E. Demand and Supply

This figure shows peak load and average load in each of the ERCOT zones from 2011 to 2014.



Annual Load Statistics by Zone

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

Total ERCOT load over the calendar year increased from 332 terawatt-hours (TWh) in 2013 to 340 TWh in 2014, an increase of 2.5 percent or an average of 960 MW every hour. Much of this increase occurred in the first quarter as extremely cold weather contributed to record levels of winter load.

Despite this increase in average load, the ERCOT coincident peak hourly demand decreased from 67,247 MW to 66,451 MW in 2014, a decrease of 795 MW, or 1.2 percent. The highest peak demand experienced in ERCOT remains 68,311 MW that occurred during August of 2011.

The changes in load at the zonal level are not the same as the ERCOT-wide changes. The growth rate of West zone average load was once again much higher, on a percentage basis, than the other zones. Peak load in the West zone increased nearly 500 MW in 2014. Peak load did not increase in any other zone.

Approximately 2.8 GW of new generation resources came online in 2014. Gas-fueled units accounted for 2.1 GW of the total additions, primarily from two new combined cycle units. The remaining gas additions were a new combined cycle unit built on an existing site of a retired gas steam unit, and the addition of a gas turbine at an existing generator location. The remaining resource additions were wind (0.7 GW) and small solar units. When unit retirements are included, the net capacity addition in 2014 was 1.6 GW. Natural gas generation continues to account for approximately 48 percent of total ERCOT installed capacity while the share of coal generation dropped slightly from 21 percent in 2013 to 20 percent in 2014.

Over the seven years from 2007 to 2014, more new wind and coal generation has been added than any other type of capacity. The sizable additions in these two categories have been more than offset by retirements of old natural gas-fired steam units. Nonetheless, the resulting installed capacity in 2014 was 1 GW more than in 2007. Comparatively, peak load in 2014 was greater than the 2007 peak load by more than 4 GW.

The figure below shows the percentage of annual generation from each fuel type for the years 2007 through 2014. The generation share from wind has increased every year, reaching 11 percent of the annual generation requirement in 2014, up from 3 percent in 2007. During the same period the percentage of generation provided by natural gas has ranged from a high of 45 percent in 2007 to a low of 38 percent in 2010. In 2014 the percentage of generation from natural gas was 41 percent, which was a very slight increase from the 2013 level.

Similarly, the percentage of generation produced by coal units was 36 percent in 2014, a small decrease from 37 percent in 2013.



**Annual Generation Mix** 

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 is approximately 23.4 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.

Increasing levels of wind resources in ERCOT also has important implications for the net load duration curve faced by the non-wind fleet of resources. For the following analysis, net load is defined as the system load minus wind production. The figure below shows the net load duration curves for the years 2011 through 2014, normalized as a percentage of peak load, and including 2007 as a point of reference. This figure shows the reduction of remaining energy demand 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. Wind generation erodes the amount of energy available to be served by baseload coal units, while doing very little to reduce the amount of capacity necessary to reliably serve peak load.



**Net Load Duration Curve** 

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

continues to increase. This outlook further reinforces the importance of efficient energy pricing during peak demand conditions and other times of system stress, particularly within the context of the ERCOT energy-only market design.

#### F. Resource Adequacy

#### 1. Long-Term Incentives: Net Revenue

One of the primary functions of the wholesale electricity market is to provide economic signals that will encourage the investment needed to maintain a set of resources that are adequate to satisfy the system's demands and reliability needs. These economic signals are evaluated by estimating the "net revenue" new resources would receive from the markets. Net revenue 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.



**Estimated Net Revenue** 

The figure above shows the results of the net revenue analysis for four types of hypothetical new units in 2013 and 2014. These are: (a) natural gas-fired combustion turbine, (b) natural gas-fired combined-cycle, (c) coal-fired generator, and (d) a nuclear unit. For the natural gas units, net

revenue is calculated by assuming the unit will produce energy in any hour for which it is profitable and by assuming it will be available to sell reserves and regulation in all other hours. For coal and nuclear technologies, net revenue is calculated solely from producing energy.

Overall, the net revenues in 2014 were higher than those in 2013 and 2012, and all three years were much lower than in 2011. This is not surprising given shortages have been very infrequent over the past three years. Shortage pricing plays a pivotal role in providing investment incentives in an energy-only market like ERCOT. In order to provide adequate incentives, some years must exhibit an extraordinary number of shortages and net revenues that are multiples of annual net revenues needed to support investment.

The figure above also shows that the 2014 net revenue for new natural gas-fired units was somewhat higher than 2013 levels, primarily because of higher gas prices during the first quarter of 2014. Net revenues for coal and nuclear technologies increased by larger amounts from 2013 to 2014 because they benefit from the increase in natural gas prices.

Despite these increases, the net revenues produced by the ERCOT markets in 2014 were lower than the estimated annualized cost of investing in any of these new technologies.

- For a new natural gas-fired combustion turbine, the estimated net revenue requirement is approximately \$80 to \$95 per kW-year. The net revenue in 2014 for a new gas turbine was calculated to be approximately \$37 per kW-year.
- For a new combined cycle unit, the estimated net revenue requirement is approximately \$110 to \$125 per kW-year. The net revenue in 2014 for a new combined cycle unit was calculated to be approximately \$57 per kW-year.
- For a new coal-fired unit, the estimated net revenue requirement is approximately \$265 to \$310 per kW-year. The net revenue in 2014 for a new coal unit was calculated to be approximately \$105 per kW-year.
- For a new nuclear unit, the estimated net revenue requirement is approximately \$450 to \$585 per kW-year. The net revenue in 2014 for a new nuclear unit was calculated to be approximately \$227 per kW-year.

These results indicate that during 2014 the ERCOT markets would not have provided revenues greater than the estimated costs of any of the types of generation technology evaluated.

Therefore, it may seem inconsistent with these results that new generation continues to be added in the ERCOT market. This can be explained by the following factors:

First, the net revenues in any one year may be higher or lower than an investor would require over the long term. In 2014, the net revenues were substantially lower than the estimated cost of entry because shortages were much less frequent than would be expected in the long-term on average. Shortage revenues play a pivotal role in motivating investment in an energy-only market like ERCOT. Hence, in some years the shortage pricing will be frequent and net revenues may substantially exceed the cost of entry, while in most other years it will be less frequent and net revenue will be less than the cost of entry.

Second, the costs of new entry used in this report are generic and reflective of the costs of new resources on a new, 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 equipment costs, possibly though large, long-term supply agreements, or by adding generation to existing sites, or through some combination of both.

Third, in addition to the equipment cost, financing structures and costs can vary greatly between suppliers. Again, the net revenue analysis assumes generic financing costs that a specific supplier may be able to improve on. 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. Suppliers will develop their own view of future expected revenue which may include a power sales contract for some amount of the output. A power sales contract could provide them with more revenue certainly than is available by relying solely on the ERCOT wholesale market. 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.

#### 2. 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 projection of reserve margins developed prior to the summer of 2015.





Source: ERCOT Capacity Demand Reserve Reports / 2015 from December 2014 and 2016-2020 from May 2015

This figure indicates that the region will have a 15.7 percent reserve margin heading into the summer of 2015. Reserve margins are now expected to exceed the target level of 13.75 percent for the next several years. These projections are higher than those developed last year. Further, this outlook is very different than in 2013 when reserve margins were expected to be below the target level of 13.75 percent for the foreseeable future.

In 2013 the expected reserve margin for 2016 was 10.4 percent, much lower than the current expectation for 2016 of 17 percent. This increase in expected reserve margin is not due to an increase in available generating resources, but rather to ERCOT's revised long-term load forecasting methodology and resulting reduction in the forecasted peak demand. The quantity of available resources expected in 2016 as shown in the May 2013 Capacity Demand Report (CDR) is nearly identical to the quantity of resources shown in the May 2015 CDR. Although the total

expected capacity of resources has not changed between the two CDRs, the mix has changed. Almost 1,700 MW of increased wind capacity expected in 2016 has been offset by reductions in the total capacity expected from natural gas and coal.

Looking beyond 2016, several new additions have been announced and meet the requirements for being included in the CDR. The bulk of this new capacity is from new gas units (greater than 5 GW) sited at locations across the ERCOT region. Wind additions also are projected to continue, with 1.5 GW of capacity shown in the CDR representing nearly 10 GW of installed wind capacity. Rounding out the additions is more than 500 MW of solar capacity.

#### 3. 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. To incent generation additions the market design must provide revenues such that the marginal resource receives revenues sufficient to make that resource economic. Generators earn revenues from three sources: energy prices during non-scarcity, energy prices during scarcity, and capacity payments. Generator revenue in ERCOT is overwhelmingly derived from energy prices under both scarcity and non-scarcity conditions.

Expectations for energy pricing under non-scarcity 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 and to maintain existing resources. This will occur when energy prices are allowed to rise substantially at 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 for 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 produce a planning reserve margin that is less than the planning reserve target.

Faced with reduced levels of generation development activity coupled with increasing loads that result in falling planning reserve margins, the PUCT has devoted considerable effort since 2012 deliberating issues related to resource adequacy. To date, the PUCT continues to support the energy-only nature of the ERCOT market and has directed market modifications to improve ERCOT's shortage pricing based on the demand for operating reserves.

The Operating Reserve Demand Curve (ORDC) is a shortage pricing mechanism that reflects the loss of load probability (LOLP) at varying levels of operating reserves multiplied by the value of lost load (VOLL). Selected as an easier to implement alternative to real-time co-optimization of energy and ancillary services, the ORDC provides a new form of shortage pricing for online and offline reserves, as well as energy. As available reserve capacity drops to 2,000 MW, payment for reserve capacity will rise to VOLL, or \$9,000 per MWh.

The initial implementation of ORDC went into effect on June 1, 2014 and included the introduction of real-time reserve on-line and off-line adders. The load-weighted real-time energy price for the period of 2014 after ORDC implementation (i.e. after June 1<sup>st</sup>) was \$35.68 per MWh. Of that total, \$0.26 per MWh (less than 1 percent) was the on-line reserve adder. The on-line reserve adder includes the off-line adder, which was \$0.09 per MWh for this time period.

#### G. 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).

#### 1. Structural Market Power

The Residual Demand Index (RDI) is used as the primary indicator of potential structural market power. The RDI measures the percentage of load that cannot be served without the resources of the largest supplier, assuming that the market could call upon all committed and quick-start capacity owned by other suppliers. When the RDI is greater than zero the largest supplier is pivotal; that is, its resources are needed to satisfy the market demand. When the RDI is less than zero, no single supplier's resources are required to serve the load as long as the resources of its competitors are available.

The RDI is a useful structural indicator of potential market power, although it is important to recognize its limitations. As a structural indicator, it does not illuminate actual supplier behavior to indicate whether a supplier may have exercised market power. The RDI also does not indicate whether it would have been profitable for a pivotal supplier to exercise market power. However, it does identify conditions under which a supplier would have the ability to raise prices significantly by withholding resources.



**Pivotal Supplier Frequency by Load Level** 

The figure above summarizes the results of the RDI analysis by displaying the percentage of time at each load level there was a pivotal supplier. At loads greater than 65 GW there was a pivotal supplier 100 percent of the time. The figure also displays the percentage of time each load level occurs. By combining these values it can be determined that there was a pivotal

supplier in approximately 23 percent of all hours of 2014, which indicates that market power is a potential concern in ERCOT and underscores the need for the current mitigation measures that address it.

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. 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.

#### 2. Evaluation of Conduct

This subsection assesses potential physical withholding and economic withholding using a variety of metrics; starting with an evaluation of potential economic withholding, which is conducted by calculating an "output gap." The output gap is defined as the quantity of energy that is not being produced by in-service capacity even though the in-service capacity is economic by a substantial margin given the real-time energy price. A participant can economically withhold resources, as measured by the output gap, by raising its energy offers so as not to be dispatched.

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

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

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



In addition to this analysis of potential economic withholding, outages, deratings, and economic units that were not committed were also evaluated to identify other means suppliers may have used to withhold resources. Very little evidence of potential physical withholding was found. Based on the analyses described above and the results of our ongoing monitoring, we find the overall performance of the ERCOT market to be competitive in 2014.

#### H. Recommendations

Overall, we find that the ERCOT market performed well in 2014. Nonetheless, we have identified and recommended a number of potential improvements. We describe these recommendations in this section.

Our recommendation to modify the Protocols related to proxy energy offer curve provisions has been addressed in NPRR 662. With this modification, available capacity without an associated energy offer will be priced at the same price as the last megawatt associated with a submitted offer, rather than being priced at the system-wide offer cap. We assert that the more appropriate price to assume for this available, but un-offered capacity is the highest price that the resource has actually submitted. This is particularly true given the recent changes raising the system-wide offer cap to \$9000 per MWh and the implementation of ORDC, under which available capacity will receive a reserve adder payment, whether it has a submitted offer or not.

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

The Operating Reserve Demand Curve (ORDC) provides a mechanism for setting realtime energy prices that reflect the expected value of lost load. However, additional benefits can be achieved by implementing real-time co-optimization of energy and ancillary services. These benefits are twofold. First, jointly optimizing all products in each interval allows the market to substitute its procurements between units on an interval-by-interval basis to minimize costs and set efficient prices. The second benefit, more fully described in Section II.D, Ancillary Services Market at page 38, would be the improved handling of situations when an entity that was selected to provide ancillary services becomes unable to fulfill that commitment, e.g. due to a generator forced outage. For these reasons we continue to recommend ERCOT implement real-time cooptimization of energy and ancillary services.

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

ERCOT has been producing non-binding generation dispatch and price projections for more than two years, but it is unclear what, if any, effect this indicative information has had on the operational actions of ERCOT or market participants. This indicative information has highlighted weaknesses in ERCOT's short term load forecasting process. ERCOT has identified improvements to its forecasting process and once those improvements have been implemented, ERCOT and stakeholders will undertake an evaluation of the benefits of implementing a multi-interval real-time market. We continue to believe there is opportunity to improve the commitment and dispatch of both load and generation resources that require longer than 5 minutes to come on line, but are available within 30 minutes. Therefore, we recommend that ERCOT evaluate improvements to this process that would allow it to facilitate better real-time generator and load commitments.

# 3. <u>Implement changes to ensure all load deployments are reflected in the real-time dispatch</u> energy and reserve prices.

When load is not being served – either because the price is higher than the load's willingness to pay, or the load has been curtailed due to emergency conditions – the energy price should reflect the value to load of not being served. Currently, when load is curtailed, the energy price reflects the cost of supply to serve the reduced amount of load. While Phase 1 of Loads in SCED made some progress in this direction, the implementation of NPRR626 will go further, by introducing a second execution of SCED in situations when loads are deployed. This second execution will determine the higher LMPs that would have occurred if the load had continued to be served. The price increment (reliability adder) will be added to settlement point prices. We will evaluate the effects of NPRR626 implementation in 2015. A further step would be to integrate bids from load resources and emergency resources in the real-time dispatch software and allow them to set prices when they are effectively marginal.

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

In the context of ongoing stakeholder discussions about Future Ancillary Services, we reintroduce 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. Although we are not recommending any changes to the current ancillary services procurement or pricing practices, inefficiencies exist in the current practices for responsive reserves. As the services and requirements for those services are re-defined, we believe it is appropriate to include this change to improve pricing efficiency and supplier incentives.

#### I. REVIEW OF REAL-TIME MARKET OUTCOMES

As is typical in other wholesale electricity markets, only a small share of the power produced in ERCOT is transacted in the spot market. However, prices in the real-time energy market are very important because they set the expectations for prices in the forward markets (including bilateral markets) where most transactions take place. Unless there are barriers preventing arbitrage of the prices between the spot and forward markets, the prices in the forward market should be directly related to the prices in the spot market (i.e., the spot prices and forward prices should converge over the long-run). Hence, artificially low prices in the real-time energy market will translate to artificially-low forward prices. Likewise, price spikes in the real-time energy market will increase prices in the forward markets. This section evaluates and summarizes electricity prices in the real-time market during 2014.

#### 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.

The ERCOT-wide price is the load-weighted average of the real-time market prices from all load zones. Ancillary services costs are estimated based on total system demand and prices in the ERCOT markets for regulation, responsive reserves, and non-spinning reserves. Uplift costs are assigned market-wide on a load-ratio share basis to pay for costs associated with reliability unit commitments and reliability must run contracts. Starting June 1, 2014, with the implementation of the Operating Reserve Demand Curve, the real-time energy price includes the Online Reserve Adder. In the figure below this adder has been separated out from the energy price.



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

Figure 1 shows the monthly average all-in price for all of ERCOT from 2011 to 2014 and the associated natural gas price. This figure indicates that natural gas prices were a primary driver of the trends in electricity prices from 2011 to 2014. Again, this is not surprising given that natural gas is a widely-used fuel for the production of electricity in ERCOT, especially among generating units that most frequently set locational marginal prices in the nodal market.

The all-in price of electricity is equal to the load-weighted average real-time energy price, plus ancillary services, and real-time uplift costs per MWh of real-time load. The largest component of the all-in cost of wholesale electricity is the energy cost. ERCOT average real-time all-in prices were 21 percent higher in 2014 than in 2013. The ERCOT-wide load-weighted average price was \$40.64 per MWh in 2014 compared to \$33.71 per MWh in 2013. The Online Reserve adder was \$0.26 per MWh for the last half of the year.

The increase in real-time energy prices was correlated with much higher fuel prices in 2014. The high correlation of natural gas prices and energy prices shown in the figure is consistent with expectations in a well-functioning market. Fuel costs constitute most of the marginal production costs for generating resources in ERCOT and competitive markets provide incentives for suppliers to submit offers consistent with marginal costs. The average natural gas price in 2014 was \$4.32 per MMBtu, a 17 percent increase compared to \$3.70 per MMBtu in 2013. Gas prices were highest in the first quarter when unusually cold weather throughout the U.S. resulted in much higher and more volatile natural gas prices.

Figure 2 shows the monthly load-weighted average prices in the four geographic ERCOT load zones during the past four years.



Figure 2: Average Real-Time Energy Market Prices

These prices are calculated by weighting the real-time energy price for each interval and each zone by the total zonal 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 contract prices.

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 3 below shows the effect that this reduction has on a monthly basis, by zone. With the CRR Auction Revenue offset included, the ERCOT-wide load-weighted average price was reduced by \$1.10 per MWh to \$39.54 per MWh in 2014.



Figure 3: Effective Real-Time Energy Market Prices

To provide additional perspective on the outcomes in the ERCOT market, the following figure compares the all-in prices in ERCOT with other organized electricity markets in the United States: New York ISO, ISO New England, Pennsylvania-New Jersey-Maryland (PJM) Interconnection, Midcontinent ISO, and California ISO.


Figure 4: 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 for economically out-ofmerit resources. Figure 4 shows that ERCOT all-in prices in 2014 were lower than CaISO and the eastern markets of New York, New England and PJM, and on par with MISO.

Figure 5 below presents price duration curves for ERCOT energy markets in each year from 2011 to 2014. A price duration curve indicates the number of hours (shown on the horizontal axis) that the price is at or above a certain level (shown on the vertical axis). The prices in this figure are the hourly load-weighted nodal settlement point prices.

Price levels during 2014 were similar to those in 2011 for most of the year, with both years having 700 to 800 hours with prices exceeding \$50 per MWh. Prices in 2012 and 2013 exceeded \$50 per MWh much less often. As described later in this section, these lower prices were a result of lower natural gas prices in those two years.



**Figure 5: ERCOT Price Duration Curve** 

To see where the prices during 2014 diverged from the previous three years, a comparison of prices for the highest 5 percent of hours in each year is presented. In 2011, energy prices for the top 100 hours were significantly higher. These higher prices were due to higher loads leading to

more shortage conditions coupled with a more effective shortage pricing mechanism implemented as part of the nodal market design. In 2012, 2013, and 2014, the price duration curves for the top 5 percent of hours are very similar and reflect fewer occasions of shortage conditions.



To better observe the effect of the highest-priced hours, the following analysis focuses on the frequency of price spikes in the real-time energy market. Figure 7 shows the average price and

the number of price spikes in each month. 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.



Figure 7: Average Real-Time Energy Prices and Number of Price Spikes

The number of price spike intervals during 2014 averaged 74 per month, an increase from the average of 54 price spike intervals per month during 2013.

To measure the impact of these price spikes on average price levels, the figure also shows average prices with and without the price spike intervals. The top portions of the stacked bars show the impact of price spikes on monthly average price levels. At \$14.09 per MWh, the impact of price spikes was the greatest in 2011. In 2012 the frequency of price spikes increased but the magnitude of their price impact decreased to \$3.63 per MWh. The magnitude decreased again in 2013 to \$3.43 per MWh. The magnitude increased in 2014, with an impact on the average energy price of \$5.28 per MWh. Of this price spike impact, \$0.20 was due to the effects of the Operating Reserve Demand Curve (ORDC) adder.

To depict how real-time energy prices vary by hour in each zone, Figure 8 shows the hourly average price duration curve in 2014 for the four ERCOT load zones.





The Houston, North and South load zones had similar prices over the majority of hours. The price duration curve for the West zone is noticeably different than the other zones, with more hours when prices exceeded \$50 per MWh. The number of hours with prices less than \$0 per MWh was very similar for all zones in 2014. This is notable since for the past several years the West zone has consistently had much more frequent occurrences of negative prices than the other zones. Significant transmission additions have lowered the frequency of depressed West zone prices due to transmission congestion during times of high wind output. However, the trend of local transmission constraints during low wind and high load conditions has continued and causes West prices to be higher than the rest of ERCOT. As discussed above in Figure 3, these higher prices are largely offset by the CRR Auction Revenues allocated to QSEs representing load.

Figure 9 shows the relationship between West zone and ERCOT average prices for 2011 through 2014.



Figure 9: West Zone and ERCOT Price Duration Curves

On the low price end, the near elimination in the number of hours when West zone prices were below the ERCOT average can be observed. Note that the minimum West zone prices have increased; that is, become "less negative." West zone prices were noticeably higher than the ERCOT average for a significant number of hours in 2014, although not to the same magnitude as they were in 2013 (which was itself a reduction from 2012). But like 2013 and 2012, the combination of more hours with higher prices, and fewer hours with less negative prices resulted in the average real-time energy price in the West zone being greater than the ERCOT average. As noted previously, however, the offset provided by CRR Auction Revenue actually brings the effective average real-time energy price in the West zone lower than the ERCOT average.

More details about the transmission constraints influencing energy prices in the West zone are provided in Section III. Transmission and Congestion.

#### B. Real-Time Prices Adjusted for Fuel Price Changes

Although real-time electricity prices are driven to a large extent by changes in fuel prices, natural gas prices in particular, they are also influenced by other factors. To clearly identify changes in electricity prices that are not driven by changes in natural gas prices, Figure 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.<sup>1</sup>



Implied heat rates in 2012 were noticeably higher for the majority of hours, as compared to the other three years. This can be explained by the very low natural gas prices experienced in 2012, and resulting pricing outcomes which were influenced by coal, not natural gas, being the

<sup>&</sup>lt;sup>1</sup> The *Implied Marginal Heat Rate* equals the *Real-Time Energy Price* divided by the *Natural Gas Price*. This methodology implicitly assumes that electricity prices move in direct proportion to changes in natural gas prices.

marginal fuel.<sup>2</sup> For most hours, there are no discernable differences between 2011, 2013, and 2014.

Taking a closer look at the implied marginal heat rates for the top five percent of hours for years 2011 through 2014 in Figure 11 also shows that the implied heat rates in 2012, 2013, and 2014 are also very similar; 2011 remains an outlier.



# Figure 11: Implied Marginal Heat Rate Duration Curve – Top Five 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 in 2013 and 2014, with annual average heat rate data for 2011 through 2014. This figure is the fuel price-adjusted version of Figure 2 in the prior subsection. Adjusting for natural gas price influence, Figure 12 shows that the annual, system-wide average implied heat rate increased in 2014 compared to 2013.

<sup>&</sup>lt;sup>2</sup> See the 2012 ERCOT SOM report at pages 12-13.



Figure 12: Monthly Average Implied Heat Rates

The monthly average implied heat rates in 2014 are generally higher than those in 2013 through May, after which they drop below the 2013 heat rates. This trend is generally consistent with rising gas prices and higher loads in early 2014 compared to the same months of 2013.

The examination of implied heat rates from the real-time energy market concluded by evaluating them at various load levels. Figure 13 below provides the average heat rate at various system load levels from 2011 through 2014.



Figure 13: 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. Although a generally positive relationship exists, there is a noticeable disparity for loads between 50 and 55 GW. During the extreme cold weather event in early February 2011, loads were at this level while prices reached \$3,000 per MWh for a sustained period of time. The higher heat rates observed at lower loads in 2012 are likely due to the displacement of generation from coal units by generation from natural gas units when low natural gas prices were experienced during that year.<sup>3</sup>

There are two noticeable differences in 2014 relative to the other years. The first is the higher implied marginal heat rate at load levels between 55 and 60 GW. This is due to scarcity pricing that occurred when load was in that range during January. The second is the lower implied

<sup>&</sup>lt;sup>3</sup> For additional explanation see the 2012 ERCOT SOM report at pages 12-13.

marginal heat rate at load levels between 60 and 65 GW. This is due to the relative lack of scarcity pricing at those load levels during the summer.

#### C. **Aggregated Offer Curves**

The next analysis compares the quantity and price of generation offered in 2014 to that of 2013. 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. Comparing 2014 to 2013, more capacity was offered at lower prices. Specifically, there was approximately 900 MW of additional capacity offered at prices less than zero. This was split between offers from wind generators (400 MW) and capacity below generators' low operating limits (500 MW). There was approximately 1,200 MW of additional capacity offered in 2014 at prices between zero and ten multiplied by the daily natural gas price. The amount of capacity offered at prices between 10 multiplied by the daily natural gas price and \$250 per MWh was similar in the both years. With smaller changes to the quantities of generation offered at prices above \$250 per MWh, the resulting average aggregated generation offer stack was roughly 2,100 MW greater in 2014 than in 2013.





The next analysis provides a similar comparison for only the summer season. As shown below in Figure 15, the changes in the aggregated offer stacks between the summer of 2013 and 2014 were similar to those just described. Comparing 2013 to 2014, there were approximately 2,300 MW additional capacity offered at prices less than 10 multiplied times the daily natural gas price; 600 MW at prices less than zero; and 1,700 MW at prices greater than zero. There was approximately 1,000 MW less capacity offered at prices between 10 multiplied by the daily natural gas price and \$250 per MWh. With smaller reductions to the quantities of generation offered at prices above \$250 per MWh, the resulting average aggregated generation offer stack for the summer season was approximately 1,000 MW greater than in 2013.



# D. Prices at the System-Wide Offer Cap

Revisions to 16 TEX. ADMIN. CODE § 25.505 raised the system-wide offer cap to \$5,000 per MWh effective June 1, 2013; \$7,000 per MWh effective June 1, 2014; and \$9,000 per MWh effective June 1, 2015. As more fully described later in Section V. Resource Adequacy,

independent of the energy offers by generators, energy prices rise toward the system-wide offer cap as available operating reserves approach minimum required levels to reflect the degradation in system reliability. Given the ERCOT market's reliance on these high real-time prices, Figure 16 below shows the aggregate amount of time when the real-time energy price was at the system-wide offer cap, displayed by month.





There were no instances in 2014 of energy prices rising to the cap after the system-wide offer cap was increased to \$7,000 per MWh on June 1. Prices during 2014 were at the system-wide offer cap for only 1.56 hours, an increase from 0.22 hours in 2013 and a slight increase from the 1.51 hours experienced in 2012. All years were much lower than the 28.44 hours at the cap experienced in 2011 and the average amount expected over the long term in an energy-only market.

The next figure provides a detailed comparison of each August's load, required reserve levels, and prices for 2011 through 2014. There were very few dispatch intervals when real-time energy

prices reached the system-wide offer cap in 2012, 2013, and 2014 compared to the relatively high frequency it occurred in 2011. Although the weather may have been similar, there were significant differences in load and available operating reserve levels, resulting in much higher prices in August 2011.



Figure 17: Load, Reserves and Prices in August

Shown on the left side of Figure 17 is the relationship between real-time energy price and load level for each dispatch interval for the months of August 2011, 2012, 2013, and 2014. ERCOT loads were greater than 65 GW for 71 hours in 2011, whereas loads reached that level for 12 hours during August 2012, 18 hours in August 2013 and 11 hours in August 2014. As previously discussed, a strong positive correlation between higher load and higher prices is expected in a well-functioning energy market. Such a relationship between higher prices and higher load is observed in this analysis. However, that relationship appears to be weaker in the past three years with more instances of higher prices occurring at lower loads.

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

On the right side of Figure 17 are data showing the relationship between real-time energy prices and the quantity of available operating reserves for each dispatch interval during August of 2011, 2012, 2013, and 2014. This figure shows a strong correlation between diminishing operating reserves and rising prices. With the lower loads in August 2012, 2013, and 2014, available operating reserves were well above minimum levels for the entire month, and there were no occurrences when the energy price reached the system-wide offer cap. In contrast, there were numerous dispatch intervals in August 2011 when the minimum operating reserve level was approached or breached, with 17.4 hours where prices reached the system-wide offer cap. It should be noted that during August 2011 there were a number of dispatch intervals when operating reserves were below minimum requirements with prices well below the level that would be reflective of the reduced state of reliability. In Section IV.C., Demand Response Capability, at page 79, an example is provided explaining why this can occur and a recommendation for improvement is offered.

#### 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 18 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 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 2013 are also presented. Comparing average real-time energy prices for 2014 with those from 2013, the effects of higher natural gas prices on average prices may be observed.



Figure 18: Real-Time Energy Price Volatility (May – August)

Outside of the hours from 15 to 18, short-term increases in average real-time energy prices are typically due to high prices resulting from generator ramp rate limitations occurring at times when significant amounts of generation are changing online status. The price effects of these ramp-limited periods were similar in 2013 and 2014. The average of the absolute value of changes in five-minute real-time energy prices, expressed as a percentage of average price, was 3.0 percent in 2014, compared to 3.4 percent in 2013, 3.6 percent in 2012 and approximately 6.2 percent for the same period in 2011. This steady decline may be attributed to a decrease in

shadow price cap intervals from 2011-2012 and the decrease in West zone price volatility from 2012 to 2014.

Expanding the view of price volatility, Figure 19 below presents the monthly variation in realtime prices. The highest price variability occurs during months when real-time prices rose to the system wide offer cap.



**Figure 19: Monthly Price Variation** 

The volatility of 15-minute settlement point prices for the four geographic load zones in 2014 was similar to that seen in 2013 and 2012, as shown below in Table 1.

Load Zone	2012	2013	2014
Houston	13.0%	14.8%	14.7%
South	13.1	15.4	15.2
North	13.9	13.7	14.1
West	19.4	17.2	15.4

Price volatility in the Houston, North, and South zones was similar in 2014 to the levels experienced in 2013. The table also shows that price volatility in the West zone has decreased, likely as a result of transmission investment in the region. Price volatility in the West zone in 2014 was similar to the other zones.

#### F. Mitigation

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 only considering transmission constraints that have been deemed competitive. These "reference prices" at each generator location are compared with that generator's mitigated offer cap, and the higher of the two is used to formulate the offer curve to be used for that generator in the second step in the dispatch process. The resulting mitigated offer curve is used by the dispatch software to determine the final output levels for each generator, taking all transmission constraints into consideration. This approach is intended to limit the ability of a specific generator to raise prices in the event of a transmission constraint that requires its output to resolve. In this subsection the quantity of capacity being mitigated in 2014 during this mitigation process is analyzed. Although executing all the time, the automatic price mitigation aspect of the two-step dispatch process only has an effect when a non-competitive transmission constraint is active. The mitigation process should limit the ability of a generator to affect price when its output is required to manage congestion. The process as initially implemented did not identify situations with sufficient competition between generators on the other (harmful) side of the constraint and would mitigate those offers as well. This unnecessary mitigation was addressed on June 12, 2013 with the implementation of NPRR520. With the introduction of an impact test to determine whether units are relieving or contributing to a transmission constraint, only the relieving units are now subject to mitigation. As shown below this had a noticeable effect on the amount of capacity subject to mitigation.

The analysis shown in Figure 20 computes how much capacity, on average, is actually mitigated during each dispatch interval. The results are provided by load level.



Figure 20: Mitigated Capacity by Load Level

The level of mitigation in 2014 was much lower than in 2012 and 2013, even after the mitigation rule changed in mid-2013. The amount of mitigated capacity averaged just over 60 MW at its highest during 2014. At similar load levels in 2013 (after the change) the mitigated amounts were well over 100 MW and exceeded 150 MW at the very highest load levels. One explanation for this reduction is that the higher gas prices in 2014 led to an increase in the mitigated offer cap and a corresponding reduction in the amount of offers over the mitigated offer cap. A second explanation is the decrease in the number of intervals during 2014 when congestion was present. As described later in Section III. Transmission and Congestion, congestion occurred in only 44 percent of intervals in 2014 compared to 55 percent in 2013.

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, as shown in Figure 21.



Figure 21: Capacity Subject to Mitigation

The effects of the rule change are very noticeable in Figure 21. Compared to 2012 when the amount of capacity subject to mitigation exceeded 1500 MW for all load levels, the amount of capacity subject to mitigation after the rule change in 2013 was always below 700 MW. In 2014, the largest amount being mitigated was lower than 350 MW. Put another way, up to 7 percent of capacity required to serve load in 2012 was subject to mitigation. After the rule change this percentage decreased to less than 1 percent. An important note about this capacity measure is that it 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.

#### G. Revenue Sufficiency

In Figure 22 the combined payments to PTP Obligation owners and effective payments to other day-ahead positions are compared to the total real-time congestion rent. For the year of 2014,

real-time congestion rent was \$692.5 million, payments for PTP Obligations (including those with links to CRR Options) were \$524.5 million and effective payments for other day-ahead positions were \$211.3 million, resulting in a shortfall of approximately \$43 million for the year. This shortfall is effectively paid by all loads, allocated on a load ratio share.





For the year of 2013, real-time congestion rent was \$481 million, payments for PTP Obligations and real-time CRRs were \$352 million and effective payments for other day-ahead positions were \$167 million, resulting in a shortfall of approximately \$37 million for the year. This type of shortfall typically results from discrepancies between transmission topology assumptions used when clearing the day-ahead market and the actual transmission topology that exists during realtime. Specifically, if the day-ahead topology assumptions allow too many real-time congestion instruments (PTP Obligations and CRRs settled at real-time prices) with impacts on a certain path, and that path constrains at a lower limit in real-time, there will be insufficient rent generated to pay all of the congestion instruments and the balance will be uplifted to load.

#### II. REVIEW OF DAY-AHEAD MARKET OUTCOMES

ERCOT's centralized 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 power, the day-ahead market also includes ancillary services and 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 clearing. These transactions are made for a variety of reasons, including satisfying the participant's own supply, managing risk by hedging the participant's exposure to the real-time market, or arbitraging with the real-time markets. For example, load serving entities can insure against volatility in the real-time market by purchasing in the day-ahead market. Finally, the day-ahead market plays a critical role in coordinating generator commitments. For all of these reasons, the performance of the day-ahead market is essential.

In this section 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 two main conditions are in place: a) there are low barriers to shifting purchases and sales between the forward and real-time markets; and b) sufficient information is available to market participants to allow them to develop accurate expectations of future real-time prices. When these conditions are met, market participants can be expected to arbitrage predictable differences between

forward prices and real-time spot prices by increasing net purchases in the lower-priced market and increasing net sales in the higher-priced market. These actions will tend to improve the convergence of forward and real-time prices.

In this subsection, price convergence between the day-ahead and real-time markets is evaluated. This analysis reveals whether persistent and predictable differences exist between day-ahead and real-time prices, which participants should arbitrage over the long-term. To measure the short-term deviations between real-time and day-ahead prices, 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. For instance, if day-ahead prices are \$70 per MWh on two consecutive days while real-time prices are \$40 per MWh and \$100 per MWh on the two days, the absolute price difference between the day-ahead market and the real-time market would be \$30 per MWh on both days, while the difference in average prices would be \$0 per MWh.

Figure 23 shows the price convergence between the day-ahead and real-time market, summarized by month. Day-ahead prices averaged \$40 per MWh in 2014 compared to an average of \$38 per MWh for real-time prices.<sup>4</sup> The average absolute difference between day-ahead and real-time prices was \$12.87 per MWh in 2014; higher than in 2013 when the average of the absolute difference was \$9.86 per MWh. 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 and higher for generators selling day-ahead. The higher risk for generators is associated with the potential of incurring a forced outage and, as a result, having to buy back energy at real-time prices. This explains why the highest premiums tend to occur during the months with the highest relative demand and highest prices.

The overall day-ahead premium increased in 2014 compared to 2013, as a result of the much higher premiums in January through March. Although peak loads during the winter are somewhat lower than those in the summer, loads during the first months of 2014 set record highs for that time of the year. Day-ahead premiums in ERCOT remain higher than observed in other

<sup>&</sup>lt;sup>4</sup> These values are simple averages, rather than load-weighted averages as presented in Figures 1 and 2.

organized electricity markets. 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 2014, it should not be expected over time that every month will always produce a day-ahead premium as 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 January and March).





In Figure 24 below, monthly day-ahead and real-time prices are shown for each of the geographic load zones. Of note is the difference in the West zone data compared to the other regions. The higher volatility in West zone pricing is likely associated with the uncertainty of forecasting wind generation output and the resulting price differences between day-ahead and real-time.



Figure 24: Day-Ahead and Real-Time Prices by Zone

# B. Day-Ahead Market Volumes

The next analysis summarizes the volume of day-ahead market activity by month. Figure 25 below shows that the volume of day-ahead purchases provided through a combination of generator-specific and virtual energy offers was approximately 50 percent of real-time load in 2014, which is similar to 2013 activity.

As discussed in more detail in the next subsection, PTP Obligations are financial instruments purchased in the day-ahead market. Although these instruments do not themselves involve the direct supply of energy, they do provide the ability to avoid the congestion costs associated with transferring the delivery of energy from one location to another. 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 2014 was almost 5 percent higher than in 2013.

Adding the aggregated transfer capacity associated with purchases of PTP Obligations to the other injections and withdrawals demonstrates that net system flow volume transacted in the day-ahead market is greater than real-time load by an average of 14 percent. The volume in excess of real-time load increased in 2014 compared to 2013, when on average the monthly net system flow volume was 12 percent greater than real-time load.



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

Figure 26 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 6 and 22. Since these times align with common bilateral transaction terms, it appears that market participants are using the day-ahead market to trade around those positions.





# C. Point to Point Obligations

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



Figure 27: Point to Point Obligation Volume

Figure 27 presents the total volume of PTP Obligation purchases divided into three categories. Compared to the previous two figures which showed net flows associated with PTP Obligations, in this figure the total volume is presented. For all PTP Obligations that source at a generator location, the capacity up to the actual generator output is considered a generator hedge. 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 profit from 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 for the day-ahead price, the owner will have made a profit. Conversely, if the price difference does not materialize in real-time, the PTP Obligation may be considered

unprofitable. The profitability of PTP Obligation holdings by the two types of participants are compared in Figure 28.



Figure 28: Average Profitability of Point to Point Obligations

This analysis shows that in aggregate the PTP Obligation holdings of both physical and financial participants were profitable. However, both financial and physical participants had four months where PTP Obligation holdings were unprofitable. It may be inferred from the data shown in Figure 28 and in Figure 29 that PTP Obligation holdings, in aggregate, were more profitable in 2014 than they were in 2013.

To conclude the analysis of PTP Obligations, Figure 29 compares the total amount paid for these instruments day-ahead, with the total amount received by holders in real-time.



#### Figure 29: Point to Point Obligation Charges and Payments

As in prior years, with the exception of 2013, the aggregated total payments received by PTP Obligation owners was greater than the amount charged to the owners to acquire them. This occurs when real-time congestion is greater than what occurred in the day-ahead. Across the year, and in eight of twelve months, the acquisition charges were less than the payments received, implying that expectations of congestion as evidenced by day-ahead purchases were less than the actual congestion that occurred 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 and Congestion.

# D. Ancillary Services Market

The primary ancillary services are up regulation, down regulation, responsive reserves, and nonspinning reserves. Market participants may self-schedule ancillary services or purchase their required ancillary services through the ERCOT markets. This subsection reviews the results of the ancillary services markets in 2014, starting with a display of the quantities of each ancillary service procured each month shown in Figure 30.



### Figure 30: Ancillary Service Capacity

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.

The amount of responsive reserve was increased by 500 MW beginning in April 2012. This 500 MW increase was balanced with the same amount of decrease in the amount of non-spinning reserves procured. Although the minimum level of required responsive reserve remains at 2,300 MW, having the additional 500 MW of responsive reserve provides a higher quality – that is, faster responding capacity available to react to sudden changes in system conditions.

Under the nodal market, ancillary service offers are co-optimized as part of the day-ahead market clearing. This means that market participants do not have to include their expectations of forgone energy sales in their ancillary services capacity offers. As a result of ancillary services clearing prices explicitly accounting for the value of energy in the day-ahead market, ancillary services prices are highly correlated with day-ahead energy prices and, by extension, with real-time energy prices.



Figure 31 above presents the average clearing prices of capacity for the four ancillary services, along with day-ahead and real-time energy prices for energy. With average energy prices varying between \$25 and \$60 per MWh, the prices of ancillary services remained fairly stable throughout the year.

In contrast to the previous data that showed the individual ancillary service capacity prices, Figure 32 shows the monthly total ancillary service costs per MWh of ERCOT load and the average real-time energy price for 2012 through 2014. This figure shows that total ancillary service costs are generally correlated with day-ahead and real-time energy price movements, which, as previously discussed, are highly correlated with natural gas price movements. This occurs for two primary reasons. First, higher energy prices increase the opportunity costs of providing reserves and, therefore, can contribute to higher ancillary service prices. Second, shortages cause both ancillary service prices and energy prices to rise sharply so increases or decreases in the frequency of shortages will contribute to this observed correlation.





The average ancillary service cost per MWh of load increased to \$1.51 per MWh in 2014 compared to \$1.03 per MWh in 2013, an increase of 47 percent. Total ancillary service costs increased from 3.0 percent of the load-weighted average energy price in 2013 to 3.7 percent in 2014.

Responsive reserve service is the largest quantity and typically the highest priced ancillary service product. Figure 33 below shows the share of the 2014 annual responsive reserve requirements, including both load and generation provided by each QSE. During 2014, 37 different QSEs provided responsive reserves at some point, with multiple QSEs providing sizable shares.



Figure 33: Responsive Reserve Providers

In contrast, Figure 34 below shows that the provision of non-spinning reserves is highly concentrated, with a single QSE providing 58 percent of the total amount of non-spinning reserves procured last year. We are not raising concerns with the competitiveness of the provision of this service during 2014; however, the fact that one party is consistently providing

the preponderance of this service should be considered in the ongoing efforts to redefine the definition and required quantities of ERCOT ancillary services. Further, it 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 allows the market to substitute its procurements between units on an interval-by-interval basis to minimize costs and set efficient prices. Additionally, it would allow higher quality reserves (e.g., responsive reserves) to be substituted for lower quality reserves (e.g., non-spin reserves), reducing the reliance upon a single entity to provide this type of lower quality reserves.





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, events can occur which make this capacity unavailable to the system. Transmission constraints can arise which make the capacity undeliverable. Outages or limitations at a generating unit can lead to failures to provide. When

either of these situations occurs, ERCOT may open a supplemental ancillary services market (SASM) to procure replacement capacity.<sup>5</sup>

Figure 35 below, presents a summary of the frequency with which ancillary service capacity was not able to be provided and the number of times that a SASM was opened in each month. The percent of time that capacity procured in the day-ahead actually provided the service in the hour it was procured for increased to 57 percent in 2014, compared to 39 percent in 2013 and 52 percent in 2012. Even though there were deficiencies in ancillary service deliveries for more than 40 percent of the hours, SASMs were opened to procure replacement capacity in only 2 percent of the total hours, down from 3 percent of the hours in 2013 and 7 percent in 2012.



Figure 35: Frequency of SASM Clearing

The primary reason that SASMs were infrequent was the dearth of ancillary service offers typically available throughout the operating day, limiting the opportunity to replace ancillary

<sup>&</sup>lt;sup>5</sup> ERCOT may also open a SASM if they change their ancillary service plan. This did not occur during 2014.

service deficiencies via a market mechanism. Without sufficient ancillary service offers available, ERCOT more frequently brings additional capacity online using reliability unit commitment (RUC) procedures.

The frequency and quantity of ancillary service deficiency, which is defined as either failure-toprovide or as undeliverable, is summarized in Table 2 below.

		Mean	Median	
Service		Deficiency	Deficiency	
	Hours Deficient	(MW)	(MW)	
2014				
<b>Responsive Reserve</b>	2929	46	20	
Non-Spin Reserve	723	48	40	
Up Regulation	686	40	20	
Down Regulation	850	34	15	
2013				
<b>Responsive Reserve</b>	3138	43	20	
Non-Spin Reserve	610	50	38	
Up Regulation	689	38	20	
<b>Down Regulation</b>	575	39	15	
2012				
<b>Responsive Reserve</b>	3756	34	15	
Non-Spin Reserve	664	36	8	
Up Regulation	750	41	25	
Down Regulation	522	48	39	
2011				
<b>Responsive Reserve</b>	4053	39	20	
Non-Spin Reserve	1254	90	39	
Up Regulation	1222	27	20	
Down Regulation	1235	22	11	

# Table 2: Ancillary Service Deficiency

The number of hours with deficiencies in responsive reserve and up regulation services decreased by 7 percent and 0.4 percent respectively in 2014 when compared to 2013. Down regulation and non-spin reserve had about 19 percent and 48 percent increases in the number of hours of deficiency in 2014. Again during 2014, responsive reserve service was deficient most
frequently. As in 2013, well over 90 percent of the deficiency occurrences were caused by failure to provide by the resource rather than undeliverability related to a transmission constraint. The change in the average magnitude of deficiency was mixed, with responsive reserve and the regulation services increasing and the non-spin reserve decreasing slightly.

The SASM procurement method, while offer based, is inefficient and problematic. Because ancillary services are not co-optimized with energy in the SASM, potential participants are required to estimate their opportunity cost rather than have the auction engine calculate it directly, which leads to resources that underestimate their opportunity costs being inefficiently preferred over resources that overestimate their opportunity costs. 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 scarcity of SASM offers. The paucity of SASM offers frequently leaves ERCOT with two choices; (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 their opportunity costs, and 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 likely substantially 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.



Figure 36: Ancillary Service Quantities Procured in SASM

The final analysis in this section, shown in Figure 36, summarizes the average quantity of each service that was procured via SASM. As previously discussed, SASM was rarely used to replace deficiencies in ancillary services in 2014. When a SASM was used in 2014, the quantity of ancillary services procured was less than that seen in 2013. An exception to that trend was responsive reserves, which were procured less frequently, but in larger quantity.

#### III. TRANSMISSION AND CONGESTION

One of the most important functions of any electricity market is to manage the flows of power over the transmission network by limiting additional power flows over transmission facilities when they reach their operating limits. The action taken to ensure operating limits are not violated is called congestion management. The effect of congestion management is to change the output level of one or more generators so as to reduce the amount of electricity flowing on the transmission line nearing its operating limit. Because the transmission system is operated such that it can withstand the unexpected outage of any element at any time, congestion management actions most often occur when a transmission element is expected to be overloaded if a particular unexpected outage (contingency) were to occur. Congestion leads to higher costs due to lower cost generation being reduced and higher priced generator(s) will vary their output is based on the generator's energy offer curve and its relative shift factors to the contingency and constraint pair. This leads to a dispatch of the most efficient resources available to reliably serve demand.

This section of the report summarizes congestion activity in 2014, 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 congestion rights market.

#### A. Summary of Congestion

The total congestion revenue generated by the ERCOT real-time market in 2014 was \$708 million. As discussed further below, the amount of time where transmission constraints had an effect on prices actually decreased in 2014. Congestion in the West zone remained about the same as it was in 2013. Completion of the CREZ transmission projects has eliminated the longstanding limitations on the export of power from the West. However, limitations on the ability to transfer power to the West, particularly under high load and low wind conditions continue. Constraints associated with oil and gas activity in the Eagle Ford Shale area and limitations serving the lower Rio Grande valley had a larger impact in 2014.

Figure 37 provides a comparison of the amount of time transmission constraints were binding or active at various load levels in 2012 through 2014. Binding transmission constraints are those for

which the dispatch levels of generating resources are being altered in order to maintain transmission flows at reliable levels. The costs associated with this re-dispatch are the system's congestion costs and are priced in its LMPs. Active transmission constraints are those that did not require a re-dispatch of generation.





The frequency of binding transmission constraints decreased again in 2014. There was a binding constraint only 44 percent of time in 2014, down from 55 percent of the time in 2013 and 60 percent in 2012. The likelihood of binding constraints increases at higher load levels, which is consistent with the results for ERCOT shown in Figure 37. However, it is noteworthy that there were binding constraints less than 90 percent of the time during the very highest load levels in 2014, much lower than in prior years. Because the overall frequency of binding constraints decreased in 2014, the effect on LMPs also decreased in 2014.

Figure 38 displays the amount of real-time congestion costs attributed to each geographic zone. Those costs associated with constraints that cross zonal boundaries, i.e. North to Houston, are shown in the ERCOT category. Higher costs associated with constraints in the South and Houston zones are the key drivers leading to increased total congestion costs in 2014.



Figure 38: Real-Time Congestion Costs

### **B.** Real-Time Constraints

The review of real-time congestion begins with describing the congested areas with the highest financial impact as measured by congestion rent. For this discussion a congested area is determined by consolidating multiple real-time transmission constraints that are defined as similar due to their geographical proximity and constraint direction. There were 350 unique constraints that were binding at some point during 2014. The median financial impact of these constraints was approximately \$316,000.

Figure 39 below displays the ten most highly valued real-time congested areas as measured by congestion rent. The Heights TNP 138/69 kV autotransformers were the most congested location in 2014 at \$74 million. These transformers are located in the industrialized area southeast of Houston and were the limiting element due to outages of other transmission facilities in the area. As shown in Table 3 below, two constraints with the Heights autotransformer as the overloaded element were designated as irresolvable constraints during 2014.



Figure 39: Top Ten Real-Time Constraints

The second highest valued congested element was the Lytton Springs 345/138 kV #1 autotransformer with impacts of \$56 million. All of the impacts occurred during the early part of the year, January through March, and were related to a planned outage that occurred during times of higher than normal loads.

Congestion in the Midland area totaled \$51 million. This congestion was most prevalent during the months of July and August. Although multiple constraints appear to be binding at different times during the year, the different names were a result of topology changes related to Sharyland

load integrating into ERCOT. The Midland East to Buffalo 138 kV line (\$37 million) was the largest contributor to congestion in the Midland area.

Taken in aggregate, congestion in the Valley area was almost as large as the most costly single constraint, totaling \$73 million. Three of the top congested elements are related to limitations on the ability to import power to serve load in the Rio Grande Valley. They are Harlingen Switch to Oleander 138 kV line (\$29 million) and Rio Hondo to East Rio Hondo 138 kV line (\$25 million) and the Valley Import (\$13 million). These constraints were often in effect during the time that other transmission facilities in the area were taken out of service to accommodate the construction of transmission upgrades in the area. A large contribution to total cost occurred in October, when a combination of generation and transmission outages led to significant congestion. Ultimately, on October 8th the situation in the Valley required that firm load be curtailed.

The next two most costly constraints were the Odessa North 138/69 kV Autotransformer (\$24 million) and the Moss Switch to Westover 138 kV line (\$20 million). The Odessa area, which consists of the two aforementioned constraints among others, remains one of the most highly congested areas due to oil and gas development activity in the area. However, the \$63 million impacts in 2014 were a significant reduction compared to what was experienced in 2013.

Congestion on the Lon Hill to Smith 69kV line west of Corpus Christi totaled \$19 million and was one of three 69kV lines in the area that were congested due to the increased loads due to oil and natural gas development in the Eagle Ford shale. The total for all three lines was \$33 million.

The last element on the list, the Hockley to Betka 138 kV line, is located north and west of Houston and is affected by North to Houston transfers.

## Irresolvable Constraints

When a constraint becomes irresolvable, ERCOT's dispatch software can find no combination of generators to dispatch in a manner such that the flows on the transmission line(s) of concern are below where needed to operate reliably. In these situations, offers from generators are not setting locational prices. Prices are set based on predefined rules, which since there are no

supply options for clearing, should reflect the value of reduced reliability for demand. To address the situation more generally, a regional peaker net margin mechanism was introduced such that once local price increases accumulate to a predefined threshold due to an irresolvable constraint; the shadow price of that constraint would drop.

As shown below in Table 3, seventeen constraints, each comprised of a contingency and overloaded element, were deemed irresolvable in 2014 and as such, had a maximum shadow price cap imposed according to the Irresolvable Constraint Methodology. Many of the irresolvable constraints were the previously discussed most costly constraints. Five constraints were deemed resolvable during the ERCOT analysis annual review and were removed from the list.

Loss of:	Overloads:	Original Max Shadow Price	Adjusted Max Shadow Price	Effective Date	Termination Date
Base case	Valley Import	\$5,000	\$2,000.00	1/1/12	-
Denton to Argyle / West Denton 138 kV lines	Jim Crystal to West Denton 69 kV line	\$2,800	\$2,000.00	1/1/12	1/29/14
Graham to Long Creek 345 kV line	Bomarton to Seymour 69 kV line	\$2,800	\$2,000.00	1/1/12	1/29/14
Odessa North to Holt 69 kV Line	Odessa Basin to Odessa North 69 kV line	\$2,800	\$2,800.00	1/1/12	-
Odessa to Morgan Creek/Quail Sw 345 kV lines	China Grove to Bluff Creek 138 kV line	\$3,500	\$2,000.00	5/3/12	-
Holt to Moss 138 kV line	Odessa North 138/69 kV transformer	\$3,500	\$2,000.00	8/6/12	1/29/14
Sun Switch to Morgan Creek 138 kV Line	China Grove to Bluff Creek 138 kV Line	\$3,500	\$2,000.00	10/11/12	-
Morgan Creek Autotransformer #4 345/148 kV	Morgan Creek Autotransformer #1 345/138 kV	\$4,500	\$2,000.00	11/2/12	-
Odessa Basin to Odessa North 69 kV line	Holt to Ector Shell Tap 69 kV line	\$2,800	\$2,320.68	1/1/13	1/29/14
Wink TNP 138 kV/69 kV Autotransformer	Wink TNP to Wink Sub 69 kV line	\$2,800	\$2,000.00	5/20/13	1/29/14
Skywest to Salt Flat Road 138 kV Line	Midland East to Buffalo 138 kV Line	\$3,500	\$2,377.57	7/24/14	-
Tejas to Greenbelt 138 kV Line	Heights TNP 138/69 kV Transformer	\$3,500	\$2,000.00	9/23/14	-
Estokta to McElmurray and Abilene South to Moore 138 kV Lines	Abilene Northwest to Ely Rea Tap 69 kV Line	\$2,800	\$2,131.06	9/26/14	-

### **Table 3: Irresolvable Constraints**

Loss of:	Overloads:	Original Max Shadow Price	Adjusted Max Shadow Price	Effective Date	Termination Date
Las Palma to Rio Hondo 138 kV Line	Harlingen to Oleander 69 kV Line	\$2,800	\$2,000.00	10/9/14	-
Las Palma to Rio Hondo 138 kV Line	Rio Hondo to East Rio Hondo 138 kV Line	\$3,500	\$2,000.00	10/10/14	-
Bakke to Unocal Parker 138 kV Line	Emma to Holt Switch 69 kV Line	\$2,800	\$2,800.00	10/27/14	-
Caddo to Apache 138 kV Line	Heights TNP 138/69 kV Transformer	\$3,500	\$2,000.00	10/28/14	-

Figure 40 presents a slightly different set of real-time congested areas. These are the most frequently occurring.



## Figure 40: Most Frequent Real-Time Constraints

Of the ten most frequently occurring constraints, four have already been described as the most costly. They are the Lon Hill to Smith 69kV line, the Odessa North 138/69 kV Autotransformer, the Heights TNP 138/69 kV autotransformers, and the Midland East to Buffalo 138 kV line. The Bruni 138/69 kV transformer constraint frequently limits the output from two wind generators located east of Laredo. The Shannon to Post Oak Switch 69 kV line is located between Fort Worth and Wichita Falls. The Gila to Highway 138 kV line is located in the

Corpus Christi area. The Jewett to Singleton and Singleton to Zenith 345kV lines provide part of the current Houston import capability.

## C. Day-Ahead Constraints

This subsection provides a review of the transmission constraints from the day-ahead market. To the extent the model of the transmission system used for the day-ahead market matches the real-time transmission system, and assuming market participants transact in the day-ahead market similarly to how they transact in real-time, the same transmission constraints are expected to appear in the day-ahead market as actually occurred during real-time.



Figure 41: Top Ten Day-Ahead Congested Areas

Figure 41 presents the top ten congested areas from the day-ahead market, ranked by the financial impact as measured by congestion rent. There is a close correlation between constraints with high day-ahead impacts and those previously described in the real-time subsection. The only two constraints that appear here that have not already been discussed are the Munday AEP to Bkem Munday 69 kV line located south and west of Wichita Falls, and the Falfurrias 138/69kV transformer located south of the Eagle Ford shale.



Figure 42: Day-Ahead Congestion Costs by Zone

As they were in real time, higher costs associated with constraints in the South and Houston zones were the key drivers leading to increased total congestion costs in the day-ahead market during 2014.

# D. Congestion Revenue Rights Market

Congestion can be significant from an economic perspective, compelling the dispatch of highercost resources because power produced by lower-cost resources cannot be delivered due to transmission constraints. 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 are acquired by 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 corresponding to the difference in locational prices of the source and sink.

Figure 43 summarizes the revenues collected by ERCOT in each month for all CRRs, both auctioned and allocated. These revenues are distributed to loads in one of two ways. Revenues from cross-zone CRRs are allocated to loads ERCOT wide. Revenues from CRRs that have the source and sink in the same geographic zone are allocated to loads within that zone. This method of revenue allocation provides a disproportionate share of CRR Auction revenues to loads located in the West zone. In 2014, CRRs with both their source and sink in the West zone accounted for 42 percent of CRR Auction revenues. This revenue was allocated to West zone loads, which accounted for only 9 percent of the ERCOT total. In comparison, in 2013, 45 percent of CRR Auction revenues were allocated to the West zone load, which accounted for 8 percent of the ERCOT total. Allocating CRR Auction revenues in this manner helps reduce the impact of the higher congestion on West zone prices.





As shown in Figure 2, the annual average real-time energy price for the West zone was \$43.58 per MWh, nearly \$3 per MWh higher than the ERCOT-wide average. The value of CRR Auction revenues distributed only to the West zone equated to \$4.40 per MWh higher than the ERCOT-wide average distribution of CRR Auction revenues. This was sufficient to offset the higher real-time prices incurred in the West load zone during 2014. In 2013 the annual average price for the West zone was \$37.99 per MWh, which was about \$4 per MWh higher than the ERCOT-wide average, and the incremental CRR Auction revenues were almost \$5.50 per MWh.

Next, Figure 44 compares the value received by CRR owners (in aggregate) to the price paid to acquire the CRRs. Although results for individual participants and specific source/sink combinations varied, the aggregated results in most months show that participants did not overpay in the auction. The exceptions were the summer months of June, July and August. Across the entire year of 2014, participants spent \$375 million to procure CRRs and received \$491 million.



Figure 44: CRR Auction Revenue and Payment Received

The next look at aggregated CRR positions adds day-ahead congestion rent to the picture.

Simply put, 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. Revenue from congestion rent creates the source of funds used to make payments to CRR owners. Figure 45 presents CRR Auction revenues, payment to CRR owners, and day-ahead congestionrent in 2013 and 2014, by month. Congestion rent for the year 2014 totaled \$528 million and payment to CRR owners was \$491 million.



Figure 45: CRR Auction Revenue, Payments and Congestion Rent

The target value of a CRR is the MW amount of the CRR multiplied by the 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 CRRs, if modeled on the day-ahead network, would cause a higher flow on a transmission line than the line's rating, thereby creating a

constraint. 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, that excess congestion rent can be used to make whole the CRR account holders that received shortfall charges.



**Figure 46: CRR Shortfalls and Derations** 

Figure 46 shows the amount of target payment, deration amount, and net shortfall charges (after make whole payments) for 2014. In 2014 the total target payment to CRRs was \$521 million; however, there were \$19 million of derations and \$11 million of shortfall charges leaving a final payment to CRR account holders of \$491 million. This corresponds to a CRR funding percentage of 94 percent.

The last look at congestion examines the price spreads for each pair of hub and load zone in more detail. These price spreads are interesting as many loads may have contracts that hedge them to the hub price and are thus exposed to the price differential between the hub and its corresponding load zone. Figure 47 presents the price spreads between the West Hub and West load zone as valued at four separate points in time – at the semi-annual CRR Auction, monthly CRR auction, day-ahead and in real-time. Since CRRs are sold for one of three defined time periods, weekday on peak, weekend on peak, and off peak, Figure 47 includes a separate comparison for each. Figure 48, Figure 49, and Figure 50 present the same information for the North, South and Houston load zones, respectively.

Of note is that the same intra-zone congestion that drives the relatively high CRR auction revenue amounts for the West zone also drives high price spreads between the West hub and the West load zone. Of the other zones only the South has price spreads approaching those of the West.







# Figure 48: North Hub to North Load Zone Price Spreads







**Figure 50:** Houston Hub to Houston Load Zone Price Spreads

### IV. DEMAND AND SUPPLY

This section reviews and analyzes the load patterns during 2014 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 Loads in 2014

The changes in overall load levels from year to year can be shown by tracking the changes in average load levels. This metric will tend to capture changes in load over a large portion of the hours during the year. It is also important to separately evaluate the changes in the load during the highest-demand hours of the year. Significant changes in 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 and inconvenience to customers that can result from interruptions to that service. Hence, both of these dimensions of load during 2014 are examined in this subsection and summarized in Figure 51.

This figure shows peak load and average load in each of the ERCOT zones from 2011 to 2014.<sup>6</sup> 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 51 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.

<sup>&</sup>lt;sup>6</sup> For purposes of this analysis, Non-Opt In Entity (NOIE) Load Zones have been included with the proximate geographic Load Zone.



Figure 51: Annual Load Statistics by Zone

Total ERCOT load over the calendar year increased from 332 terawatt-hours (TWh) in 2013 to 340 TWh in 2014, an increase of 2.5 percent or an average of 960 MW every hour. Much of this increase occurred in the first quarter as extremely cold weather contributed to record levels of winter load.

Despite the increase in average load, the ERCOT coincident peak hourly demand decreased from 67,247 MW to 66,451 MW in 2014, a decrease of 795 MW, or 1.2 percent. The highest peak demand experienced in ERCOT remains 68,311 MW that occurred during August of 2011.

The changes in load at the zonal level are not the same as the ERCOT-wide changes. The growth rate of West zone average load was once again much higher, on a percentage basis, than the other zones. Similarly, peak load in the West zone increased nearly 500 MW in 2014. Peak load did not increase in any other zone.

To provide a more detailed analysis of load at the hourly level, Figure 52 compares load duration curves for each year from 2011 to 2014. A load duration curve shows the number of hours (shown on the horizontal axis) that load exceeds a particular level (shown on the vertical axis). ERCOT has a fairly smooth load duration curve, typical of most electricity markets, with low to moderate electricity demand in most hours, and peak demand usually occurring during the late afternoon and early evening hours of days with exceptionally high temperatures.





As shown in Figure 52, the load duration curve for 2014 is higher than in 2013 for all but the very highest load hours in the year. This is consistent with the aforementioned 2.5 percent load increase from 2013 to 2014. Still noticeable are the much higher loads experienced in 2011. Even with three years of energy growth the highest 2,000 hours of loads experienced in 2011 remain the highest.

To better illustrate the differences in the highest-demand periods between years, Figure 53 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 95<sup>th</sup> percentile of hourly load. From 2011 to 2014, the peak load value averaged nearly 18 percent greater than the load at the 95<sup>th</sup> 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.





## **B.** Generation Capacity in ERCOT

The generation mix in ERCOT is evaluated in this subsection. The distribution of capacity among the ERCOT zones is similar to the distribution of demand with the exception of the large amount of wind capacity in the West zone. The North zone accounts for approximately 37 percent of capacity, the South zone 28 percent, the Houston zone 21 percent, and the West zone 14 percent. Excluding mothballed resources and including only the fraction of wind capacity available to reliably meet peak demand,<sup>7</sup> the North zone accounts for approximately

<sup>&</sup>lt;sup>7</sup> The percentages of installed capacity to serve peak demand assume wind availability of 12 percent for noncoastal wind and 56 percent for coastal wind.

40 percent of capacity, the South zone 32 percent, the Houston zone 21percent, and the West zone 7 percent. Figure 54 shows the installed generating capacity by type in each of the ERCOT zones.<sup>8</sup>



Approximately 2.8 GW of new generation resources came online in 2014. Gas-fueled units accounted for 2.1 GW of the total additions, primarily from two new combined cycle units. The remaining gas additions were a new combined cycle unit built on an existing site of a retired gas steam unit, and the addition of a gas turbine at an existing generator location. The remaining resource additions were wind (0.7 GW) and small solar units. When unit retirements are included, the net capacity addition in 2014 was 1.6 GW. Natural gas generation continues to account for approximately 48 percent of total ERCOT installed capacity while the share of coal generation dropped slightly from 21 percent in 2013 to 20 percent in 2014.

<sup>&</sup>lt;sup>8</sup> For purposes of this analysis, generation located in a NOIE Load Zone has been included with the proximate geographic Load Zone.

By comparing the current mix of installed generation capacity to that in 2007, as shown in Figure 55, the effects of longer term trends can be seen. Over these seven years, more new wind and coal generation has been added than any other type of capacity.<sup>9</sup> The sizable additions in these two categories have been more than offset by retirements of old natural gas-fired steam units. Nonetheless, the resulting installed capacity in 2014 was 1 GW more than in 2007. Comparatively, peak load in 2014 was greater than the 2007 peak load by more than 4 GW.



Figure 55: Installed Capacity by Type: 2007 to 2014

The shifting contribution of coal and wind generation is evident in Figure 56, which shows the percentage of annual generation from each fuel type for the years 2007 through 2014. The generation share from wind has increased every year, reaching 11 percent of the annual generation requirement in 2014, up from 3 percent in 2007. During the same period the percentage of generation provided by natural gas has ranged from a high of 45 percent in 2007 to

<sup>&</sup>lt;sup>9</sup> Wind capacity is shown at its full installed capacity in this chart.

a low of 38 percent in 2010. In 2014 the percentage of generation from natural gas was 41 percent, which was a very slight increase from the 2013 level.<sup>10</sup> Similarly, the percentage of generation produced by coal units ranged from a high of 40 percent in 2010 to a low of 34 percent in 2012. The percentage of generation from coal was 36 percent in 2014, a small decrease from 37 percent in 2013.



Figure 56: Annual Generation Mix

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 is approximately 23.4 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.

<sup>&</sup>lt;sup>10</sup> Natural gas provided 40.5 percent of total generation in 2013, and 41.1 percent in 2014.

The generation mix in 2012 remains notable due to the reduced share of coal generation. For the first time the combination of coal and nuclear units provided less than 50 percent of the annual energy requirements. The reduced contribution from coal in 2012 was directly related to relatively low natural gas prices experienced that year. Low natural gas prices allow efficient gas units to produce electricity at lower costs than most coal units in ERCOT, leading to the noticeable displacement of coal observed in 2012. As natural gas prices increased in 2013 and 2014 the amount of coal displacement has decreased and the generation share from coal has increased.

## 1. Wind Generation

The amount of wind generation installed in ERCOT exceeded 12 GW by the end of 2014. 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. This subsection will more fully describe the characteristics of wind generation in ERCOT.



**Figure 57: Average Wind Production** 

The average profile of wind production is negatively correlated with the load profile, with the highest wind production occurring during non-summer months, and predominately during off-peak hours. Figure 57 shows average wind production for each month in 2013 and 2014, with the average production in each month shown separately in four hour blocks.

The completion of the CREZ lines in late 2013 eliminated what had been a longstanding constraint limiting the export of wind from the West zone. There continue to be localized constraints limiting wind generation at certain locations.

Examining wind generation in total masks the different wind profiles that exist for locations across ERCOT. Wind developers have more recently been attracted to site facilities along the Gulf Coast of Texas due to the higher correlation of winds with electricity demands. 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 58: Summer Wind Production vs. Load

Figure 58 presents data for the summer months of June through August, comparing the average output for wind generators located in coastal and non-coastal areas in ERCOT across various load levels. It shows a strong negative relationship between non-coastal wind output and increasing load levels. It further shows that the output from wind generators located in the coastal area of the South zone is much more highly correlated with peak electricity demand.

The growing numbers of solar generation facilities in ERCOT also have an expected generation profile highly correlated with peak summer loads. Figure 59 below 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. The total installed capacity of solar generation is much smaller than that of wind generation. However, its production as a percentage of installed capacity is the highest, nearing 80 percent in the early afternoon, and producing more than 60 percent of its installed capacity during peak load hours.





The contrast between coastal wind and non-coastal wind is also clearly displayed in Figure 59. Coastal wind produced greater than 50 percent of its installed capacity during summer peak hours while output from non-coastal wind was between 20 and 30 percent during summer peak hours.



Figure 60 shows the wind production and estimated curtailment quantities for each month of 2012 through 2014. This figure reveals that the total production from wind resources continued to increase and the quantity of curtailments was reduced in 2014. The volume of wind actually produced was estimated as 99.5 percent of the total available wind in 2014, up slightly from 98.9 percent in 2013 and 96 percent in 2012.

Increasing levels of wind resources in ERCOT also has important implications for the net load duration curve faced by the non-wind fleet of resources. Net load is defined as the system load minus wind production. Figure 61 shows the net load duration curves for the years 2011 through 2014, normalized as a percentage of peak load, and including 2007 as a point of reference.



Figure 61: Net Load Duration Curves

This figure 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. Wind generation erodes the amount of energy available to be served by baseload coal units, while doing very little to reduce the amount capacity necessary to reliably serve peak load.

Even with the increased development activity in the coastal area of the South zone, nearly 80 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 relatively low system demand. This profile results in only modest reductions of the net load relative to the actual load during the hours of highest demand, but much more significant reductions in the net load relative to the actual load relative to the actual load in the other hours of the year.

Focusing on the left side of the net load duration curve shown in Figure 62, the difference between peak net load and the 95<sup>th</sup> percentile of net load has averaged 12 GW the past three years.





On the right side of the net load duration curve, the minimum net load has dropped from approximately 20 GW in 2007 to below 16 GW last year, even with sizable growth in total annual load. This continues to put operational pressure on the 23.5 GW of nuclear and coal-fired generation currently installed in ERCOT.

Thus, although the peak net load and reserve margin requirements are projected to continue to increase and create an increasing need for non-wind capacity to meet net load and reliability requirements, the non-wind fleet can expect to operate for fewer hours as wind penetration continues to increase. This outlook further reinforces the importance of efficient energy pricing

during peak demand conditions and other times of system stress, particularly within the context of the ERCOT energy-only market design.

### 2. Generator Commitments

One of the important characteristics of any electricity market is the extent to which it results in the efficient commitment of generating resources. Under-commitment can cause apparent shortages in 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 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 it is important to note that ERCOT's day-ahead market is only financially binding. That is, although a generator's offer to sell is selected (cleared) in the day-ahead market there is no corresponding requirement to actually start that unit. Nonetheless, the generator will be financially responsible for replacing its offered capacity if it does not start the unit.

The following figure compares the amount of on-line reserves in 2014 and 2013. The amount of on-line reserves is equal to the amount of the capacity committed in excess of expected demand. Figure 63 displays available online reserves aggregated by total system load levels and shows the expected pattern of declining reserves as system load increases. Further, at all but the very highest system loads, there were more online reserves in 2014 than in 2013. This indicates that more capacity was online during 2014.



Figure 63: Average On-line Reserves

Two possible explanations for the increase in capacity commitments in 2014 are: (1) response to the increased payment made to online reserves with the implementation of ORDC on June 1, and (2) increased hedging behavior by entities wanting to avoid potential exposure to \$7,000 per MWh prices resulting from the higher system-wide offer cap.

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



Figure 64: Frequency of Reliability Unit Commitments

Figure 64 summarizes, by month, the number of hours with units committed via the reliability unit commitment process. There was increased reliance on the reliability unit commitment process in 2014. During 2014, the number of hours with at least one unit receiving a reliability unit commitment instruction was 19 percent. The increase was noticeable given the relatively low occurrences in 2012 and 2013 when the number of hours with at least one unit receiving a reliability unit commitment instruction was 3 percent and 5 percent, respectively. During 2011, approximately one third of the hours had at least one unit committed by ERCOT through the reliability unit commitment process.

One cause for the increase in reliability unit commitment activity in 2014 was related to maintaining reliable service in the Rio Grande Valley. Almost half (47 percent) of the hours with at least one unit receiving a reliability unit commitment instruction in 2014 were related to conditions in the Valley. Another reason for the increase in 2014 was due to the more extreme weather conditions during the winter (January through March). Natural gas curtailments to
power plants are more common as the temperature drops. In these situations it is not unusual for ERCOT to use the reliability unit commitment process to ensure generation capacity using fueloil is available.

The low number of hours in 2012 and 2013 can be attributed, in part, to the less extreme weather and resulting lower load levels experienced. There also was an operational change midway through 2011 which contributed to the reduced frequency of reliability unit commitments. During the initial months of operating the nodal market, it was common for ERCOT to commit units that were providing non-spin reserves if they were needed to resolve congestion. This practice was greatly reduced starting in July 2011.

The majority of reliability unit commitment instructions are to resolve localized transmission constraints. Less than 18 percent of the unit hours of RUC instructions in 2014 were for system-wide capacity requirements and these hours were primarily during the period from January through March.

The next analysis compares the average dispatched output of the reliability committed units with their operational limits. Figure 65 below shows that the quantity of reliability unit commitment generation decreased in 2014; even though, as previously described, the frequency of reliability unit commitment increased in 2014.



Figure 65: Reliability Unit Commitment Capacity

There was less variation in the average quantity of reliability committed capacity in 2014. The average quantity dispatched was generally between 100 and 200 MW for all but the summer months.

Factors contributing to the high average capacity in October 2013 included an unseasonably warm day leading to system-wide capacity deficiency and localized generation requirements because of North to Houston and Valley import transmission constraints. April 2013 capacity needs were primarily in the Dallas-Fort Worth area for voltage support. The large amounts of reliability unit committed capacity in April 2012 were related to brief generator outages resulting in reactive power deficiencies in the Dallas-Fort Worth area. This was similar to the situation that existed during October 2011. The larger quantity of committed capacity in February 2011 was a result of ERCOT operator action taken to attempt to ensure overall capacity adequacy

during both the extreme cold weather event that occurred early in the month, and a subsequent bout of cold weather that occurred one week later.

# 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. Unlike active participation in ERCOT-administered markets, self-dispatch by demand is not directly tracked by ERCOT.

### 1. 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 in each year. As of December 2014, approximately 3,154 MW of capability were qualified as Load Resources.

Figure 66 shows the amount of responsive reserves provided from load resources on a daily basis in 2014. The high level of participation by demand response in the ancillary service markets sets ERCOT apart from other operating electricity markets. For reliability reasons the maximum amount of responsive reserves that can be provided by load resources is limited to 50 percent of the total, or 1,400 MW. The fifty percent limit has been maintained even as the total amount of responsive reserves increased from 2,300 MW to 2,800 MW in April of 2012.



Figure 66: Daily Average of Responsive Reserves Provided by Load Resources

Figure 66 shows amounts of responsive reserves that were either self-scheduled or offered by load resources. The quantity of offers submitted by load resources exceeds the 50 percent limit most of the time. Times when this is not the case generally correspond with periods of expected high real-time prices. Since load resources provide capacity by reducing their consumption, they have to actually be consuming energy to be eligible to provide the capacity service. During periods of expected high prices the price paid for the energy can exceed the value received from providing responsive reserves. Noticeable reductions can be seen in February 2011 and the summer months of 2011. Reductions in the amount of offers are also observed every year around October, which generally 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.

## 2. Reliability Programs

There are two main reliability programs in which demand can participate in ERCOT, Emergency Response Service and transmission provider load management programs. The Emergency Response Service (ERS) product is defined by PUCT Rule enacted in March of 2012. It replaced a previously defined emergency load service that was created in 2007. As originally conceived in 2007, the program would have ERCOT procure 500 to 1000 MW of load that would submit to being curtailed during emergency conditions, just prior to the forced curtailment of firm load. Although \$20 million was initially allocated to fund this procurement, less than 500 MW of loads offered to be included.

Several program changes have been implemented over the years, so that now almost \$50 million is spent annually to procure, on average, slightly less than 800 MW. The amount of ERS procured ranged from 600 to 1000MW across the various periods in the 2014 program year. Beginning with the auction covering the first period of program year 2014 (February 1 – May 31) the program was modified from a pay as bid auction to a clearing price auction, increasing participation and providing a clearer incentive to load to submit offers based on their costs to curtail, including opportunity cost. ERS was deployed only once in 2014. The time weighted average price paid in 2014 to providers of ERS service was \$7.15 per MWh. As a point of comparison, the average paid to providers of responsive reserve service was \$14.22 per MWh.

Beyond ERS there are slightly more than 200 MW of load participating in load management programs administered by transmission providers. 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 curtailments are actually controlled by ERCOT and occur during EEA Level 2.

# 3. 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 by actions taken specifically to avoid the allocation of transmission costs. Of these two methods, the more significant are actions taken specifically to avoid the allocation of transmission costs.

Transmission costs have for decades been allocated to all loads in ERCOT on the basis of load contribution to the highest 15-minute system demand during each of the four months, June – September. This allocation mechanism is routinely referred to as four coincident peak, or 4CP. Over the last three years these transmission costs have risen by more than 60 percent, thus significantly increasing an already substantial incentive to reduce load during probable peak intervals in the summer. It is estimated that over 800MW of load is actively pursuing reduction during these intervals.

# Pricing During Load Deployments

During times when there are shortages of supply offers available for dispatch and Responsive Reserves are deployed, that is, converted to energy as one of the last steps taken before shedding firm load, the value of the foregone reserves – which is much higher than the marginal cost of the most expensive online generator – should be reflected in energy prices to achieve efficient economic signals governing investment in generation, demand response, and transmission. Unfortunately, ERCOT's dispatch software does not recognize that load has been curtailed, and computes prices based on supplying only the remaining load. A good example of this situation occurred on August 4, 2011. Figure 67 displays available reserves and the system price for that afternoon and shows that even though reserves were below required levels, system prices dropped to \$60 per MWh. At this level, prices are being set based on supply offers and do not reflect the value of the load that is being curtailed to reliably serve the remaining system demand.



Figure 67: Pricing During Load Deployments

In 2014 ERCOT took the first step toward including the actions taken by load during the realtime energy market. The first phase of "Loads in SCED" allows those controllable loads that can respond to 5-minute dispatch instructions to specify the price at which they no longer wish to consume. This change was implemented in June of 2014. Although an important first step, there are very few loads that can respond to price in this manner.

We recommend that ERCOT implement system changes that will ensure that all demand response that is actively deployed by ERCOT be able to set the price at the value of load at times when such deployments are necessary to reliably serve the remaining system demand. This includes load resources and ERS providers being deployed for the services they contracted to provide or when firm load is involuntarily curtailed. It may be possible to integrate load bids and emergency resources in the real-time dispatch software and allow them to set prices when they are effectively marginal. Alternatively, it may be adequate to address this concern through administrative shortage pricing rules.

## V. RESOURCE ADEQUACY

One of the primary functions of the wholesale electricity market is to provide economic signals that will facilitate the investment needed to maintain a set of resources that are adequate to satisfy the system's demands and reliability needs. 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. 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. The section concludes with a discussion of the impacts of the Operating Reserve Demand Curve implemented last year.

## 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. Net revenues from the real-time energy and ancillary services markets provide economic signals that help inform suppliers' decisions to invest in new generation or retire existing generation. 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 their investment decisions on their 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. Variable costs (fuel and O&M) were assumed to be \$24 per MWh for the coal unit and \$8 per MWh for the nuclear. A total outage rate (planned and forced) of 10 percent was assumed for each technology.

The next two figures provide an historical perspective of the net revenues available to support new natural gas combustion turbine (Figure 68) and combined cycle generation (Figure 69).



Figure 68: Combustion Turbine Net Revenues

Based on updated estimates of investment costs for new units,<sup>11</sup> 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 updated estimates of annual fixed costs have been reduced to reflect lower power plant equipment costs and further reduced to reflect Texas-specific construction costs. The net revenue in 2014 for a new gas turbine was calculated to be approximately \$37 per kW-year, below the estimated cost of new gas turbine generation.



**Figure 69: Combined Cycle Net Revenues** 

For a new combined cycle gas unit, the updated estimate of net revenue requirement is approximately \$110 to \$125 per kW-year, also reflecting lower power plant equipment costs and

<sup>11</sup> Estimated annual fixed costs are derived from the EIA estimates released April 12, 2013 available here: http://www.eia.gov/forecasts/capitalcost/.

further reduced to reflect Texas-specific construction costs. The net revenue in 2014 for a new combined cycle unit was calculated to be approximately \$57 per kW-year, also below the estimated cost of new combined cycle generation.

Even though net revenues for the Houston and South zones in 2008 may have appeared to be sufficient to support new natural gas-fired generation, the higher prices actually resulted from extremely inefficient transmission congestion management and inefficient pricing mechanisms associated with the deployment of non-spinning reserves, thereby contributing to higher than warranted net revenues. Discounting the effect that the 2008 results would have had on forward price signals, 2011 has been the only year with net revenues that would have been sufficient to support either new gas turbine or combined cycle generation.

Figure 70 expands the net revenue analysis to include coal and nuclear generation in addition to natural gas-fired combustion turbine and combined-cycle generation. Estimated net revenues for the four types of generation are compared below for 2013 and 2014.



Figure 70: Estimated Net Revenue by Zone and Unit Type

For the natural gas units, net revenue is calculated by assuming the unit will produce energy in any hour for which it is profitable and by assuming it will be available to sell reserves and regulation in all other hours. For coal and nuclear technologies, net revenue is calculated solely from producing energy.

Overall, the net revenues in 2014 were higher than those in 2013 and 2012, and all three years were much lower than in 2011. This is not surprising given shortages have been very infrequent over the past three years. Shortage pricing plays a pivotal role in providing investment incentives in an energy-only market like ERCOT. In order to provide adequate incentives, some years must exhibit an extraordinary number of shortages and net revenues that are multiples of annual net revenues needed to support investment.

As previously described, the 2014 net revenue for the natural gas-fired technologies was somewhat higher than 2013 levels, primarily because of higher gas prices during the first quarter of 2014. Net revenues for coal and nuclear technologies increased by larger amounts from 2013 to 2014 because they benefit from the increase in natural gas prices.

Despite these increases, the net revenues produced by the ERCOT markets in 2014 were lower than the annualized cost of investing in any of these new technologies.

- For a new coal unit, the estimated net revenue requirement is approximately \$265 to \$310 per kW-year. The net revenue in 2014 for a new coal unit was calculated to be approximately \$105 per kW-year.
- For a new nuclear unit, the estimated net revenue requirement is approximately \$450 to \$585 per kW-year. The net revenue in 2014 for a new nuclear unit was calculated to be approximately \$227 per kW-year.

Prior to 2005, net revenues were well below the levels necessary to justify new investment in coal and nuclear generation. Higher natural gas prices from 2005 through 2008 resulted in sustained energy prices high enough to support new entry for these technologies. The production costs of coal and nuclear units did not change significantly over this period, leading to a dramatic rise in net revenues. However, natural gas prices peaked in 2008, resulting in reduced net revenues for coal and nuclear technologies since then. Even with the higher energy prices experienced in 2011, net revenues for these technologies were calculated to be less than the

estimated cost of new entry. Very low natural gas prices and few occurrences of shortage pricing during 2012 resulted in calculated net revenue for coal and nuclear to be well below the estimated cost of new entry. Although natural gas prices increased in 2013, the calculated net revenue for coal and nuclear technologies was less than the estimated cost of new entry. Similarly, net revenue for coal and nuclear technologies in 2014 was again less than the estimated cost of new entry.

The net revenues in 2014 were higher than those in 2013 and 2012, and all three years were much lower than in 2011. These results indicate that during 2014 the ERCOT markets would not have provided sufficient revenues to support profitable investment in any of the types of generation technology evaluated. Therefore, it may seem inconsistent with these results that new generation continues to be added in the ERCOT market. This can be explained by the following factors:

First, the net revenues in any one year may be higher or lower than an investor would require over the long term. In 2014, the net revenues were substantially lower than the cost of entry because 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 the 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.

Second, the costs of new entry used in this report are generic and reflective of the costs of new resources on a new, 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 equipment costs, possibly through large, long-term supply agreements, or by adding generation to existing sites, or some combination of both.

Third, in addition to the equipment cost, financing structures and costs can vary greatly between suppliers. Again, the net revenue analysis assumes generic financing costs that a specific supplier may be able to improve on. 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. Suppliers will develop their own view of future expected revenue which may include a power sales contract for some amount of the output. A long-term power sales contract could provide them with more revenue certainly than is available by relying completely on the ERCOT wholesale market. 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.

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-fired 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 two types of natural gas generators for the ERCOT North zone, PJM, two locations within the New York ISO, and the Midcontinent ISO. Most of these locations are central locations with the exception of New York City, which is significantly affected by congestion. Figure 71provides a comparison of net revenues for a combustion turbine and Figure 72 provides the same comparison for a combined cycle unit.



Figure 71: 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. The nascent capacity market in MISO contributed a small amount to net revenues in 2014. Net revenues for all other regions are calculated for central locations. However, there are load pockets within each market where net revenue and the cost of new investment may be higher. The NYC zone of NYISO is presented as 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. Finally, resource investments are driven primarily by forward price expectations, so historical net revenue analyses do not provide a complete picture of the future pricing expectations that will spur new investment.



Figure 72: Combined Cycle Net Revenue Comparison Between Markets

Both figures indicate that across all markets net revenues increased in 2014. The increases were more noticeable in PJM and NYISO, primarily due to higher energy revenues as a result of the extreme winter weather experienced in those regions early in 2014.

Over the long-run, markets should provide sufficient net revenue to allow generation owners to receive a return of, and on an investment in a new generating unit when that unit is needed. In the short-run, if the net revenues produced by the market are not sufficient to justify entry, then one or more of these conditions exist:

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

Likewise, the opposite would be true if the markets provide excessive net revenues in the shortrun. The persistence of excessive net revenues in the presence of a capacity surplus is an indication of competitive issues or market design flaws.

In an energy only market, net revenues are expected to be less than required to support new investment in most years. However, in the small number of years that are much worse than normal, the sharp increase in the frequency of shortage pricing should cause the net revenues in that year to be multiples of the annual level required to support investment. This pattern over the long run must create an expectation that net revenues, on average, will support new investments.

# B. Effectiveness of the Scarcity Pricing Mechanism

The Public Utility Commission of Texas (PUCT) adopted rules in 2006 that define the parameters of an energy-only market. These rules included a Scarcity Pricing Mechanism (SPM) that increased the system-wide offer cap in multiple steps until it reached \$3,000 per MWh shortly after the implementation of the nodal market. In accordance with the IMM's charge to conduct an annual review,<sup>12</sup> this subsection assesses the SPM in 2014 under ERCOT's energy-only market structure.

Revisions to 16 TEX. ADMIN. CODE § 25.505 were adopted in 2012 that specified the following increases to the system-wide offer cap:

- \$5,000 per MWh beginning on June 1, 2013,
- \$7,000 per MWh beginning on June 1, 2014, and
- \$9,000 per MWh beginning on June 1, 2015.

As shown in Figure 16 on page 16, there have been very brief periods when energy prices rose to the cap since the system-wide offer cap was increased to greater than \$3,000 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 result in reducing the system-wide offer cap. PNM also serves as a simplified measure of the annual net revenue of a hypothetical

<sup>&</sup>lt;sup>12</sup> See 16 Tex. Admin. Code § 25.505(g)(6)(D).

peaking unit.<sup>13</sup> Under the current rule, if the PNM for a year reaches a cumulative total of \$300,000 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.<sup>14</sup> Figure 73 shows the cumulative PNM results for each year from 2006 through 2014 and shows that PNM in 2014 was similar to the level of 2009.



Figure 73: Peaker Net Margin

As previously described, the net revenue required to satisfy the reduced estimates of the annual fixed costs (including capital carrying costs) of a new gas turbine unit ranges from \$75,000 to \$90,000 per MW-year. Thus, as shown in Figure 73 and consistent with the previous findings in this section relating to net revenue, the PNM was slightly below the levels estimated to support new entry in 2014.

<sup>&</sup>lt;sup>13</sup> 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.

<sup>&</sup>lt;sup>14</sup> 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 PNM threshold for 2014 and each subsequent year will be set to \$315,000 per MW-year based on the analysis prepared by Brattle dated June 1, 2012, unless there is a change identified in the cost of new entry of new generation plants.

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.

#### 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 projection of reserve margins developed prior to the summer of 2015.



**Figure 74: Projected Reserve Margins** 

Source: ERCOT Capacity Demand Reserve Reports / 2015 from December 2014 and 2016-2020 from May 2015

Figure 74 above indicates that the region will have a 15.7 percent reserve margin heading into the summer of 2015. Reserve margins are now expected to exceed the target level of 13.75 percent for the next several years. These projections are higher than those developed last year. Further, this outlook is very different than in 2013, when reserve margins were expected to be below the target level of 13.75 percent for the foreseeable future. In 2013 the expected

reserve margin for 2016 was 10.4 percent, much lower than the current expectation for 2016 of 17 percent. This increase in expected reserve margin is not due to an increase in available generating resources, but rather to ERCOT's revised long-term load forecasting methodology and resulting reduction in the forecasted peak demand. The quantity of available resources expected in 2016 as shown in the May 2013 Capacity Demand Report (CDR) is nearly identical to the quantity of resources shown in the May 2015 CDR. Although the total expected capacity of resources has not changed between the two CDRs, the mix has changed. Almost 1,700 MW of increased wind capacity expected in 2016 has been offset by reductions in the total capacity expected from natural gas and coal.

The figure to the right presents a comparison of ERCOT's peak demand forecasts from recent CDR reports. Comparing the May 2013 forecast with the December 2014 forecast, the difference in peak demand expected in 2016 is greater than 4,000 MW.

Looking beyond 2016, several new additions have been announced that



meet the requirements for being included in the CDR. The bulk of this new capacity is from new gas units (greater than 5 GW) sited at locations across the ERCOT region. Wind additions are also projected to continue, with 1.5 GW of capacity shown in the CDR representing nearly 10 GW of installed wind capacity. Rounding out the additions is more than 500 MW of solar capacity.

To compare the situation in ERCOT with other regions, Figure 76 provides the anticipated reserve margins for the North American Electric Reliability Council (NERC) regions in the

United States for the summer of 2015, as of the most recent NERC report in November 2014.<sup>15</sup> Figure 76 shows that required, or reference level reserve margins center around 15 percent across other regions. These regions run the gamut from traditional bundled, regulated utility service territories to fully competitive, centrally operated wholesale markets. There are large differences in the level of planning reserves expected for the summer of 2015. However, reserve margins are lower in nearly every region this year compared to last. ERCOT is unique in that its anticipated reserve margin remains very close to its target level. Even with the forecasted additions, ERCOT is projected to sustain lower reserve margins than many other regions. This makes it important to ensure that the ERCOT market is designed to provide adequate economic signals to remain near this target, which is discussed below



### **Figure 76: Reserve Margins in Other Regions**

<sup>&</sup>lt;sup>15</sup> Data from NERC 2014 Long-Term Reliability Assessment (November 2014) available at <u>http://www.nerc.com/pa/RAPA/ra/Reliability% 20Assessments% 20DL/2014LTRA\_ERATTA.pdf</u>. For the most recent projected reserve margins for ERCOT, please see Figure 74 and the associated discussion supra.

### 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 are continually increasing, 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-scarcity, energy prices during scarcity and capacity payments. The capacity payments generators receive in ERCOT are related to the provision of ancillary services. As discussed in the net revenue subsection, ancillary service payments are a small contributor: \$5 - \$10 per kW-year. Setting them aside, generator revenue in ERCOT is overwhelmingly derived from energy prices under both scarcity and non-scarcity conditions.

Expectations for energy pricing under non-scarcity 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 at 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 produce a planning reserve margin that is less than the planning reserve target.

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

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

A market design that adheres to the pricing principles stated above will set prices that reflect each of these actions. When there is a shortage of supply in the market, the marginal action first taken by the system operator is generally to sacrifice operating reserves requirements (*i.e.*, dispatch reserves for energy). Diminished operating reserves results in an increased probability of outage, which has a real cost to electricity consumers. In this case, the value of the foregone reserves – which is much higher than the marginal cost of the most expensive online generator – should be reflected in energy prices to achieve efficient economic signals governing investment in generation, demand response, and transmission.

Faced with reduced levels of generation development activity coupled with increasing loads that result in falling planning reserve margins, the PUCT has devoted considerable effort since 2012 deliberating issues related to resource adequacy. To date, the PUCT continues to support the energy-only nature of the ERCOT market and has directed market modifications to improve ERCOT's shortage pricing based on the demand for operating reserves.

# E. Operating Reserve Demand Curve Implementation

The Operating Reserve Demand Curve (ORDC) is a shortage pricing mechanism that reflects the loss of load probability (LOLP) at varying levels of operating reserves multiplied by the value of lost load (VOLL). Selected as an easier to implement alternative to real-time co-optimization of energy and ancillary services, the ORDC provides a new form of shortage pricing for online and offline reserves, as well as energy. As the quantity of reserves decreases, payments will increase. As conceptualized below in Figure 77, once available reserve capacity drops to 2,000 MW, payment for reserve capacity will rise to VOLL, or \$9,000 per MWh.



Figure 77: Operating Reserve Demand Curve

The initial implemetation of ORDC went into effect on June 1, 2014 and included the introduction of real-time reserve on-line and off-line adders. Since real-time co-optimization of energy and ancillary services was not implemented, a mechanism was needed to ensure that resources are indifferent between providing energy and reserves in real-time. This is accomplished using an ancillary service imbalance settlement, with adjustments to the price floors that had previously been in place for the energy associated with capacity providing ancillary services. There is no longer a price floor associated with regulation, responsive reserves, or off-line non spin. The price floor associated with on-line non-spin was reduced to \$75 per MWh. The price floor associated with RUC capacity is now set at \$1500 per MWh.

The load-weighted real-time energy price for the period of 2014 after ORDC implementation (i.e. after June 1<sup>st</sup>) was \$35.68 per MWh. Of that total, \$0.26 per MWh (less than 1 percent) was the on-line reserve adder. The on-line reserve adder includes the off-line adder, which was \$0.09 per MWh for this time period.

Figure 78 presents the online reserve adder amount and associated reserve level for every 15 minute settlement period after June 1.



Figure 78: Online Reserve Adder

Although the pricing impacts due to ORDC implementation have so far been very small, the concern remains that prices resulting from ORDC will rise to levels approaching the VOLL when the available reserves are at levels where the LOLP is less than 1.0 and involuntary load curtailment is not imminent. This situation would likely lead to inefficient actions by participants. We will evaluate this concern going forward as the ORDC is fully implemented.

Finally, we continue to recommend that ERCOT implement a system to co-optimize energy and ancillary services because this would improve the efficiency of ERCOT's dispatch, more fully utilize its resources, and allow for improvements in its shortage pricing.

# VI. 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 is consistent with observations in prior years. This section also includes a summary of the Voluntary Mitigation Plans in effect during 2014. 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 2014.

### A. Structural Market Power Indicators

The market structure is analyzed by using the Residual Demand Index (RDI), a statistic that measures the percentage of load that could not be satisfied without the resources of the largest supplier. The RDI is used to measure the percentage of load that cannot be served without the resources of the largest supplier, assuming that the market could call upon all committed and quick-start capacity owned by other suppliers.<sup>16</sup> When the RDI is greater than zero, the largest supplier is pivotal (i.e., its resources are needed to satisfy the market demand). When the RDI is less than zero, no single supplier's resources are required to serve the load as long as the resources of its competitors are available.

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

<sup>&</sup>lt;sup>16</sup> For the purpose of this analysis, "quick-start" includes off-line simple cycle gas turbines that are flagged as online 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.

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

Figure 79 shows the RDI relative to load for all hours in 2014. 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. It is possible that they also control other capacity through bilateral arrangements, although we do not know whether this is the case. To the extent that the resources scheduled by the largest QSEs are not controlled by or provide revenue to the QSE, the RDIs will tend to be slightly overstated.





Figure 80 below summarizes the results of the RDI analysis by displaying the percentage of time at each load level there was a pivotal supplier. At loads greater than 65 GW there was a pivotal supplier 100 percent of the time. The figure also displays the percentage of time each load level

occurs. By combining these values it can be determined that there was a pivotal supplier in approximately 23 percent of all hours of 2014, which indicates that market power is a potential concern in ERCOT and underscores the need for the current mitigation measures that address it.



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

## Voluntary Mitigation Plans

Voluntary Mitigation Plans (VMPs) existed for two market participants – NRG and Calpine – during 2014. 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 in the ERCOT real-time energy market through economic withholding. The same restrictions are not required in forward energy markets (e.g., the ERCOT day-ahead market) because the price in forward energy markets is derived from the real-time energy prices. Because the 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.

NRG's plan, initially approved in June 2012 and modified in May 2014, 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 for each natural gas-fired 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 difference between the high sustained limit and the low sustained limit for each natural gas-fired 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. 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 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.

Allowing offers up to these high levels 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, both NRG's and Calpine's VMPs contains a requirement that these offers, if offered in any hour of an operating day, must be offered in the same price/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 these two VMPs is the timing of termination. The approved VMPs for NRG and Calpine may each be terminated after three business days' notice. PURA §39.157(a) 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..." 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.

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 could be large.

The figure below shows the amount of surplus capacity available in each hour of every day during 2011, 2012, 2013 and 2014. 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 and regulation up capacity, minus load. Over the past four years there were 13 hours with no surplus capacity, with all but one hour occurring in 2011. These correspond to times when 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 81. There were 491 hours over the past four years with less than 4,000 MW of surplus capacity.<sup>17</sup> 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 VMPs granted to NRG and Calpine afford them the flexibility to raise their offers on a combined 1,200 MW of capacity. During the past four years this amount of capacity would have been pivotal in 61 hours.



<sup>&</sup>lt;sup>17</sup> Surplus capacity was less than 4000 MW for 296 hours in 2011, 154 hours in 2012, 15 hours in 2013, and 26 hours in 2014.

The effects of such actions became much more pronounced after June 21, 2013 when changes to real-time mitigation measures went into effect. These changes narrowed the scope of mitigation addressing the previously discussed issue where mitigation measures were being applied much more broadly than intended or necessary in the ERCOT real-time energy market.<sup>18</sup> Although "small fish" market participants have always been allowed to offer all of their capacity at prices up to the system-wide offer cap, the effect on market outcomes of a large "small fish" offering substantial quantities at high prices became more noticeable after the scope of mitigation was narrowed.

### **B.** Evaluation of Supplier Conduct

The previous subsection presented a structural analysis that supports inferences about potential market power. In this subsection actual participant conduct is evaluated to assess whether market participants have attempted to exercise market power through physical or economic withholding. First examined are unit deratings and forced outages to detect physical withholding, this is followed by an evaluation of "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. The strategy is profitable only if the withholding firm's incremental profit due to higher price is greater than the lost profit from the foregone sales of its withheld capacity.

### 1. Generation Outages and Deratings

Some portion of installed capability 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

<sup>&</sup>lt;sup>18</sup> Refer to Section I.F. Mitigation at page 22.

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 capability 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 capability level due to technical or environmental factors (e.g., component equipment failures or ambient temperature conditions). Wind generators rarely produce at their 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 this subsection, long-term and short-term deratings are evaluated.

Figure 82 shows a breakdown of total installed capability for ERCOT on a daily basis during 2014. 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.





Outages and deratings of non-wind generators fluctuated between 3 and 30 GW, as shown in Figure 82, while wind unavailability varied between 2 and 12 GW. Short-term planned outages were largest in March and April and small during the summer, which is consistent with expectations. Short-term forced outages also declined during the summer. Short-term deratings peaked during October.

The quantity of long-term (greater than 30 days) unavailable capacity, peaked in March at 3.8 GW, reduced to less than 1 GW during the summer months, and increased to almost 3.5 GW in November. This pattern reflects the choice by some owners to mothball certain generators on a seasonal basis, maintaining the units' operational status only during the high load summer season when more costly units have a higher likelihood of operating.

The next analysis focuses specifically on short-term planned and forced outages and deratings 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 83 shows the average magnitude of the outages and deratings lasting less than 30 days for the year and for each month during 2014.



Figure 83: Short-Term Outages and Deratings

Figure 83 shows that total short-term deratings and outages were as large as 13.7 percent of installed capacity in April, and averaged less than 5 percent during the summer. Most of this fluctuation was due to anticipated planned outages. The amount of capacity unavailable during 2014 averaged 7.8 percent of installed capacity. This is an increase from the 7.0 percent experienced in 2013, and the 5.0 and 6.0 percent experienced in 2012 and 2011. 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.
## 2. Evaluation of Potential Physical Withholding

Physical withholding occurs when a participant makes resources unavailable for dispatch that are otherwise physically capable of providing energy and that are economic at prevailing market prices. This can be done either by derating a unit or declaring it as forced out of service. Because generator deratings and forced outages are unavoidable, the goal of the analysis in this 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 79 and Figure 80 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 84 shows the average relationship of short-term deratings and forced outages as a percentage of total installed capacity to real-time load level for large and small suppliers. Portfolio size is important in determining whether individual suppliers have incentives to withhold available resources. Hence, the patterns of outages and deratings of large suppliers can be usefully evaluated by comparing them to the small suppliers' patterns.

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 84: Outages and Deratings by Load Level and Participant Size June to August, 2014

Figure 84 suggests that as demand for electricity increases, all market participants tend to make more capacity available to the market. For large suppliers, the combined short-term derating and forced outage rates decreased from 6.5 percent at low demand levels to approximately 3.2 percent at load levels above 65 GW. These are larger than for small suppliers at all load levels, which at first look may be seen as a competitive concern. However, large supplier outage rates are roughly the same as they were in 2013, whereas small supplier outage rates increased 68 percent. Given the overall low magnitude of outage rates for all suppliers, these results raise no competitiveness concerns.

## 3. Evaluation of Potential Economic Withholding

To complement the prior analysis of physical withholding, this subsection evaluates potential economic withholding by calculating an "output gap." The output gap is defined as the quantity of energy that is not being produced by in-service capacity even though the in-service capacity is

economic by a substantial margin given the real-time energy price. A participant can economically withhold resources, as measured by the output gap, by raising its energy offers so as not to be dispatched.

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

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

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



Figure 85: Incremental Output Gap by Load Level and Participant Size – Step 1

The results of the analysis shown in Figure 85 indicate barely detectable quantities of capacity being potentially economical withheld.

Figure 86 shows the ultimate output gap, measured by the difference between a unit's operating level and the output level had the unit been competitively offered to the market. In the second step of the dispatch, the after-mitigation offer curve is used to determine dispatch instructions and locational prices. As previously illustrated, even though the offer curve is mitigated there is still the potential for the mitigated offer curve to be increased as a result of a high first-step reference price being influenced by a market participant raising prices.

Similar to the previous analysis, Figure 86 shows very small quantities of capacity that would be considered part of this output gap.



Figure 86: Incremental Output Gap by Load Level and Participant Size – Step 2

The output gap of several of the largest suppliers were also examined for year 2014, and unlike the findings in 2013, found to be consistently low for all suppliers across all load levels.