2009 STATE OF THE MARKET REPORT
FOR THE
ERCOT WHOLESALE ELECTRICITY MARKETS

POTOMAC ECONOMICS, LTD.

Independent Market Monitor for the
ERCOT Wholesale Market

July 2010

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

A. Introduction

This report reviews and evaluates the outcomes of the ERCOT wholesale electricity markets in 2009, and is submitted to the Public Utility Commission of Texas ("PUCT") and the Electric Reliability Council of Texas ("ERCOT") pursuant to the requirement in Section 39.1515(h) of the Public Utility Regulatory Act. It includes assessments of the incentives provided by the current market rules and procedures, and analyses of the conduct of market participants. This report also assesses the effectiveness of the scarcity pricing mechanism pursuant to the provisions of PUCT Substantive Rule 25.505(g). Key findings in the report include the following:

★ The average wholesale electricity price was $34.03 per MWh in 2009, which is 56 percent lower than the 2008 average price of $77.19 per MWh. This is the lowest annual average price experienced in the ERCOT wholesale market since 2002.

★ All-in wholesale electricity prices for the ERCOT market in 2009 were lower than in the organized wholesale electricity markets in California, New England, the New York ISO, and the PJM Interconnection.

★ Lower wholesale electricity prices provide benefits to consumers in the short-term. However, pricing outcomes in 2009 continued to inadequately reflect market conditions during times of operating reserve scarcity. During such shortage conditions when demand for energy and operating reserves cannot be met with available resources, prices should rise sharply to reflect the value of diminished reliability as reserves are used to meet energy needs. Although these shortage conditions occur in only a handful of hours each year, efficient shortage pricing is critical to the long-term success of the ERCOT energy-only market.

★ As a result of inadequate shortage pricing and the fact that the number of shortage intervals in 2009 were roughly one-half of that experienced in 2008, estimated net revenues in 2009 were substantially below the levels required to support market entry for natural gas combined-cycle and combustion turbine resources at all locations in the ERCOT region. Estimated net revenues for nuclear and coal resources were also insufficient to support new entry in 2009, although these results were more affected by the reduction in natural gas prices and associated reduction in wholesale energy prices than by pricing outcomes during shortage conditions.

★ Ancillary service costs generally track wholesale energy price movements, and therefore were significantly lower in 2009 than in recent years.

★ Load participation in the responsive reserve market declined in late 2008 and in 2009 relative to prior years, likely as a result of general economic conditions.

★ Interzonal price disparities were larger in 2008 and 2009 than in prior years, primarily as a result of increased wind capacity in the West Zone and inefficiencies that are inherent to the zonal market design.

★ The number of hours in which coal was the marginal (*i.e.*, price-setting) fuel in the ERCOT region was much higher in 2009 than in prior years. This increase can be attributed to (1) increased wind resource production; (2) a slight reduction in demand in 2009 due to the economic downturn; and (3) periods when natural gas prices were very low thereby making coal and natural gas combined-cycle resources competitive from an economic dispatch standpoint.

★ The ERCOT wholesale market performed competitively in 2009, with the competitive performance measures showing a trend of increasing competitiveness over the period 2005 through 2009.

In addition to these key findings, the report generally confirms prior findings that the current market rules and procedures are resulting in systemic inefficiencies. Our previous reports regarding ERCOT electricity markets have included a number of recommendations designed to improve the performance of the current ERCOT markets.[1] Some of these recommendations have

[1] "ERCOT State of the Market Report 2003", Potomac Economics, August 2004 ("2003 SOM Report"); "2004 Assessment of the Operation of the ERCOT Wholesale Electricity Markets", Potomac Economics, November 2004; "ERCOT State of the Market Report 2004", Potomac Economics, July 2005 ("2004 SOM Report"); "ERCOT

been implemented. Given the approaching implementation of the nodal market design in December 2010, no additional recommendations for the current market design are offered at this time. In particular, implementation of the nodal market will provide the following improvements:

- The nodal market design will fundamentally improve ERCOT's ability to efficiently manage transmission congestion, which is one of the most important functions in electricity markets.
- The wholesale market should function more efficiently under the nodal market design by providing better incentives to market participants, facilitating more efficient commitment and dispatch of generation, and improving ERCOT's operational control of the system. The congestion on all transmission paths and facilities will be managed through market-based mechanisms in the nodal market. In contrast, under the current zonal market design, transmission congestion is most frequently resolved through non-transparent, non-market-based procedures.
- Under the nodal market, unit-specific dispatch will allow ERCOT to more fully utilize generating resources than the current market, which frequently exhibits price spikes even when generating capacity is not fully utilized.
- The nodal market will allow ERCOT to increase the economic and reliable utilization of scarce transmission resources well beyond that attainable in the zonal market.
- The nodal market will significantly improve the ability to efficiently and reliably integrate the ever-growing quantities of intermittent resources, such as wind and solar generating facilities.
- The nodal market will produce price signals that better indicate where new generation is most needed (and where it is not) for managing congestion and maintaining reliability.

State of the Market Report 2005", Potomac Economics, July 2006 ("2005 SOM Report"); "ERCOT State of the Market Report 2006", Potomac Economics, August 2007 ("2006 SOM Report"), "ERCOT State of the Market Report 2007", Potomac Economics, August 2008 ("2007 SOM Report"); and "ERCOT State of the Market Report 2008", Potomac Economics, August 2009 ("2008 SOM Report").

In the long-term, these enhancements to overall market efficiency should translate into substantial savings for consumers.

B. Review of Market Outcomes

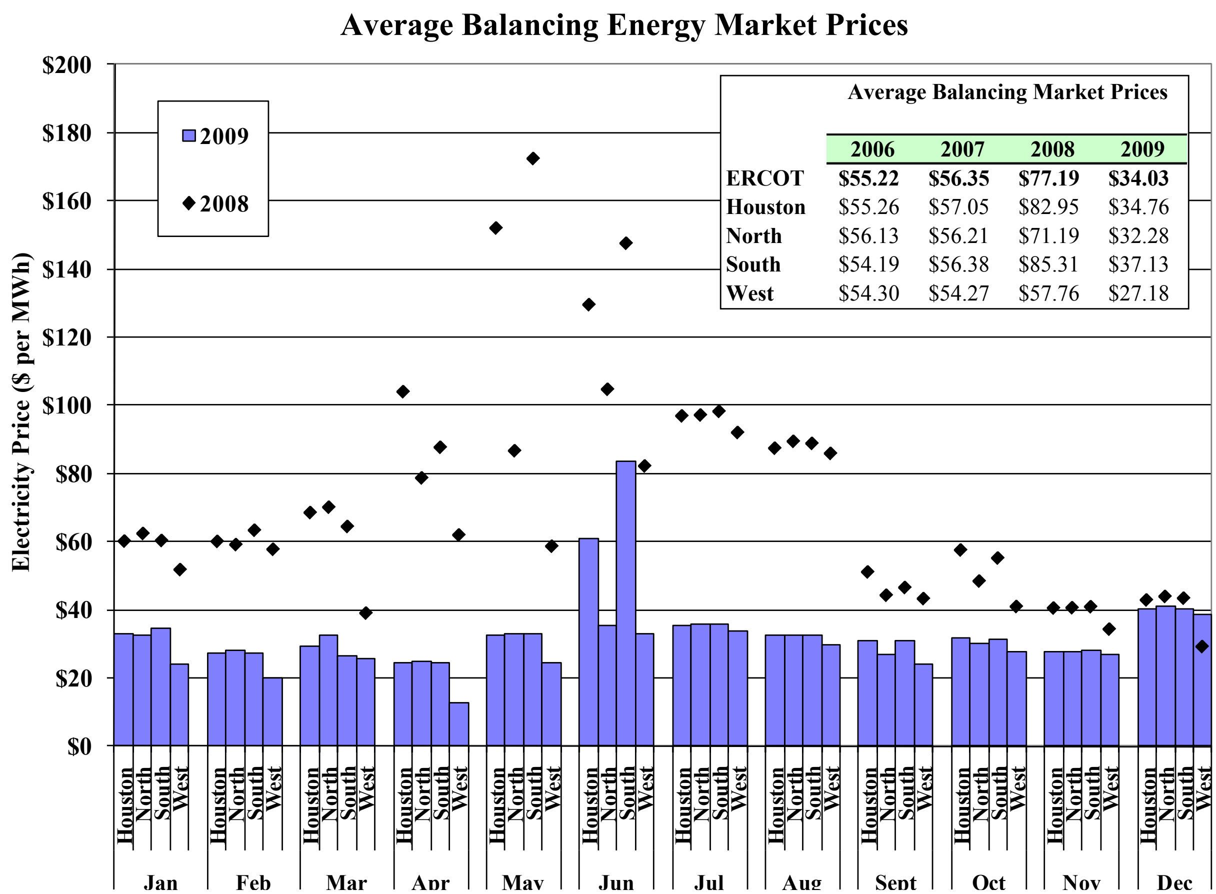
1. Balancing Energy Prices

The balancing energy market allows participants to make real-time purchases and sales of energy to supplement their forward bilateral contracts. While on average only a relatively small portion of the electricity produced in ERCOT is cleared through the balancing energy market, its role is critical in the overall wholesale market. The balancing energy market governs real-time dispatch of generation by altering where energy is produced to: a) balance supply and demand; b) manage interzonal congestion, and c) displace higher-cost energy with lower-cost energy given the energy offers of the Qualify Scheduling Entities ("QSEs").

In addition, the balancing energy prices also provide a vital signal of the value of power for market participants entering into forward contracts. Although most power is purchased through forward contracts of varying duration, the spot prices emerging from the balancing energy market should directly affect forward contract prices.

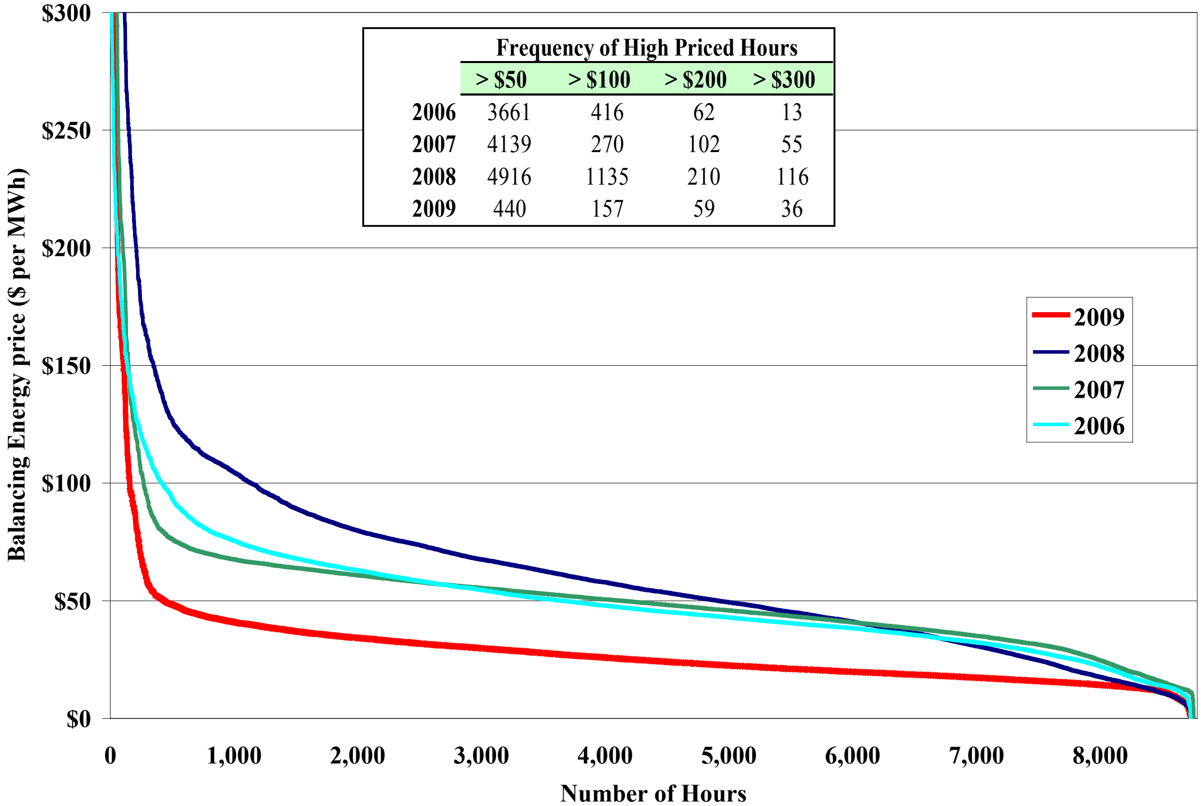
As shown in the following figure, ERCOT average balancing energy market prices were 56 percent lower in 2009 than in 2008, with an ERCOT-wide load weighted average price of $34.03 per MWh in 2009 compared to $77.19 per MWh in 2008. April through August experienced the highest balancing energy market price reductions in 2009, averaging 66 percent lower than the prices in the same months in 2008. With the exception of the West Zone in December, the balancing energy prices in 2009 were lower in every month in all zones than in 2008.

The average natural gas price fell 56 percent in 2009, averaging $3.74 per MMBtu in 2009 compared to $8.50 per MMBtu in 2008. Natural gas prices reached a maximum monthly average of $12.37 per MMBtu in July 2008, and reached a minimum monthly average of $2.93 per MMBtu in September 2009. Hence, the changes in energy prices from 2008 to 2009 were largely a result of natural gas price movements.



The following figure shows the price duration curves for the ERCOT balancing energy market each year from 2006 to 2009. 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 hourly load-weighted average prices for the ERCOT balancing energy market.

ERCOT Price Duration Curve

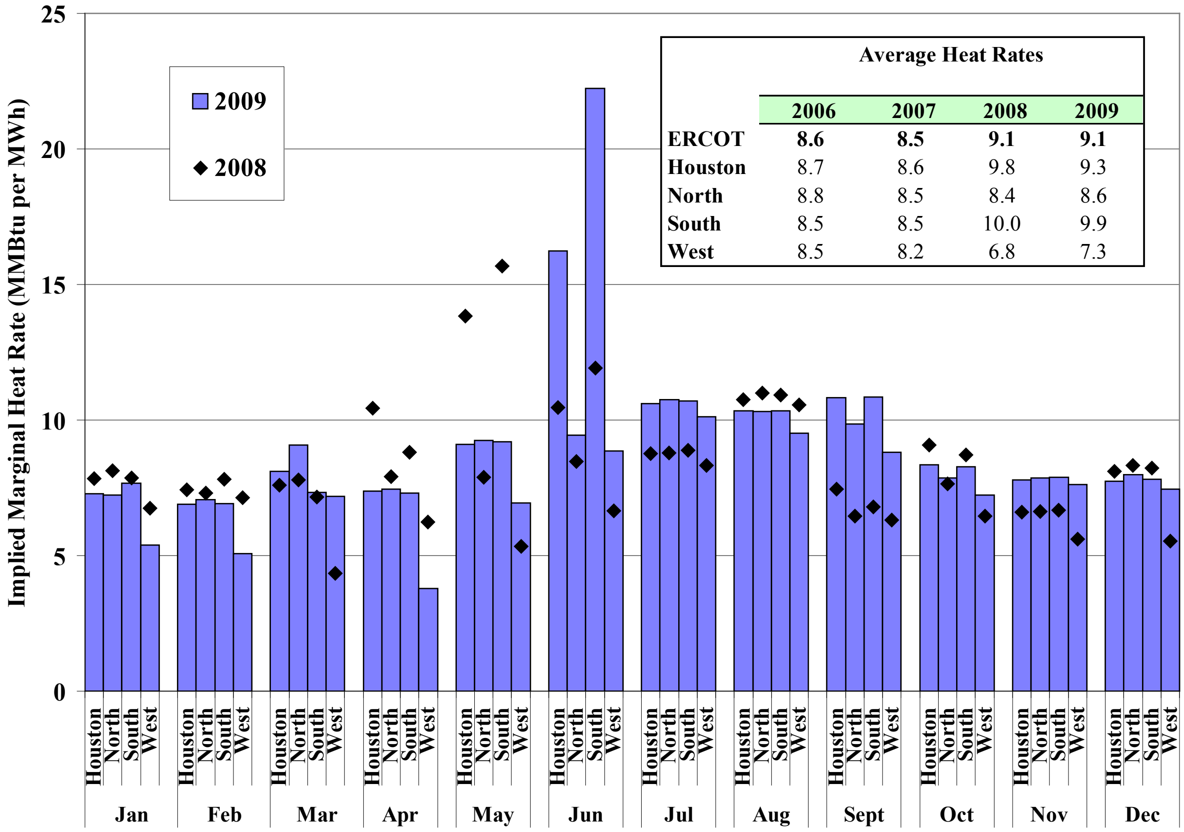


	Frequency of High Priced Hours			
	> $50	> $100	> $200	> $300
2006	3661	416	62	13
2007	4139	270	102	55
2008	4916	1135	210	116
2009	440	157	59	36

Balancing energy prices exceeded $50 per MWh in 440 hours in 2009 compared to more than 4,900 hours in 2008. These year-to-year changes reflect lower natural gas prices in 2009 that directly affect electricity prices in a broad range of hours.

Although fuel price fluctuations are the dominant factor driving electricity prices in the ERCOT wholesale market, fuel prices alone do not explain all of the price outcomes. The following figure presents ERCOT balancing energy market prices adjusted for natural gas price fluctuations to better highlight variations in electricity prices not related to fuel costs.

Monthly Average Implied Marginal Heat Rate



Adjusted for gas price influence, the above figure shows that average implied heat rate for all hours of the year was comparable in 2009 to 2008.[2] The average implied heat rate was significantly higher in 2008 than in 2009 during the months of April and May due to significant zonal congestion on the North to South and North to Houston interfaces that materialized in these months in 2008. Similarly, the magnitude of zonal congestion on the North to South interface increased significantly in late June 2009, causing the implied heat rate in June to be significantly higher in 2009 than in 2008. The implied heat rate in July was higher in 2009 than in 2008, primarily because of a stretch of extremely high temperatures and load levels, including the setting of a new record peak demand of 63,400 MW on July 13, 2009. Finally, the implied heat rate in September was much lower in 2008 than in 2009 because of the landfall of Hurricane Ike in September 2008 that resulted in widespread and prolonged loss of load in the Houston area.

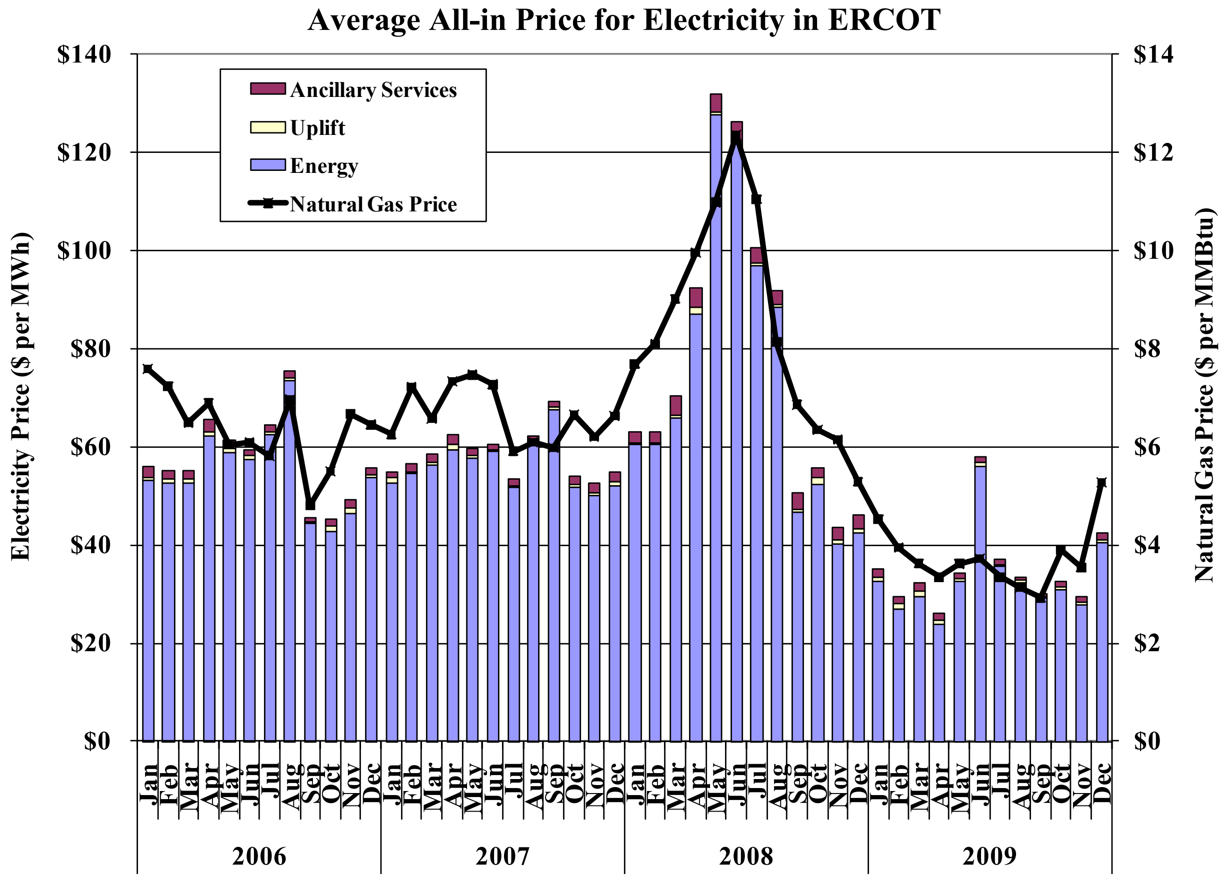
2 The *Implied Marginal Heat Rate* equals the *Balancing Energy Market Price* divided by the *Natural Gas Price.*

The report evaluates two other aspects of the balancing energy prices: 1) the correlation of the balancing energy prices with forward electricity prices in Texas, and 2) the primary determinants of balancing energy prices. Natural market forces should push forward market prices to levels consistent with expectations of spot market prices. Day-ahead prices averaged $38 per MWh in 2008 compared to an average of $35 per MWh for real-time prices. Although the day-ahead and real-time prices exhibited relatively good average convergence in 2009, the average absolute price difference increased during the months of June and July 2009.

The price volatility in June 2009 was due in large part to the significant and unpredictable transmission congestion experienced in that timeframe that caused average real-time prices to exceed day-ahead prices in June 2009. In contrast, average day-ahead prices were significantly higher than real-time prices in July 2009, which may be associated with transmission congestion expectations based on the experience in the prior month, as well as real-time pricing expectations associated with the extremely high temperatures and loads experienced during July 2009. The introduction of the nodal market, which will include an integrated day-ahead market, should also improve the convergence between day-ahead and real-time energy prices.

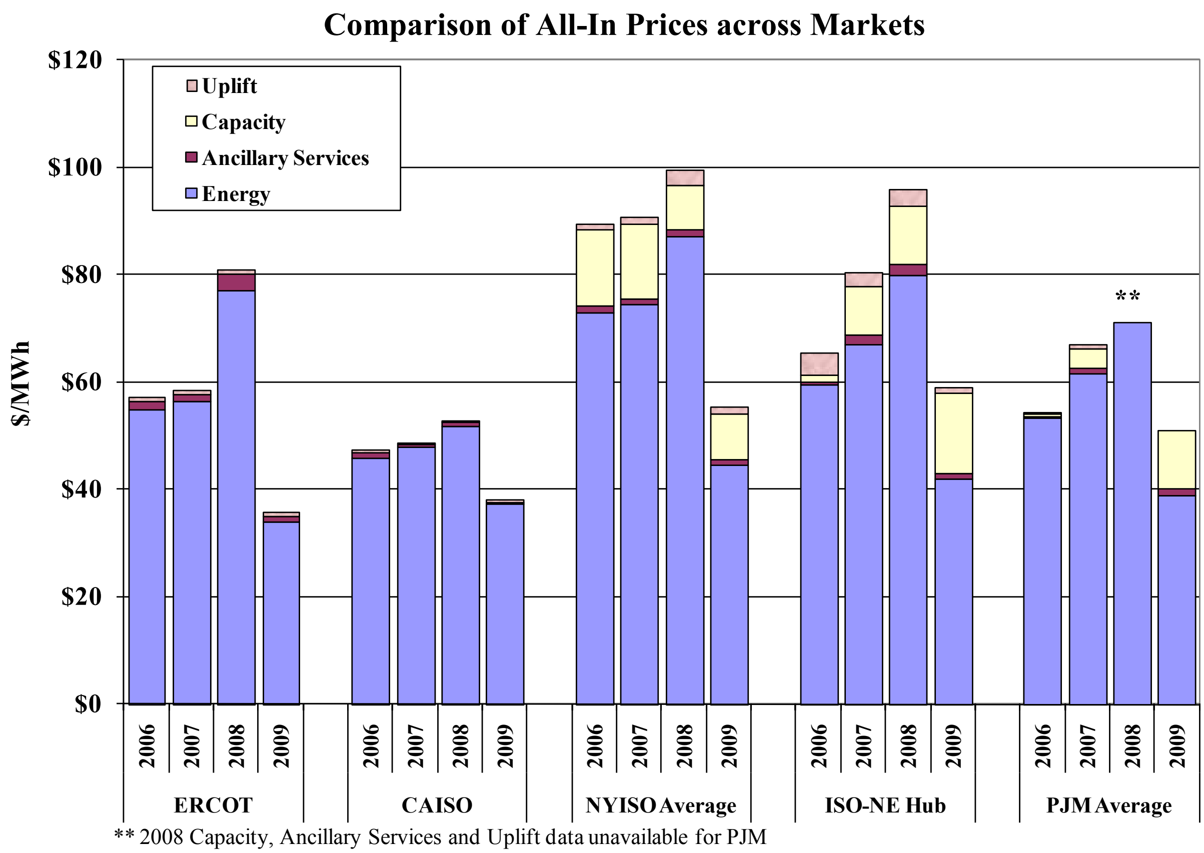
2. All-In Electricity Prices

In addition to the costs of energy, loads incur costs associated with operating reserves, regulation, and uplift. The uplift costs include payments for out-of-merit capacity (“OOMC”), Replacement Reserve (“RPRS”), out-of-merit energy (“OOME”), and reliability must run agreements (“RMR”), but exclude administrative charges such as the ERCOT fee. These costs, regardless of the location of the congestion, are borne proportionally by all loads within ERCOT.



The monthly average all-in energy prices for the past four years are shown in the figure above along with the monthly average price of natural gas. This figure indicates that natural gas prices were the primary driver of the trends in electricity prices from 2006 to 2009. Average natural gas prices decreased in 2009 by 56 percent from 2008 levels. The average all-in price for electricity was $80.97 in 2008 and $35.09 in 2009, a decrease of 56 percent.

To provide additional perspective on the outcomes in the ERCOT market, our next analysis compares the all-in price metrics for ERCOT and other electricity markets. The following figure compares the all-in prices in ERCOT with other organized electricity markets in the U.S.: California ISO, New York ISO, ISO New England, and PJM. For each region, the figure reports the average cost (per MWh of load) for energy, ancillary services (reserves and regulation), capacity markets (if applicable), and uplift for economically out-of-merit resources.



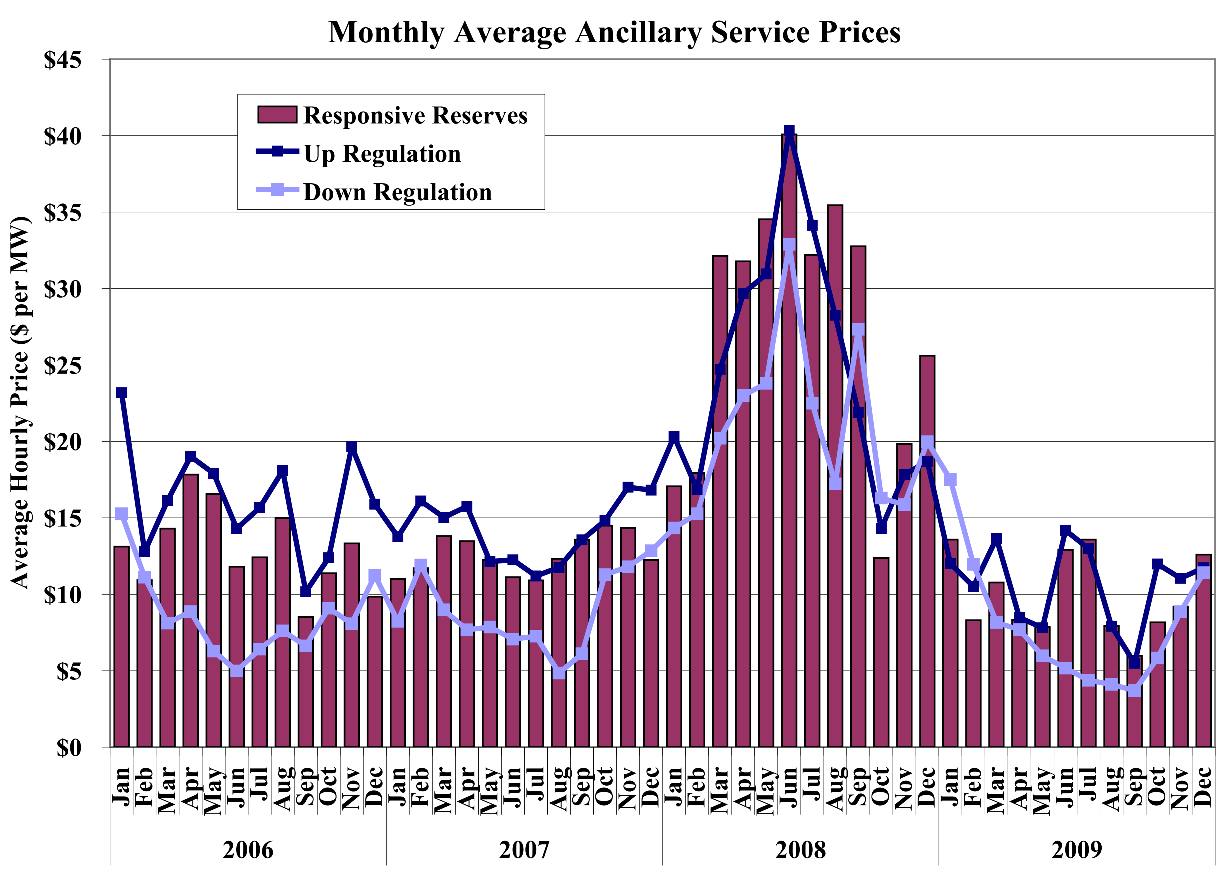
This figure shows that energy prices decreased in wholesale electricity markets across the U.S. in 2009, primarily due to decreases in fuel costs, and that the ERCOT market experienced the lowest all-in wholesale prices of any of these markets in 2009.

3. Ancillary Services Markets

The primary ancillary services are up regulation, down regulation, and responsive reserves. Historically, ERCOT has also procured non-spinning reserves as needed during periods of increased supply and demand uncertainty. However, beginning in November 2008, ERCOT began procuring non-spinning reserves across all hours based on its assessment of "net load" error, where "net load" is equal to demand minus wind production. QSEs may self-schedule ancillary services or purchase their required ancillary services through the ERCOT markets. This section reviews the results of the ancillary services markets in 2009.

Because ancillary services markets are conducted prior to the balancing energy market, participants must include their expected value of foregone sales in the balancing energy market in their offers for responsive reserves and regulation. Providers of both responsive reserves and up regulation can incur such opportunity costs if they reduce the output from economic units to make the capability available to provide these services. Likewise, providers of down regulation can incur opportunity costs in real-time if they receive instructions to reduce their output below the most profitable operating level. Further, because generators must be online to provide regulation and responsive reserves, there is an economic risk during low price periods of operating uneconomically at minimum output levels (or having to operate above minimum output levels if providing down regulation). The figure below shows the monthly average prices for regulation and responsive reserve services from 2006 to 2009.

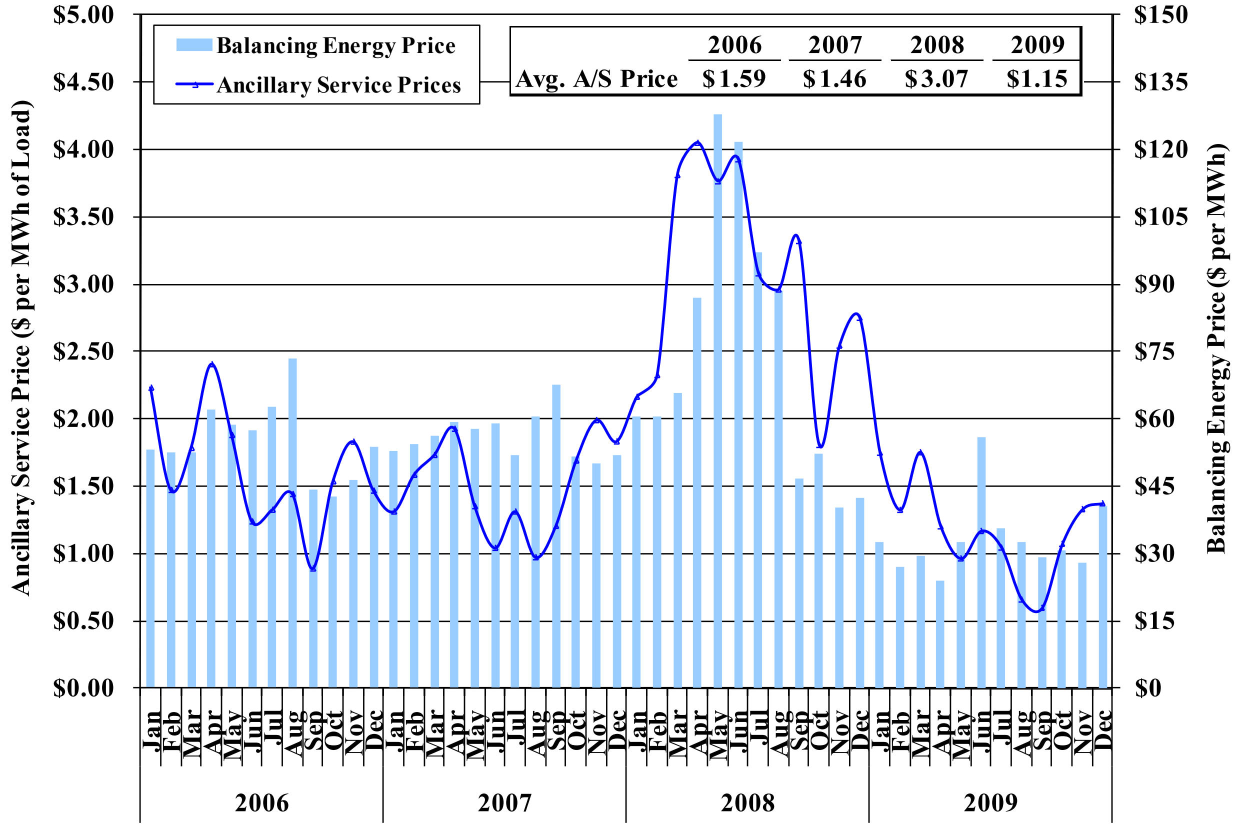
This figure shows that ancillary service capacity prices generally returned to levels seen in 2006 and 2007 after reaching significantly higher levels in 2008. These price movements can be primarily attributed to the variations in energy prices that occurred over the same timeframe.



The current Nodal Protocols specify that energy and ancillary services will be jointly optimized in a centralized day-ahead market. This is likely to improve the overall efficiency of the day-ahead unit commitment. Additionally, although not possible to implement at the inception in the nodal market, we also recommend the development of real-time markets that co-optimize energy and reserves to further enhance the efficient dispatch of resources and pricing in real-time.

While the previous figure shows the individual ancillary service capacity prices, the following figure shows the monthly total ancillary service costs per MWh of ERCOT load and the average balancing energy price for 2006 through 2009.

Ancillary Service Costs per MWh of Load



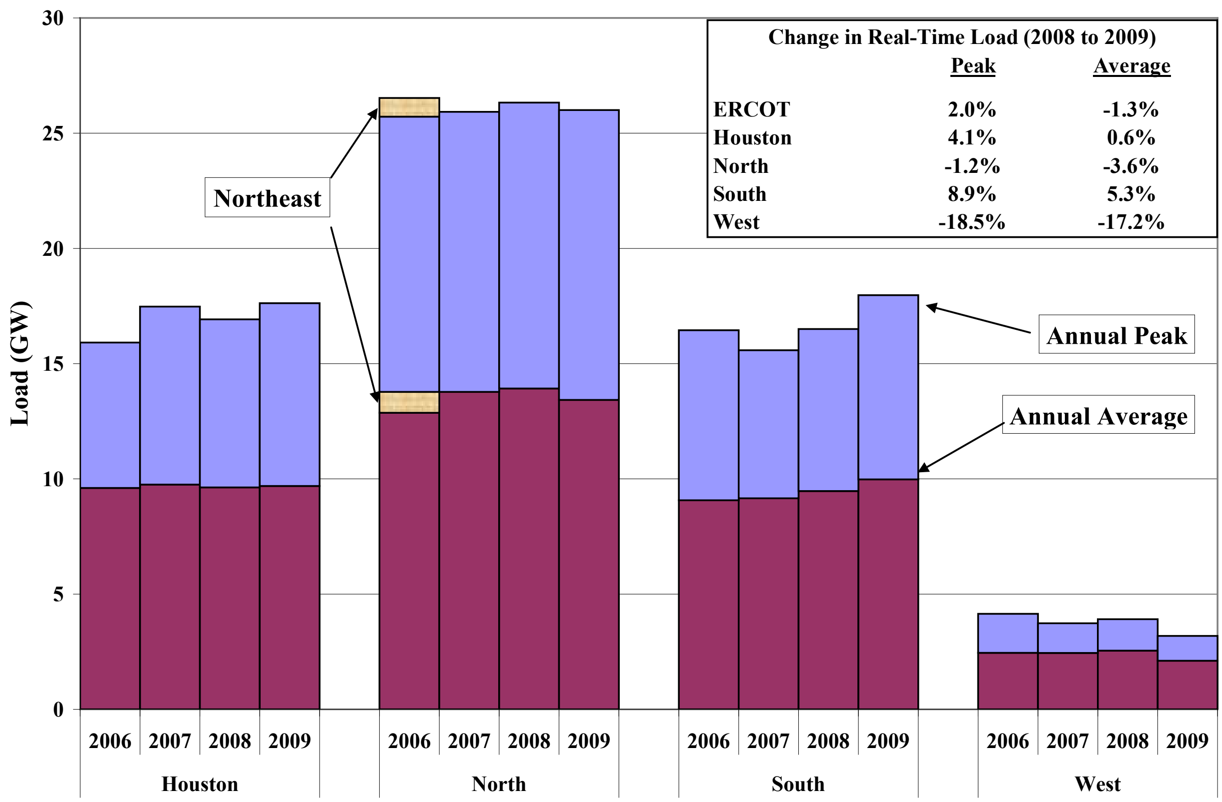
This figure shows that total ancillary service costs are generally correlated with balancing energy price movements which, as previously discussed, are highly correlated with natural gas price movements. The average ancillary service cost per MWh of load decreased to $1.15 per MWh in 2009 compared to $3.07 per MWh in 2008, a decrease of more than 63 percent. Ancillary service costs were equal to 4.0 and 3.5 percent of the load-weighted average energy price in 2008 and 2009, respectively.

C. Demand and Resource Adequacy

1. ERCOT Loads in 2009

This section examines changes in average and peak load levels in 2009 of these dimensions of load during 2009. The following figure shows peak load and average load in each of the ERCOT zones from 2006 to 2009. This figure indicates that in each zone, as in most electrical systems, peak demand significantly exceeds average demand. The North Zone is the largest zone (about 38 percent of the total ERCOT load);[3] the South and Houston Zones are comparable (with about 28 percent) while the West Zone is the smallest (with about 6 percent of the total ERCOT load).

Annual Load Statistics by Zone



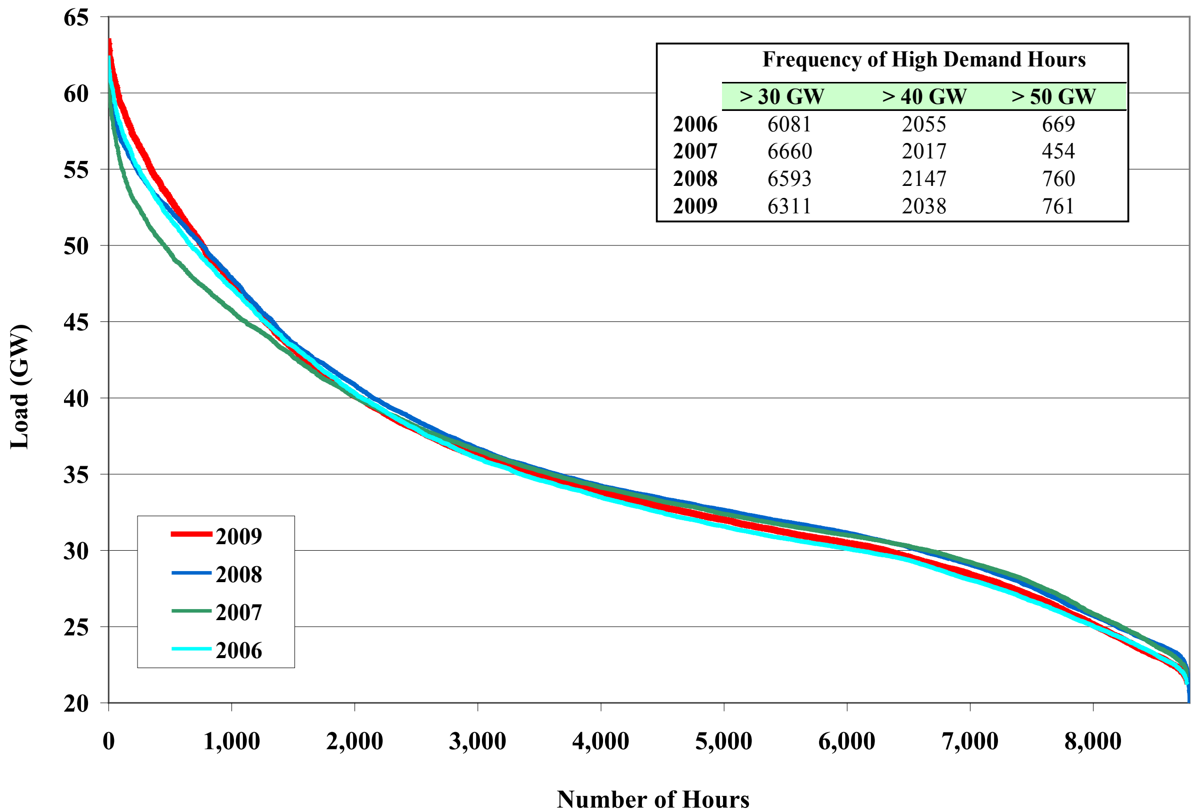
Some of the changes in zonal peak and average loads from 2008 to 2009 can be attributed to changes to the zonal definitions that resulted in some loads moving to a different zone in 2009. Overall, the ERCOT average load decreased from 312,401 GWh in 2008 to 308,278 GWh in 2009, a decrease of 1.3 percent. In contrast, the ERCOT coincident peak demand increased from 62,174 MW in 2008 to 63,400 MW in 2009, an increase of 2.0 percent.

---

3 The Northeast Zone was integrated into the North Zone in 2007.

To provide a more detailed analysis of load at the hourly level, the next figure compares load duration curves for each year from 2006 to 2009. 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, as most hours exhibit low to moderate electricity demand, with peak demand usually occurring during the afternoon and early evening hours of days with exceptionally high temperatures.

ERCOT Load Duration Curve – All Hours

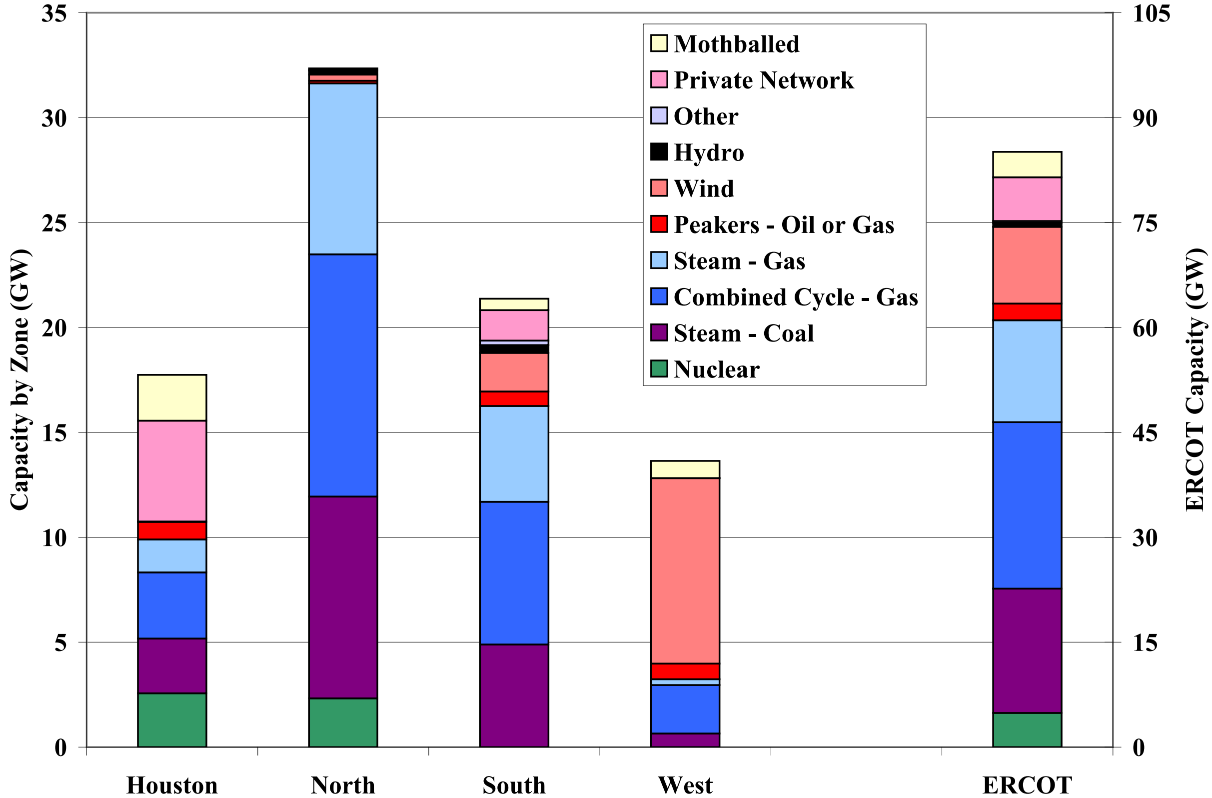


As shown in the figure above, the load duration curve for 2009 is slightly lower than in 2008 at load levels less than 45 GW, which accounts for approximately 85 percent of the hours in 2009 and is consistent with the load reduction of 1.3 percent from 2008 to 2009. However, the number of high demand hours (more than 50 GW) in 2008 and 2009 are at comparable levels (760 and 761 hours respectively). Load exceeded 58 GW in 160 hours in 2009, more than double the hours in 2008.

2. Generation Capacity in ERCOT

This section evaluates the generation mix in ERCOT. With the exception of the wind resources in the West Zone and the nuclear resources in the North and Houston Zones, the mix of generating capacity is relatively uniform in ERCOT. The following figure shows the installed generating capacity by type in each of the ERCOT zones.

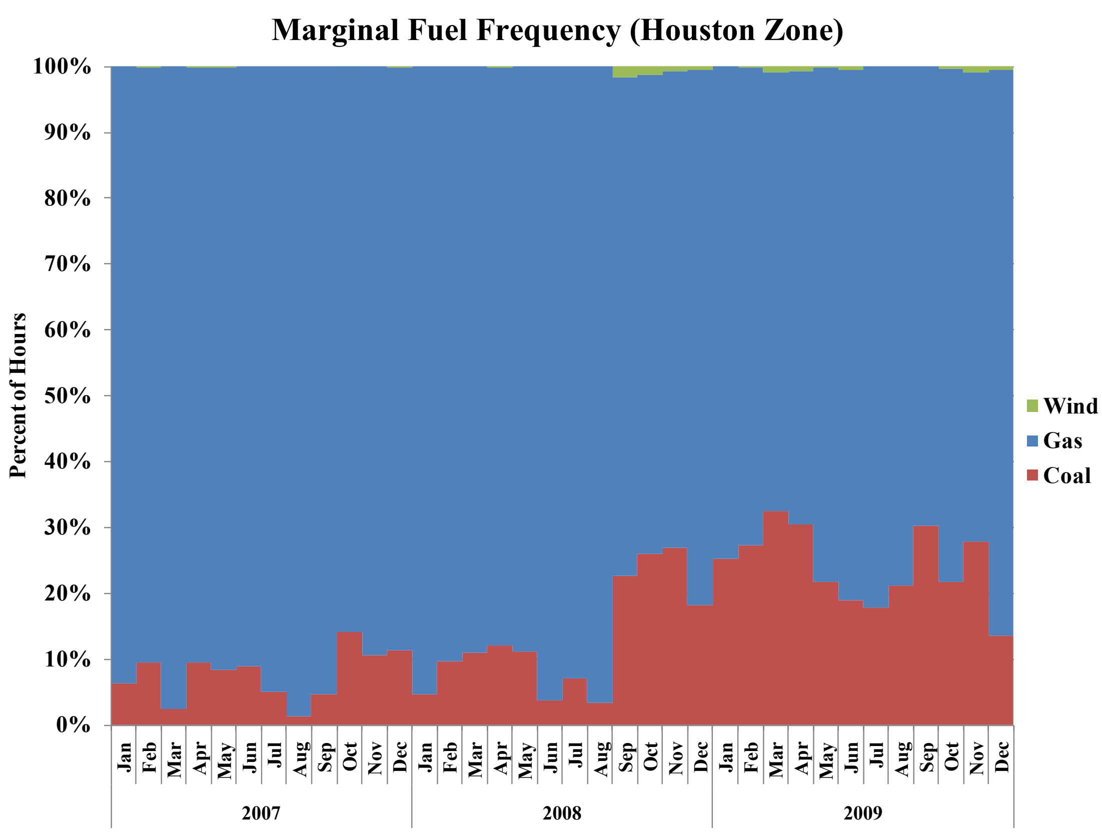
Installed Capacity by Technology for each Zone



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.

While ERCOT has coal/lignite and nuclear plants that operate primarily as base load units, its reliance on natural gas resources makes it vulnerable to natural gas price spikes. There is approximately 22.6 GW of coal and nuclear generation in ERCOT. Because there are very few hours when ERCOT load drops as low as 20 GW, natural gas resources will be dispatched and set the balancing energy spot price in most hours. Hence, although coal-fired and nuclear units combined produce approximately half of the energy in ERCOT, they have historically played a much less significant role in setting spot electricity prices. However, with the significant increases in wind capacity that has a lower marginal production cost than coal and lignite, the frequency at which coal and lignite are the marginal units in ERCOT is expected to increase in the future, particularly during the off-peak hours in the spring and fall, and even more as additional transmission capacity is added that will accommodate increased levels of wind production in the West Zone.

The figure below shows the marginal fuel frequency for the Houston Zone, for each month from 2007 through 2009. The marginal fuel frequency is the percentage of hours that a generation fuel type is marginal and setting the price at a particular location.



As shown in the figure above, the frequency at which coal was the price setting fuel for the Houston Zone experienced a significant and sustained increase beginning in September 2008. This increase can be attributed to (1) increased wind resource production; (2) a slight reduction in demand in 2009 due to the economic downturn; and (3) periods when natural gas prices were very low thereby making coal and combined-cycle natural gas resources competitive from an economic dispatch standpoint. As significant additional wind, coal and potentially nuclear resources are added to the ERCOT region and transmission constraints that serve to limit existing wind production are alleviated, it is likely that the marginal fuel frequency of coal will increase in coming years.

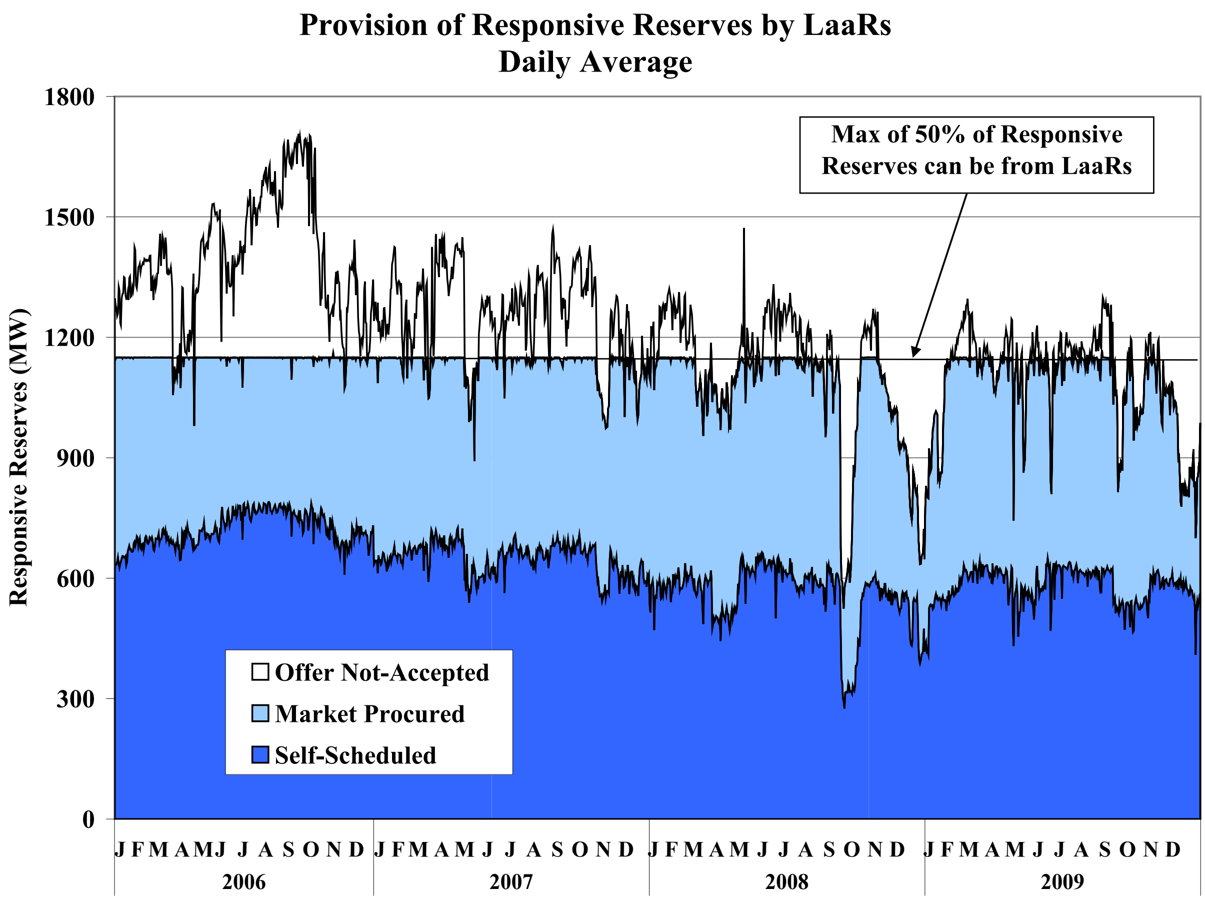
3. Load Participation in the ERCOT Markets

The ERCOT Protocols allow for loads to participate in the ERCOT-administered markets as either Load acting as Resources ("LaaRs") or Balancing Up Loads ("BULs"). LaaRs are loads that are qualified by ERCOT to offer responsive reserves, non-spinning reserves, or regulation into the day-ahead ancillary services markets and can also offer blocks of energy in the balancing energy market.

As of December 2009, over 2,200 MW of capability were qualified as LaaRs. In 2009, LaaRs were permitted to supply up to 1,150 MW of the responsive reserves requirement. Although the participants with LaaR resources are qualified to provide non-spinning reserves and up balancing energy in real-time, LaaR participation in the non-spinning reserve and balancing energy market was negligible in 2009.[4] This is not surprising because the value of curtailed load tends to be relatively high, and providing responsive reserves offers substantial revenue with very little probability of being deployed. In contrast, resources providing non-spinning reserves have a much higher probability of being curtailed. Hence, most LaaRs will have a strong preference to provide responsive reserves over non-spinning reserves or balancing energy. The following figure shows the daily average provision of responsive reserves by LaaRs in the ERCOT market from 2006 through 2009.

---

4 Although there was no active participation in the balancing energy market, loads can and do respond to market prices without actively submitting a bid to ERCOT. This is often referred to as passive load response.



The high level of participation by demand response participating in the ancillary service markets sets ERCOT apart from other operating electricity markets. The figure above shows that the amount of responsive reserves provided by LaaRs has held fairly constant at 1,150 MW since the beginning of 2006. Exceptions include a decrease in September of 2008 corresponding to the Texas landfall of Hurricane Ike and a more prolonged reduction from November 2008 through January 2009 that was likely a product of the economic downturn and its effect on industrial operations.

4. Net Revenue Analysis

Net revenue is defined as the total revenue that can be earned by a new generating unit less its variable production costs. It represents the revenue that is available to recover a unit's fixed and capital costs. Hence, this metric shows the economic signals provided by the market for investors to build new generation or for existing owners to retire generation. In long-run equilibrium, the markets should provide sufficient net revenue to allow an investor to break-even on an investment in a new generating unit, including a return of and on the investment.

In the short-run, if the net revenues produced by the market are not sufficient to justify entry, then one of three conditions likely exists:

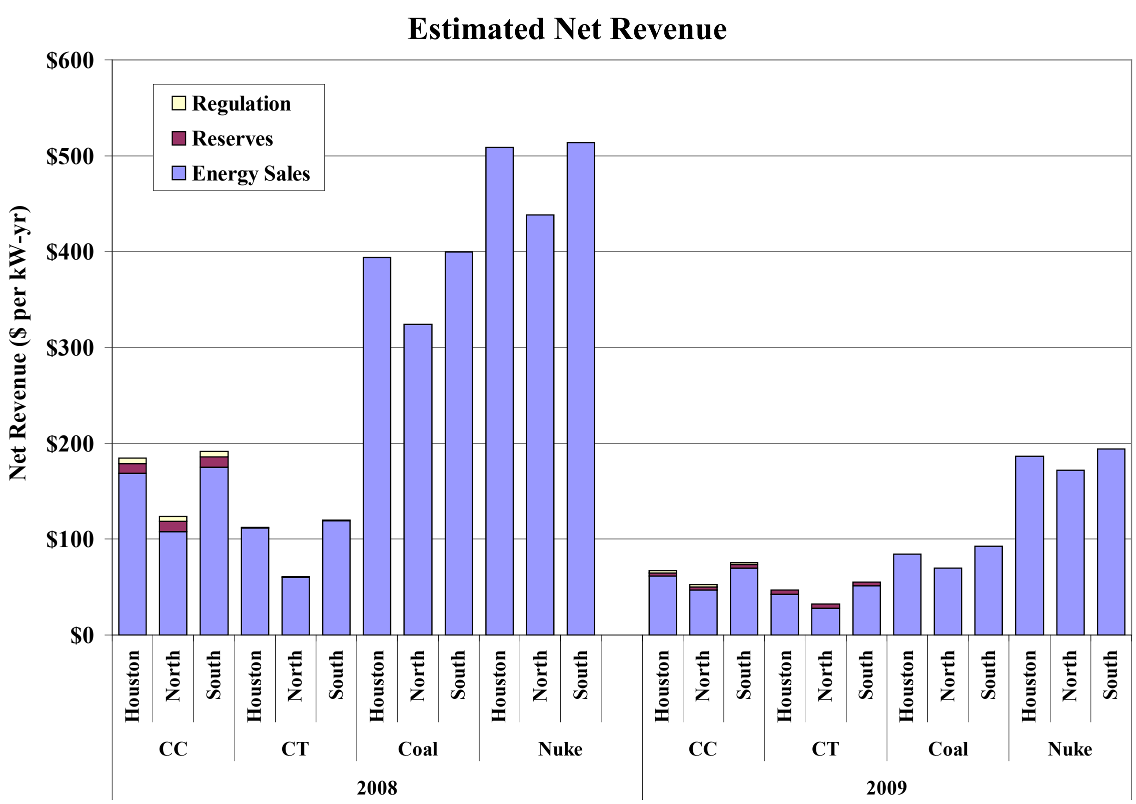
(i) New capacity is not currently needed because there is sufficient generation already available;

(ii) Load levels, and thus energy prices, are temporarily low due to mild weather or economic conditions; or

(iii) Market rules are causing revenues to be reduced inefficiently.

Likewise, the opposite would be true if the markets provide excessive net revenue in the short-run. Excessive net revenue that persists for an extended period in the presence of a capacity surplus is an indication of competitive issues or market design flaws.

The report estimates the net revenue that would have been received in 2008 and 2009 for four types of units: a natural gas combined-cycle generator, a simple-cycle gas turbine, a coal unit, and a nuclear unit.



The figure above shows that the net revenue decreased substantially in 2009 compared to each zone compared in 2008. Based on our estimates of investment costs for new units, the net revenue required to satisfy the annual fixed costs (including capital carrying costs) of a new gas turbine unit ranges from $70 to $95 per kW-year. The estimated net revenue in 2009 for a new gas turbine was approximately $55, $47 and $32 per kW-year in the South, Houston and North Zones, respectively. For a new combined cycle unit, the estimated net revenue requirement is approximately $105 to $135 per kW-year. The estimated net revenue in 2009 for a new combined cycle unit was approximately $76, $67 and $52 per kW-year in the South, Houston and North Zones, respectively. These values indicate that the estimated net revenue in 2009 was well below the levels required to support new entry for a new gas turbine or a combined cycle unit in the ERCOT region. Prior to 2005, net revenues were well below the levels necessary to justify new investment in coal and nuclear generation. However, high natural gas prices through 2008 allowed energy prices to remain at levels high enough to support new entry for these technologies. The production costs of coal and nuclear units did not change significantly over this period, leading to a dramatic rise in net revenues. With the significant decline in natural gas and energy prices in 2009, these results changed dramatically from recent years. For a new coal unit, the estimated net revenue requirement is approximately $190 to $245 per kW-year. The estimated net revenue in 2009 for a new coal unit was approximately $93, $84 and $70 per kW-year in the South, Houston and North Zones, respectively. For a new nuclear unit, the estimated net revenue requirement is approximately $280 to $390 per kW-year. The estimated net revenue in 2009 for a new nuclear unit was approximately $194, $187 and $172 per kW-year in the South, Houston and North Zones, respectively. These values indicate that the estimated net revenue for a new coal and nuclear unit in the South, Houston and North Zones was well below the levels required to support new entry in 2009.

5. Effectiveness of the Scarcity Pricing Mechanism

The PUCT adopted rules in 2006 that define the parameters of an energy-only market. These rules include a Scarcity Pricing Mechanism ("SPM") that relaxed the existing system-wide offer cap by gradually increasing it to $1,500 per MWh on March 1, 2007, $2,250 per MWh on March 1, 2008, and to $3,000 per MWh shortly after the implementation of the nodal market. Additionally, market participants controlling less than five percent of the capacity in ERCOT by definition do not possess market power under the PUCT rules. Hence, these participants can submit very high-priced offers that, per the PUCT rule, will not be deemed to be an exercise of market power. However, because of the competition faced by the smaller market participants, the quantity offered at such high prices – if any – is very small.

Unlike markets with a long-term capacity market where fixed capacity payments are made to resources across the entire year regardless of the relationship of supply and demand, the objective of the energy-only market design is to allow energy prices to rise significantly higher during legitimate shortage conditions (*i.e.*, when the available supply is insufficient to simultaneously meet both energy and operating reserve requirements) such that the appropriate price signal is provided for demand response and new investment when required. During non-shortage conditions (*i.e.*, most of the time), the expectation of competitive energy market outcomes is no different in energy-only than in capacity markets.

Consistent with the previous findings relating to net revenue, the PNM reached the level sufficient for new entry in only one of the last four years (2008). In 2008, the peaker net margin and net revenue values rose substantially, surpassing the level required to support new peaker entry. However, a significant portion of the net revenue increase in 2008 was associated with extremely inefficient transmission congestion management and inefficient pricing mechanisms associated with the deployment of non-spinning reserves.[5] Both of these issues were corrected in the zonal market and will be further improved with the implementation of the nodal market in late 2010. With these issues addressed, the peaker net margin dropped substantially in 2009, decreasing to $46,650 per MW-yr from $101,774 per MW-yr in 2008. Net revenues also dropped substantially for other technologies largely due to significant decreases in natural gas prices in 2009, but decreased natural gas price are not the driver for the reduction in net revenues for peaking resources. Beyond the correction of the market design inefficiencies that existed in 2008, there were three other factors that influenced the effectiveness of the SPM in 2009:

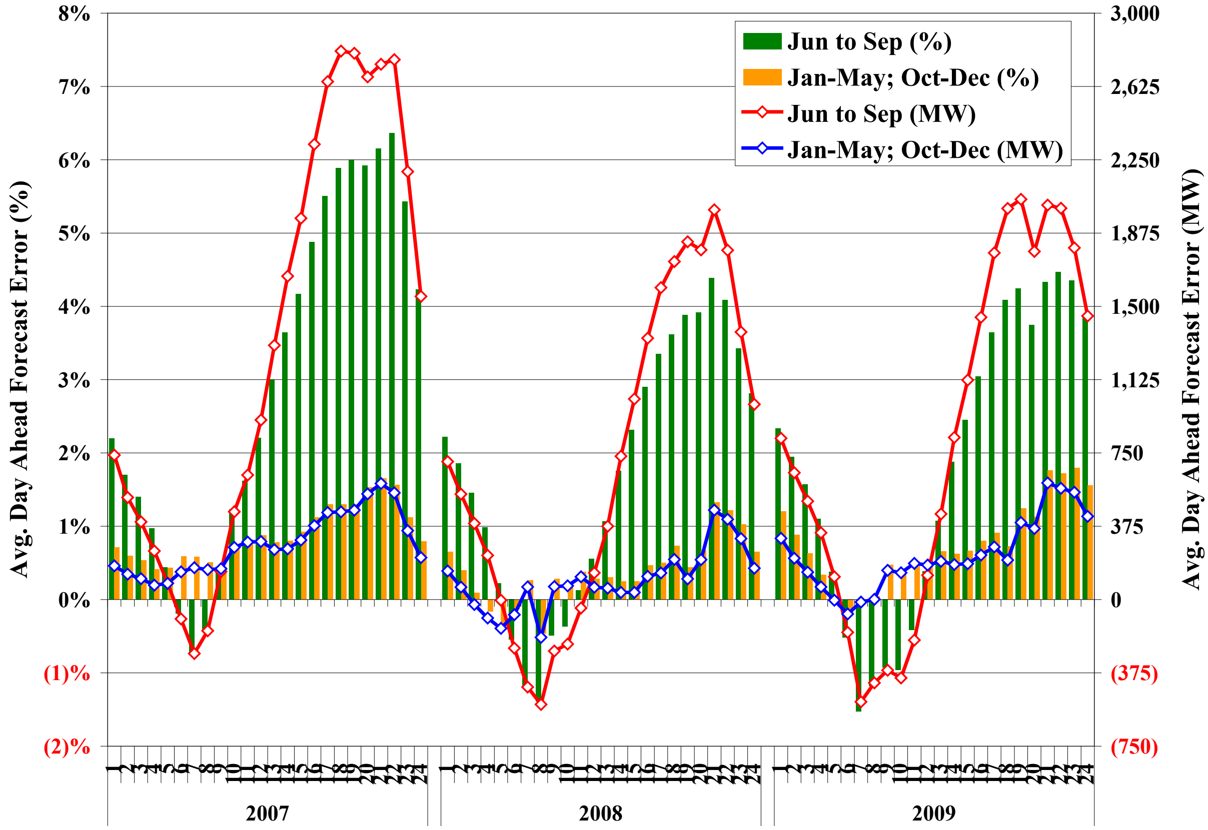
- A continued strong positive bias in ERCOT's day-ahead load forecast -- particularly during summer on-peak hours – that creates the tendency to regularly commit online resources in excess of the quantity required to meet expected demand and operating reserve requirements;
- The implementation of PRR 776, which allowed for quick start gas turbines providing non-spinning reserves to offer the capacity into the balancing energy market; and

5 *See* 2008 ERCOT SOM Report at 81-87.

- The dependence on market participants to submit offers at or near the offer cap to produce scarcity level prices during legitimate operating reserve shortage conditions

The following figure shows the ERCOT day-ahead load forecast error by hour in 2007 through 2009, with the summer and non-summer months presented separately. In this figure, positive values indicate that the day-ahead load forecast was greater than the actual load in real-time.

Day-Ahead Load Forecast Error



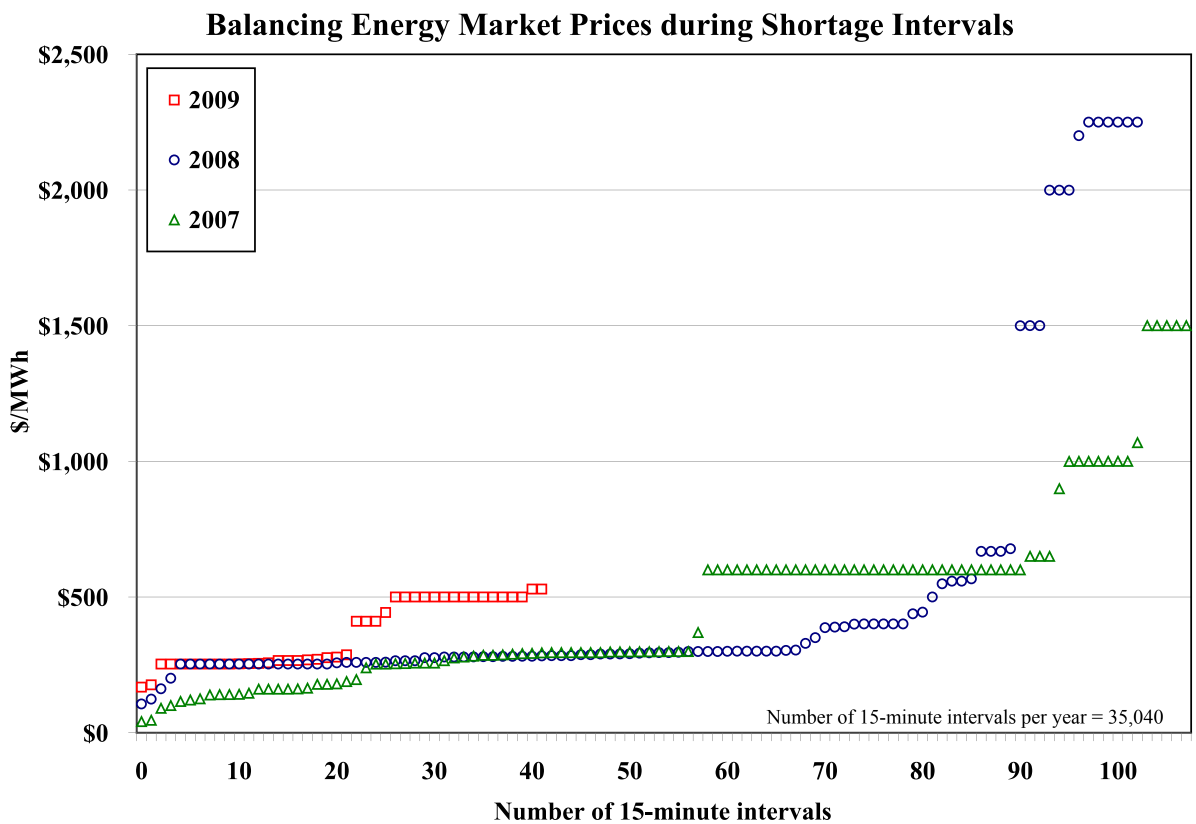
The existence of such a strong and persistent positive bias in the day-ahead load forecast will tend to lead to an inefficient over-commitment of resources and to the depression of real-time prices relative to a more optimal unit commitment. To the extent load uncertainty is driving the bias in the day-ahead load forecast, such uncertainty is more efficiently managed through the procurement of ancillary services such as non-spinning reserve, or through supplemental commitments of short-lead time resources at a time sufficiently prior to, but closer to real-time as uncertainty regarding real-time conditions diminishes.

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

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

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

Under the PUCT rules governing the energy-only market, the mechanism that allows for such pricing during shortage conditions relies upon the submission of high-priced offers by small market participants. The following figure shows the balancing market clearing prices during the 15-minute shortage intervals in 2007 through 2009.



The 42 shortage intervals in 2009 are significantly fewer than the 108 and 103 shortage intervals that occurred in 2007 and 2008, respectively. This reduction can be primarily attributed to the implementation of PRR 776, which allows more timely access to non-spinning reserves through the balancing energy market, thereby reducing the probability of transitional shortages of the core operating reserves. As shown in the figure above, prices during these 42 shortage intervals in 2009 ranged from $168 per MWh to $529 per MWh, with an average price of $364 per MWh and a median price of $283 per MWh.

Although each of the data points in the figure above represents system conditions in which the market was in shortage, the pricing outcomes are widely varied, with the majority of prices reflecting the marginal offer of the most expensive generation resource dispatched as opposed to the value of foregone operating reserves. These results indicate that relying exclusively upon the submission of high-priced offers by market participants was generally not a reliable means of producing efficient scarcity prices during shortage conditions in 2007 through 2009. In fact, although the current system-wide offer cap is $2,250 per MWh, there no hours in 2009 where an offer was submitted by a market participant that approached the offer cap. There were only 33 hours with an offer that exceeded $1,000 per MWh, and the average of the highest offers submitted by any market participant in all hours in 2009 was approximately $400 per MWh.

Despite the mixed and widely varied results of the SPM, private investment in generation capacity in ERCOT has continued, although such investment has been dominated by baseload (non-natural gas fueled) and wind generation. As indicated in the net revenue analyses, these investments are largely driven by significant increases in natural gas prices in the four years prior to 2009. In contrast, private investment in peaking resources in ERCOT has been relatively thin. In an energy-only market, net revenue expectations for peaking resources are much more sensitive the effectiveness of the shortage pricing mechanism than to factors such as the magnitude of natural gas prices, and efficient shortage pricing is a particularly critical element in the ERCOT energy-only market to ensure that the long-term resource adequacy requirements are achieved.

D. Transmission and Congestion

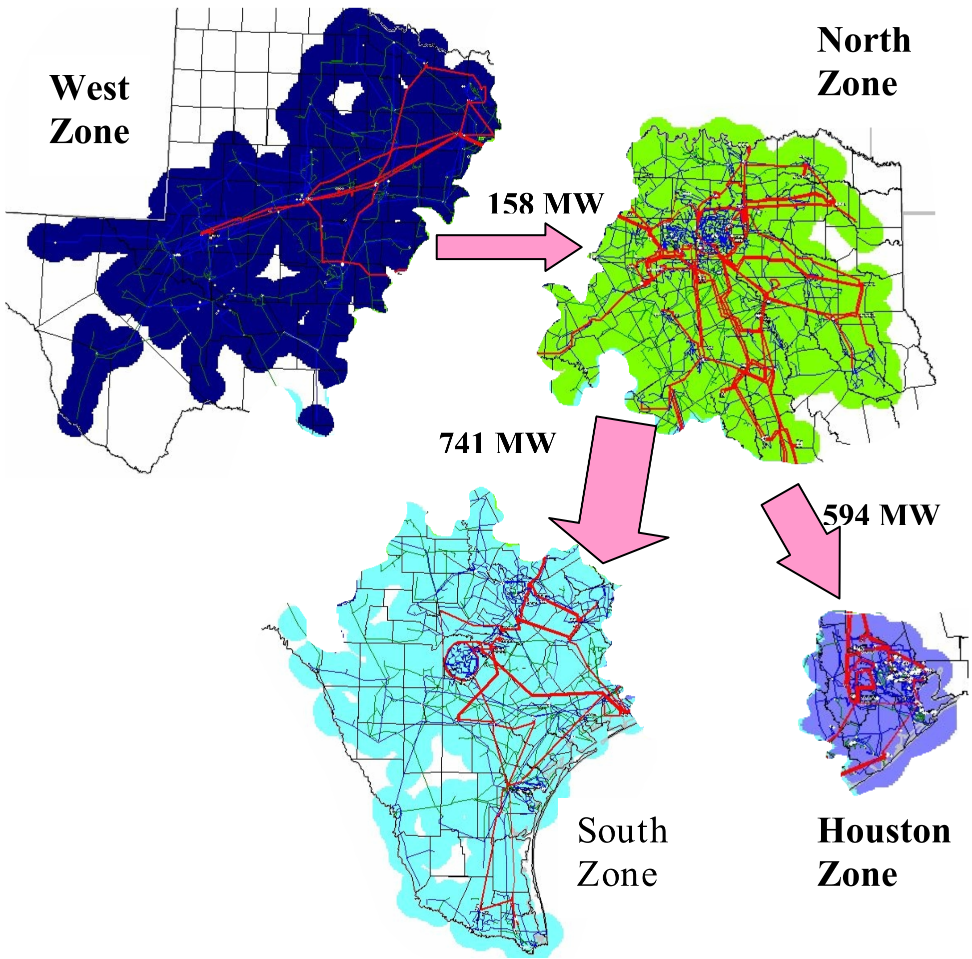
One of the most important functions of any electricity market is to manage the flows of power over the transmission network, limiting additional power flows over transmission facilities when they reach their operating limits. In ERCOT, constraints on the transmission network are managed in two ways. First, ERCOT is made up of zones with the constraints between the zones managed through the balancing energy market. The balancing energy market increases energy production in one zone and reduces it in another zone to manage the flows between the two zones when the interface constraint is binding (*i.e.*, when there is interzonal congestion). Second, constraints within each zone (*i.e.*, local congestion) are managed through the redispatch of individual generating resources. The report evaluates the ERCOT transmission system usage and analyzes the costs and frequency of transmission congestion.

1. Electricity Flows between Zones and Interzonal Congestion

The balancing energy market uses the Scheduling, Pricing, and Dispatch ("SPD") software that dispatches energy in each zone to serve load and manage congestion between zones. The SPD model embodies the market rules and requirements documented in the ERCOT protocols. To manage interzonal congestion, SPD uses a simplified network model with four zone-based locations and five transmission interfaces. The transmission interfaces are referred to as

Commercially Significant Constraints ("CSCs"). The following figure shows the average flows modeled in SPD during 2009 over each of these CSCs.

Average Modeled Flows on Commercially Significant Constraints



When interzonal congestion exists, higher-cost energy must be produced within the constrained zone because lower-cost energy cannot be delivered over the constrained interfaces. When this occurs, participants must compete to use the available transfer capability between zones. To allocate this capability in the most efficient manner possible, ERCOT establishes a clearing price for each zone and the price difference between zones is charged for any interzonal transactions.

The analysis of these CSC flows in this report indicates that:

- The simplifying assumptions made in the SPD model can result in modeled flows that are considerably different from actual flows.

- A considerable quantity of flows between zones occurs over transmission facilities that are not defined as part of a CSC. When these flows cause congestion, it is beneficial to create a new CSC to better manage congestion over that path.
- The differences between SPD-modeled flows and actual flows on CSCs create operational challenges for ERCOT that result in the inefficient use of scarce transmission resources.

Inter-zonal congestion was most frequent in 2009 on the West to North CSC, followed by the North to Houston and the North to South CSCs.

The North to Houston CSC was binding in 625 15-minute intervals with an annual average shadow price of $2.01 per MW.[6] These values represent a significant reduction in both the frequency and magnitude of congestion compared to 2008 when the North to Houston CSC was binding in 1,447 intervals with an annual average shadow price of $20.

The North to South CSC was binding in 387 15-minute intervals with an annual average shadow price of $8.39 per MW. Like the North to Houston CSC, these values represent a significant reduction in both the frequency and magnitude of congestion compared to 2008 when the North to South CSC was binding in 2,531 intervals with an annual average shadow price of $22.

The decreased congestion on the North to Houston and North to South CSCs in 2009 is primarily attributable to the implementation of PRR 764 in June 2008 that revised the definition of valid zonal transmission constraints and improved the efficiency of transmission congestion management within the context of the zonal market model.[7]

A significant percentage of the congestion on the North to South CSC occurred during June 2009. During this timeframe, the ERCOT market experienced very high temperatures and associated increases in load levels, as well as a number of outages at baseload generating facilities, particularly in the South Zone. This combination of events led to an increase in the frequency of congestion on the North to South CSC as well as local congestion related to import limitations into the San Antonio area from the north. In the zonal model, the most effective

6 The shadow price of a transmission constraint represents the marginal value of the use a transmission element. The shadow price of a transmission element will be zero unless the transmission element is bieing used to its full capacity, in which case it will have a positive shadow price.

7 *See* 2008 ERCOT SOM Report at 81-87.

resolution to North to South congestion is to increase generation in the South Zone. However, effective zonal congestion management on the North to South CSC was affected by the local congestion in the San Antonio area, which is most effectively resolved by increasing generation in and South of San Antonio, and decreasing generation north of San Antonio. Because most of the generation resources located north of San Antonio required to decrease output to manage the local congestion in the San Antonio area are also in the South Zone that was broadly required to increase output to manage the zonal North to South congestion, competing reliability objectives were present that complicated the simultaneous resolution of both the North to South zonal congestion and the intrazonal San Antonio import-related congestion. Faced with these competing reliability objectives and the inability to resolve both reliability issues within the context of the zonal model and its bifurcated process of zonal and local transmission congestion management, ERCOT implemented a temporary transmission switching solution in late June that effectively increased the transfer capability on the North to South CSC, thereby resolving these competing reliability objectives under the atypical load and generator outage conditions experienced at that time.

The West to North CSC was binding in 3,121 15-minute intervals in 2009. This was more frequent than any other CSC in 2009 and, with the exception of the same CSC in 2008 that was binding for 5,320 intervals, more frequent than any other CSC since the inception of single control area operations in 2001. The primary reason for the high frequency of congestion on the West to North CSC in 2008 and 2009 is the significant increase in installed wind generation relative to the load in the West Zone and limited transmission export capability to the broader market.

Although the marginal production cost of wind generators is near zero, the operating economics are affected by federal production tax credits and state renewable energy credits, which lead to negative-priced offers from most wind generators. Thus, when transmission congestion occurs that requires wind generators to curtail their output, negative balancing energy market prices will result in the West Zone. The hourly average balancing energy market price in the West Zone was less than zero in over 700 hours during 2009.

Although the frequency of zonal transmission congestion on the West to North CSC was very high in 2009 compared to other zonal constraints, the frequency of congestion on this constraint was lower than in 2008. However, zonal congestion data do not provide a complete view of the congestion situation in the West Zone. While the quantity of zonal curtailments for wind resources in the West Zone decreased from 604,000 MWh in 2008 to 442,000 MWh in 2009, the quantity of local curtailments increased significantly, rising from 812,000 MWh in 2008 to over 3,400,000 MWh in 2009. Hence, while curtailments in the West Zone associated with zonal congestion decreased in 2009, total congestion-related curtailments in the West Zone increased significantly in 2009.

Given the current transmission infrastructure and the level of existing wind facilities in the West Zone, the quantity of wind production that can be reliably accommodated in the West Zone will continue to be significantly limited for several years until the planned transmission improvements identified through the Competitive Renewable Energy Zone ("CREZ") project can be completed.

2. Transmission Congestion Rights and Payments

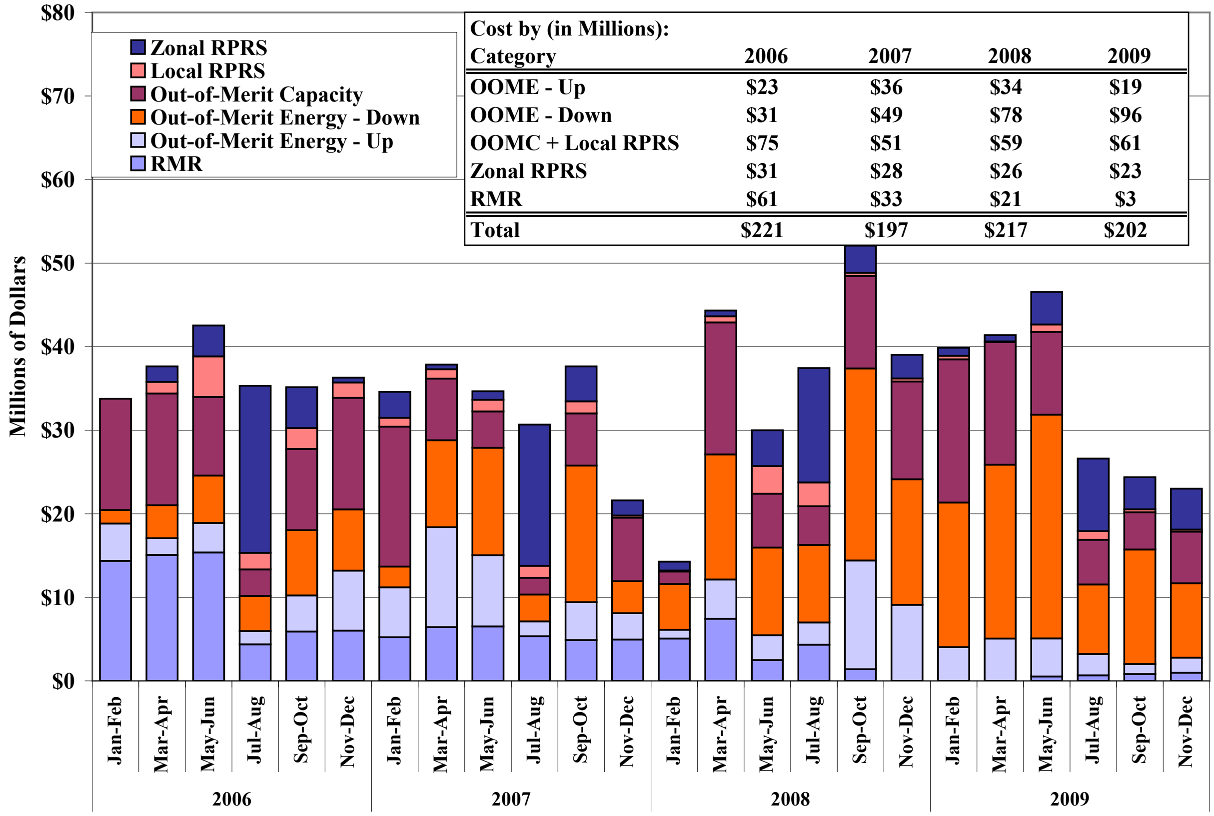
Participants in Texas can hedge against congestion in the balancing energy market by acquiring Transmission Congestion Rights ("TCRs") between zones, which entitle the holder to payments equal to the difference in zonal balancing energy prices. Because the modeled limits for the CSC interfaces vary substantially, the quantity of TCRs defined over a congested CSC frequently exceeds the modeled limits for the CSC. When this occurs, the congestion revenue collected by ERCOT will be insufficient to satisfy the financial obligation to the holders of the TCRs and the revenue shortfall is collected from loads through uplift charges. Payments to TCR holders have consistently exceeded the congestion rents that have been collected from the balancing market in 2006 through 2009. In 2009 congestion rents covered only 72 percent of the payments to TCR holders, with an annual net revenue shortfall of $53 million.

In a perfectly efficient system with no uncertainty, the average congestion cost in real-time should equal the auction price of the congestion rights. In the real world, however, we would expect reasonably close convergence with some fluctuations from year to year due to uncertainties. Market participants generally under-estimated the value of congestion by a wide margin in 2008, particularly during the first half of the year. These outcomes were likely influenced by the congestion management procedures that were applied during the first half of the year and modified by the implementation of PRR 764 in June 2008. In 2009, market participant over-estimated the value of congestion on the West to North and North to Houston CSCs, but once again underestimated the value of congestion on the North to South CSC. This was likely due to the unexpected nature of the contributors leading to congestion on this CSC.

3. Local Congestion and Local Capacity Requirements

ERCOT manages local (intrazonal) congestion by using out-of-merit dispatch ("OOME up" and "OOME down"), which causes units to depart from their scheduled output levels. When insufficient capacity is committed to meet local or system reliability requirements, ERCOT commits additional resources to provide the necessary capacity in either the day-ahead market or in the adjustment period (the adjustment period includes the hours after the close of the day-ahead market up to one hour prior to real-time). Capacity required for local reliability constraints is procured through either the Replacement Reserve Service market ("Local RPRS") or as out-of-merit capacity ("OOMC"). Capacity required for system reliability requirements (*i.e.,* the requirement that the total system-wide online capacity be greater than or equal to the sum of the ERCOT load forecast plus operating reserves in each hour) is procured through either the RPRS market ("Zonal RPRS") or as OOMC. ERCOT also enters into RMR agreements with certain generators needed for local reliability that may otherwise be mothballed or retired. When RMR units are called out-of-merit, they receive revenues specified in the agreements rather than standard OOME or OOMC payments. The following figure shows the out-of-merit energy and capacity costs, including RMR costs, from 2006 to 2009.

Expenses for Out-of-Merit Capacity and Energy



Cost by (in Millions): Category	2006	2007	2008	2009
OOME - Up	$23	$36	$34	$19
OOME - Down	$31	$49	$78	$96
OOMC + Local RPRS	$75	$51	$59	$61
Zonal RPRS	$31	$28	$26	$23
RMR	$61	$33	$21	$3
Total	$221	$197	$217	$202

The results in the figure above show that overall uplift costs for RMR units, OOME units, OOMC/Local RPRS and Zonal RPRS[8] units were $202 million in 2009, which is a $15 million decrease over the $217 million in 2008. OOME Down and RMR costs accounted for the most significant portion of the change in 2009. OOME down increased from $78 million in 2008 to $96 million in 2009. These values represent significant increases in OOME Down costs from 2006 and 2007, and are primarily attributable to increases in OOME Down instructions for wind resources in the West Zone. RMR costs decreased from $21 million in 2008 to $3 million in 2009. This figure also shows that the highest Zonal RPRS costs occur in July and August when electricity demand in the ERCOT region is at its highest levels.

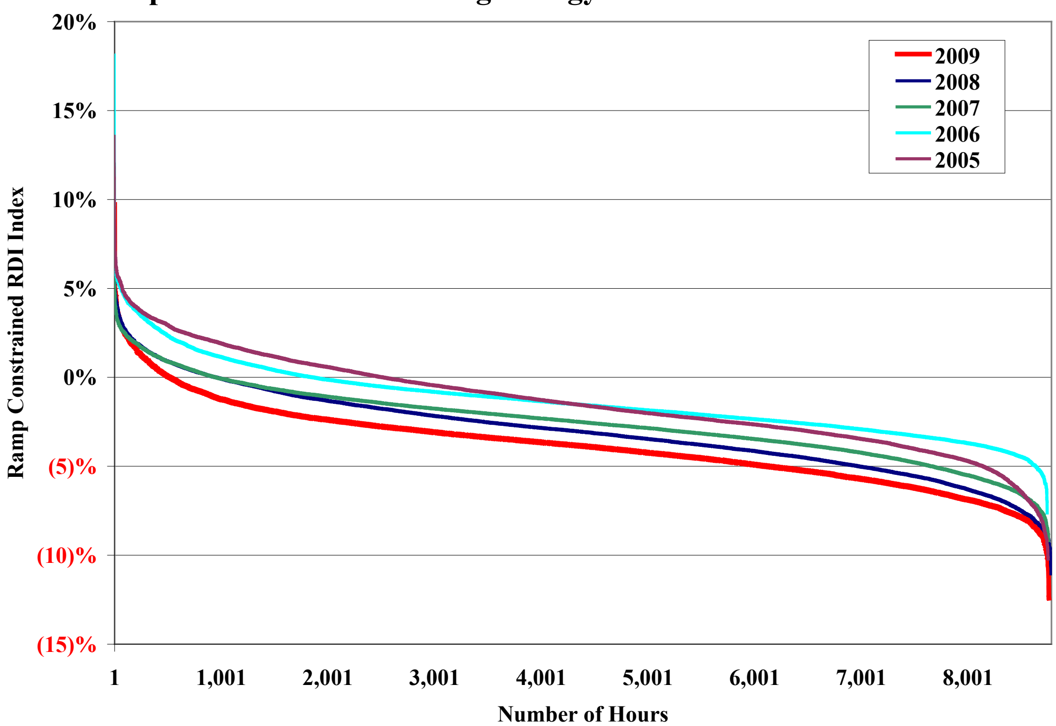
[8] Zonal RPRS for system adequacy is deployed at the second stage of the RPRS run, which is affected by the deployment at the first stage of the RPRS run, or the local RPRS deployment. Because ERCOT Protocols allocate the costs of local and zonal RPRS in the same manner, we have included both as local congestion costs.

E. Analysis of Competitive Performance

The report evaluates two aspects of market power, structural indicators of market power and behavioral indicators that would signal attempts to exercise market power. The structural analysis in this report focuses on identifying circumstances when a supplier is "pivotal," *i.e.*, when its generation is essential to serve the ERCOT load and satisfy the ancillary services requirements.

The pivotal supplier analysis indicates that the frequency with which a supplier was pivotal in the balancing energy market decreased in 2009 compared to 2008. The following figure shows the ramp-constrained balancing energy market Residual Demand Index ("RDI") duration curves for 2005 through 2009. When the RDI is greater than zero, the largest supplier's balancing energy offers are necessary to prevent a shortage of offers in the balancing energy market.

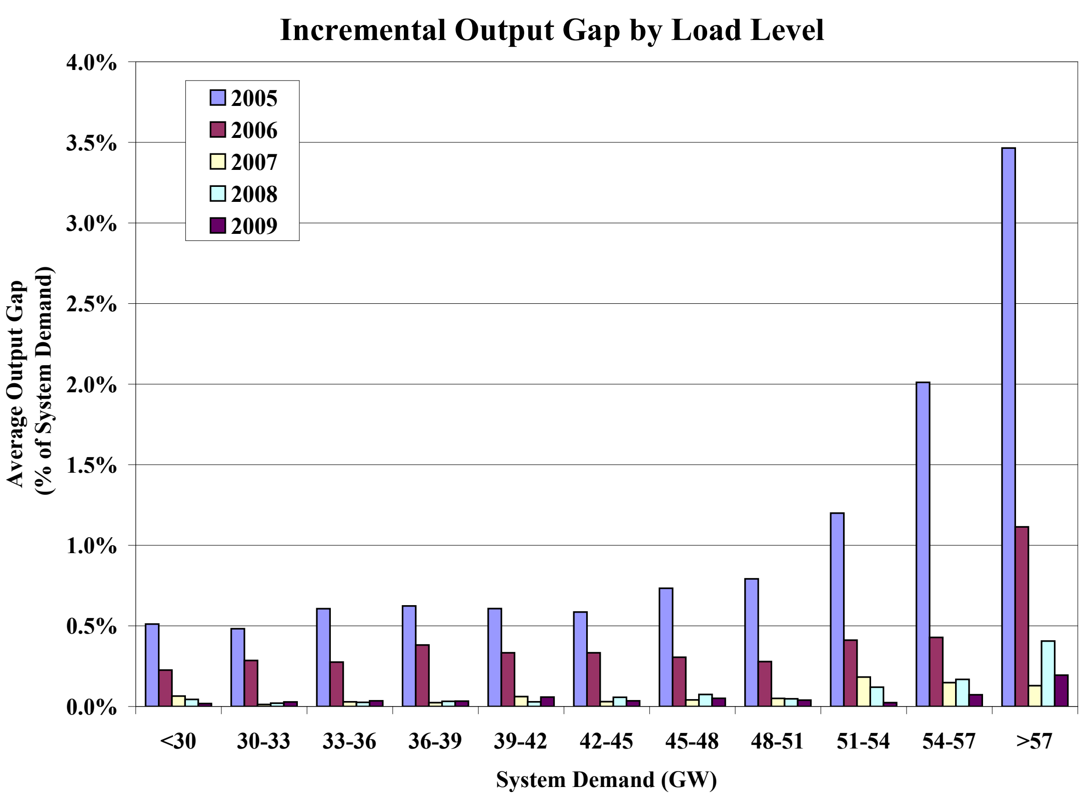
Ramp-Constrained Balancing Energy Market RDI Duration Curve



The frequency with which at least one supplier was pivotal (*i.e.*, an RDI greater than zero) has fallen consistently over the last five years from 29 and 21 percent of the hours in 2005 and 2006, respectively, to less than 11 percent of the hours in 2007 and 2008, to less than 6 percent of the hours in 2009. These results highlight the trend of continued improvement in the structural competitiveness of the balancing energy market over the last five years.

A behavioral indicator that evaluates potential economic withholding is measured 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 balancing energy price. A participant can economically withhold resources, as measured by the output gap, by raising its balancing energy offers so as not to be dispatched or by not offering unscheduled energy in the balancing energy market.

The figure below compares the real-time load to the average incremental output gap for all market participants as a percentage of the real-time system demand from 2005 through 2009.



The figure above shows that the competitiveness of supplier offers improved considerably in 2006 compared to 2005, followed by even more substantial improvement in 2007 through 2009. In 2009, the overall magnitude of the incremental output gap remains very small and does not raise significant economic withholding concerns.

Overall, based upon the analyses in this section, we find that the ERCOT wholesale market performed competitively in 2009.

I. REVIEW OF MARKET OUTCOMES

A. Balancing Energy Market

1. Balancing Energy Prices During 2009

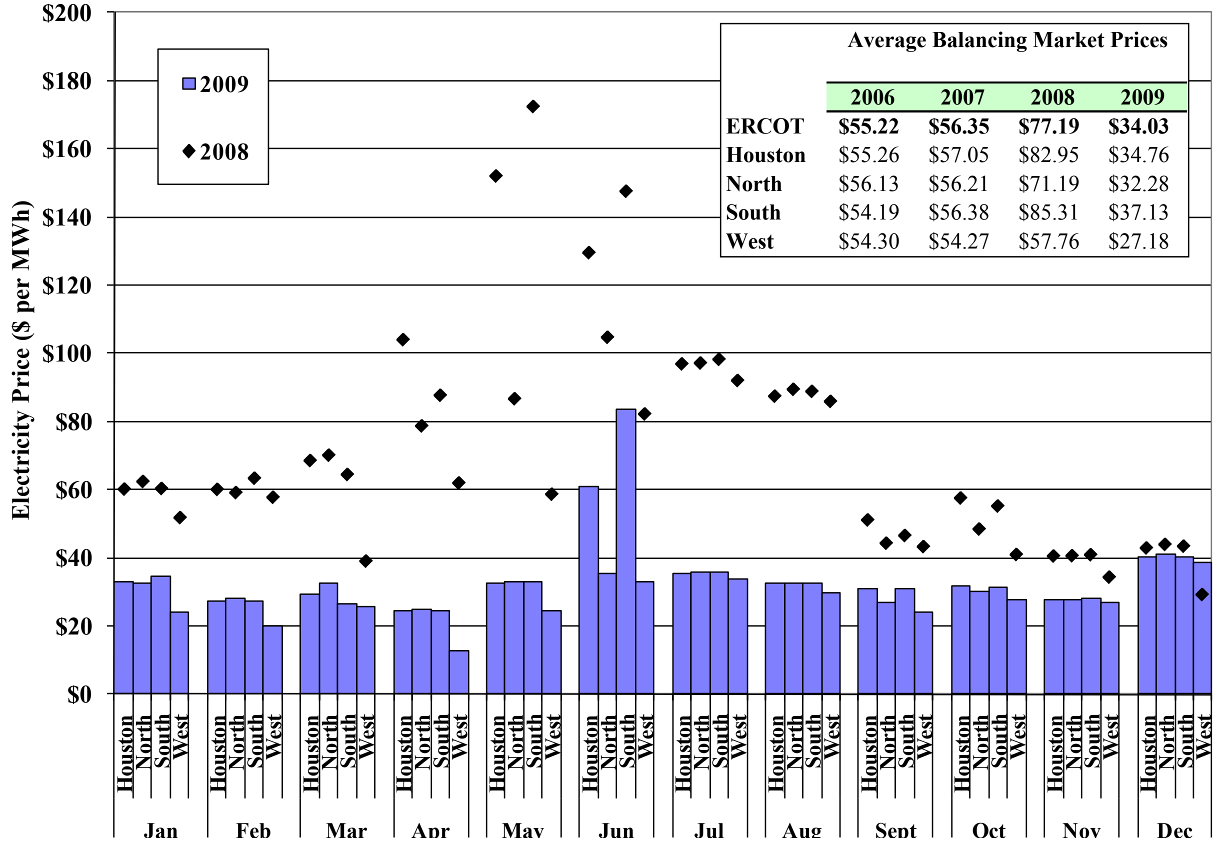
The balancing energy market is the spot market for electricity in ERCOT. As is typical in other wholesale markets, only a small share of the power produced in ERCOT is transacted in the spot market, although at times such transactions can exceed 10 percent of total demand. Although most power is purchased through bilateral forward contracts, outcomes in the balancing energy market are very important because of the expected pricing relationship between spot and forward markets (including bilateral markets).

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 balancing energy market will translate to artificially-low forward prices. Likewise, price spikes in the balancing energy market will increase prices in the forward markets. This section evaluates and summarizes balancing energy market prices during 2009.

To summarize the price levels during the past four years, Figure 1 shows the monthly load-weighted average balancing energy market prices in each of the ERCOT zones during 2008 and 2009, with annual summary data for 2006 and 2007.[9]

[9] The load-weighted average prices are calculated by weighting the balancing energy price for each interval and each zone by the total zonal load in that interval. For this evaluation, balancing energy prices are load-weighted since this is the most representative of what loads are likely to pay (assuming that balancing energy prices are generally consistent with bilateral contract prices).

Figure 1: Average Balancing Energy Market Prices



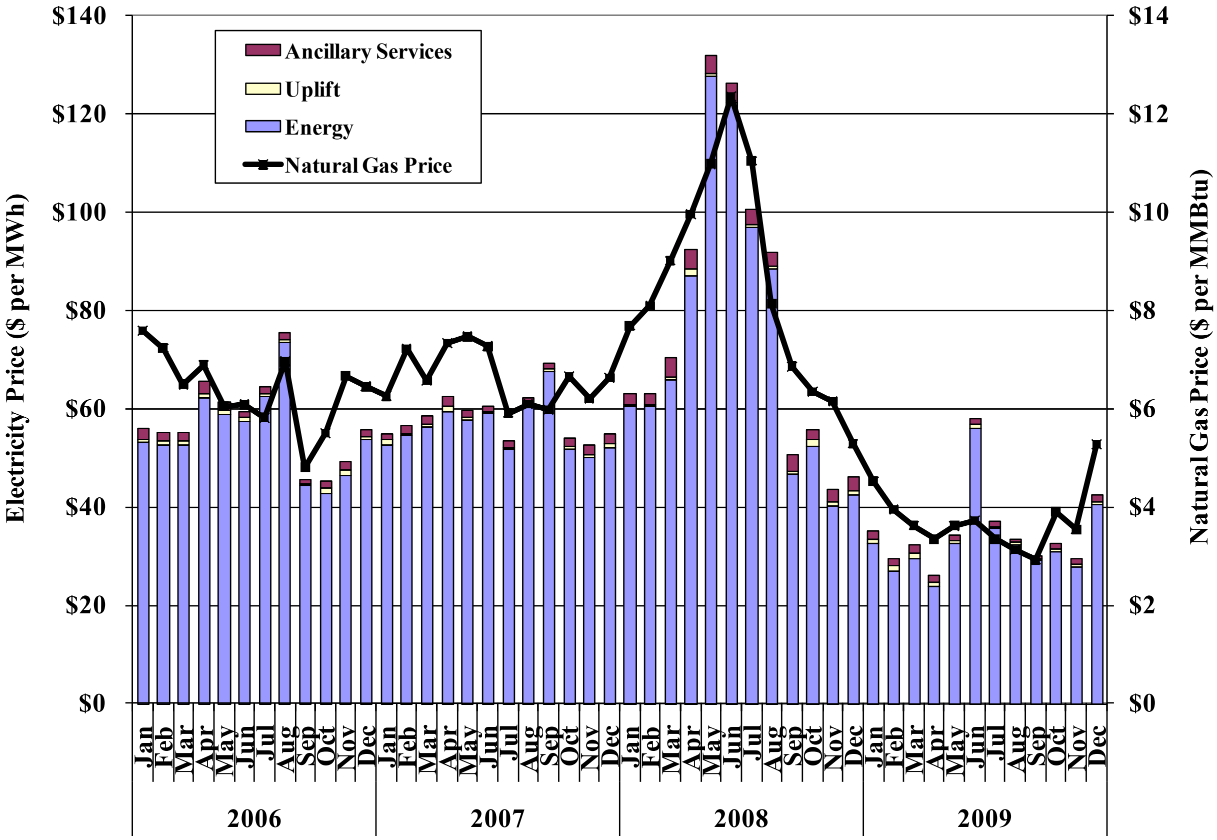
	2006	2007	2008	2009
ERCOT	**$55.22**	**$56.35**	**$77.19**	**$34.03**
Houston	$55.26	$57.05	$82.95	$34.76
North	$56.13	$56.21	$71.19	$32.28
South	$54.19	$56.38	$85.31	$37.13
West	$54.30	$54.27	$57.76	$27.18

ERCOT average balancing energy market prices were 56 percent lower in 2009 than in 2008, with an ERCOT-wide load weighted average price of $34.03 per MWh in 2009 compared to $77.19 per MWh in 2008. April through August experienced the highest balancing energy market price reductions in 2009, averaging 66 percent lower than the prices in the same months in 2008. With the exception of the West Zone in December, the balancing energy prices were lower in every month in all zones in 2009 than in 2008.

The average natural gas price fell 56 percent in 2009, averaging $3.74 per MMBtu in 2009 compared to $8.50 per MMBtu in 2008. Natural gas prices reached a maximum monthly average of $12.37 per MMBtu in July 2008, and reached a minimum monthly average of $2.93 per MMBtu in September 2009. Hence, the changes in energy prices from 2008 to 2009 were largely a function of natural gas price movements.

The next analysis evaluates the total cost of serving load in the ERCOT wholesale market. In addition to the costs of energy, loads incur costs associated with ancillary services and "uplift".[10] We have calculated an average all-in price of electricity for ERCOT that is intended to reflect wholesale energy costs as well as these additional costs. Figure 2 shows the monthly average all-in price for all of ERCOT from 2006 to 2009 and the associated natural gas price.

Figure 2: Average All-in Price for Electricity in ERCOT



The components of the all-in price of electricity include:

- Energy costs: Balancing energy market prices are used to estimate energy costs, under the assumption that the price of bilateral energy purchases converges with balancing energy market prices over the long-term, as discussed above.

- Ancillary services costs: These are estimated based on the demand and prices in the ERCOT markets for regulation, responsive reserves, and non-spinning reserves.

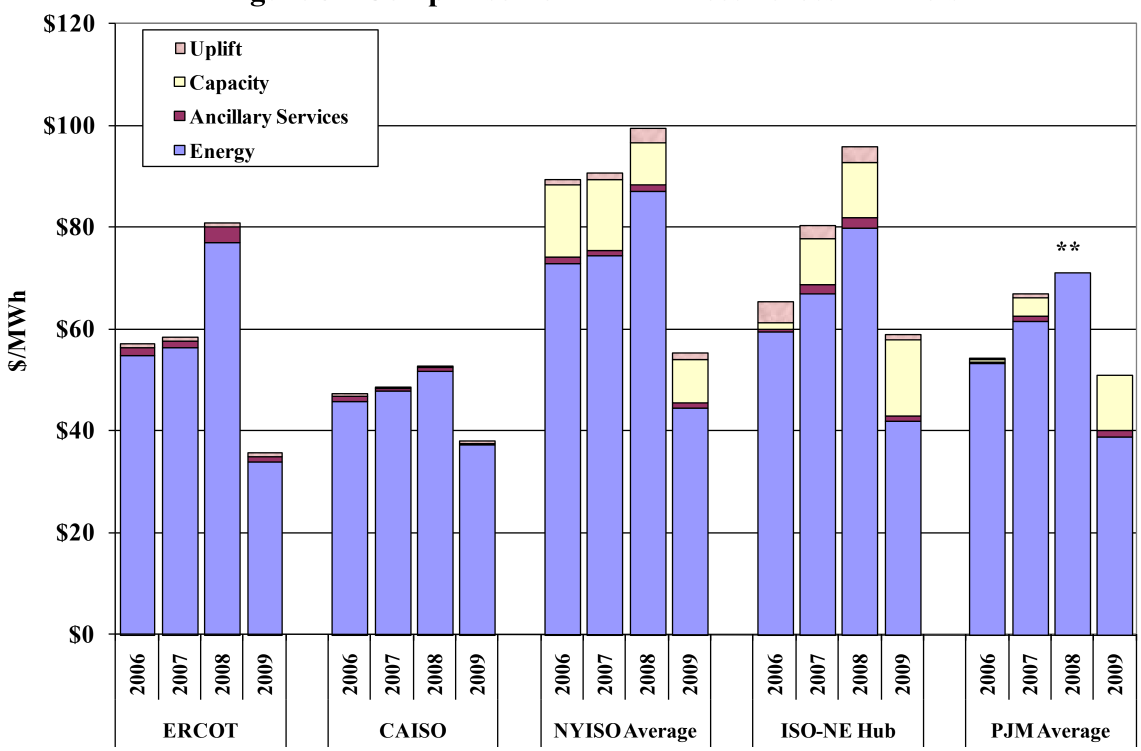
10 As discussed in more detail in Section III, uplift costs are costs that are allocated to load that pay for out-of-merit dispatch, out-of-merit commitment, and Reliability Must Run contracts.

- Uplift costs: Uplift costs are assigned market-wide on a load-ratio share basis to pay for out-of-merit energy dispatch, out-of-merit commitment, replacement reserve services and Reliability Must Run contracts.

Figure 2 indicates that natural gas prices were a primary driver of the trends in electricity prices from 2006 to 2009. 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 the balancing energy market prices.

To provide additional perspective on the outcomes in the ERCOT market, our next analysis compares the all-in price metrics for ERCOT and other electricity markets. The following figure compares the all-in prices in ERCOT with other organized electricity markets in the U.S.: California ISO, New York ISO, ISO New England, and PJM. For each region, the figure reports the average cost (per MWh of load) for energy, ancillary services (reserves and regulation), capacity markets (if applicable), and uplift for economically out-of-merit resources.

Figure 3: Comparison of All-in Prices across Markets

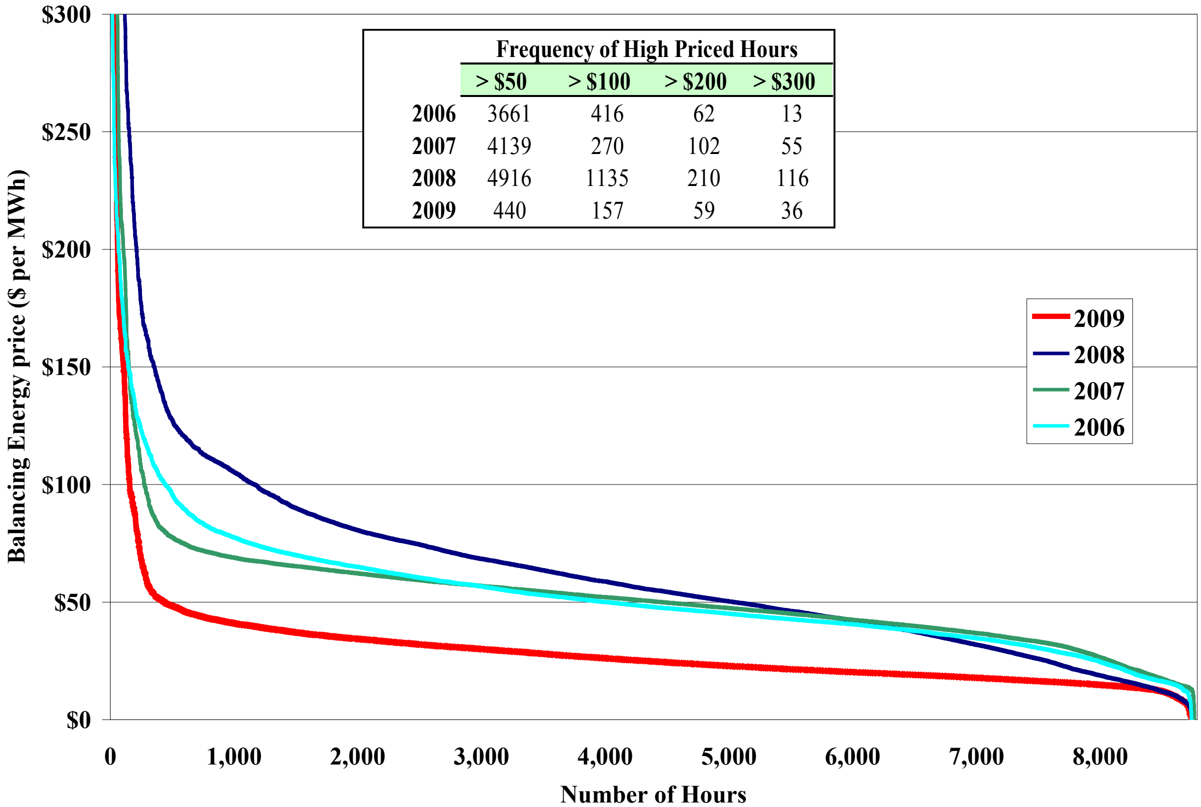


** 2008 Capacity, Ancillary Services and Uplift data unavailable for PJM

Figure 3 shows that energy prices increased in wholesale electricity markets across the U.S. in 2009, primarily due to decreases in fuel costs, and that the ERCOT market experienced the lowest all-in wholesale prices of any of these markets in 2009

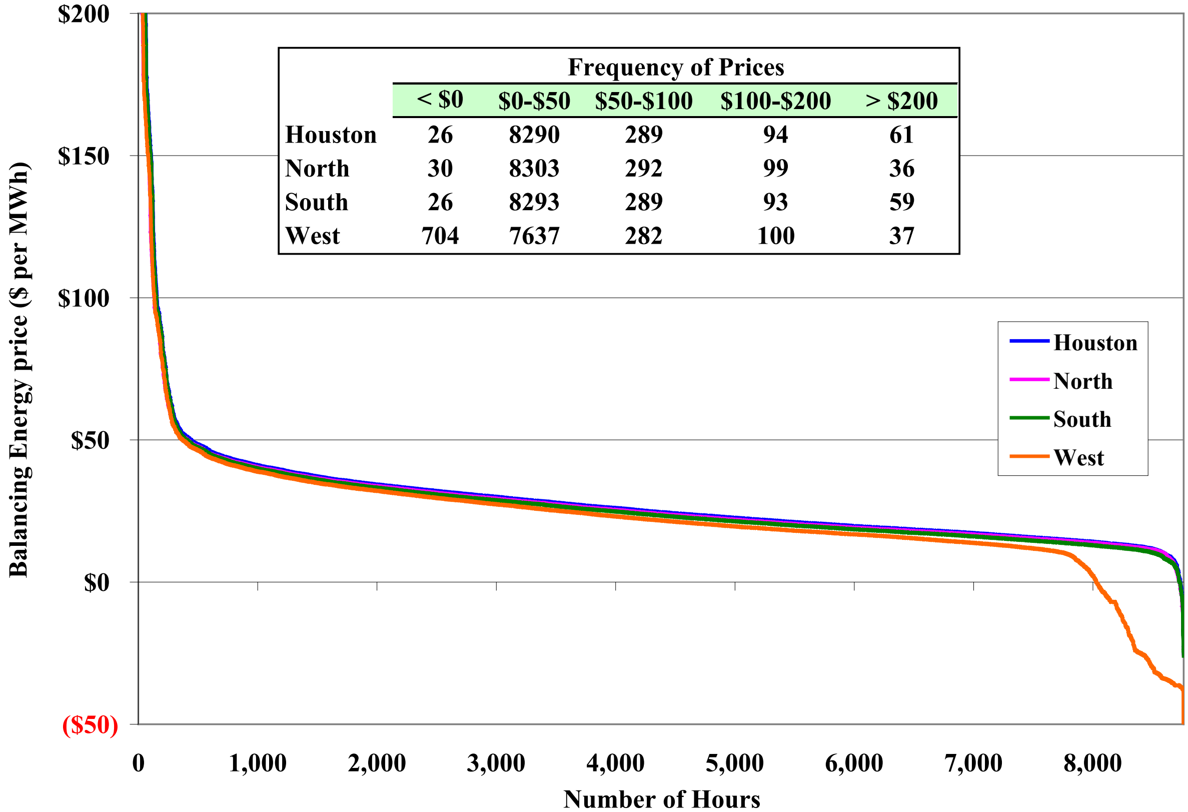
Figure 4 presents price duration curves for the ERCOT balancing energy market in each year from 2006 to 2009. A price duration curve indicates the number of hours (shown on the horizontal axis) that the price is at or above a certain level (shown on the vertical axis). The prices in this figure are hourly load-weighted average prices for the ERCOT balancing energy market.

Figure 4: ERCOT Price Duration Curve



Balancing energy prices exceeded $50 per MWh in only 440 hours in 2009 compared to more than 4,900 hours in 2008. These year-to-year changes reflect lower natural gas prices in 2009 that affect electricity prices in a broad range of hours.

Figure 5: Zonal Price Duration Curves

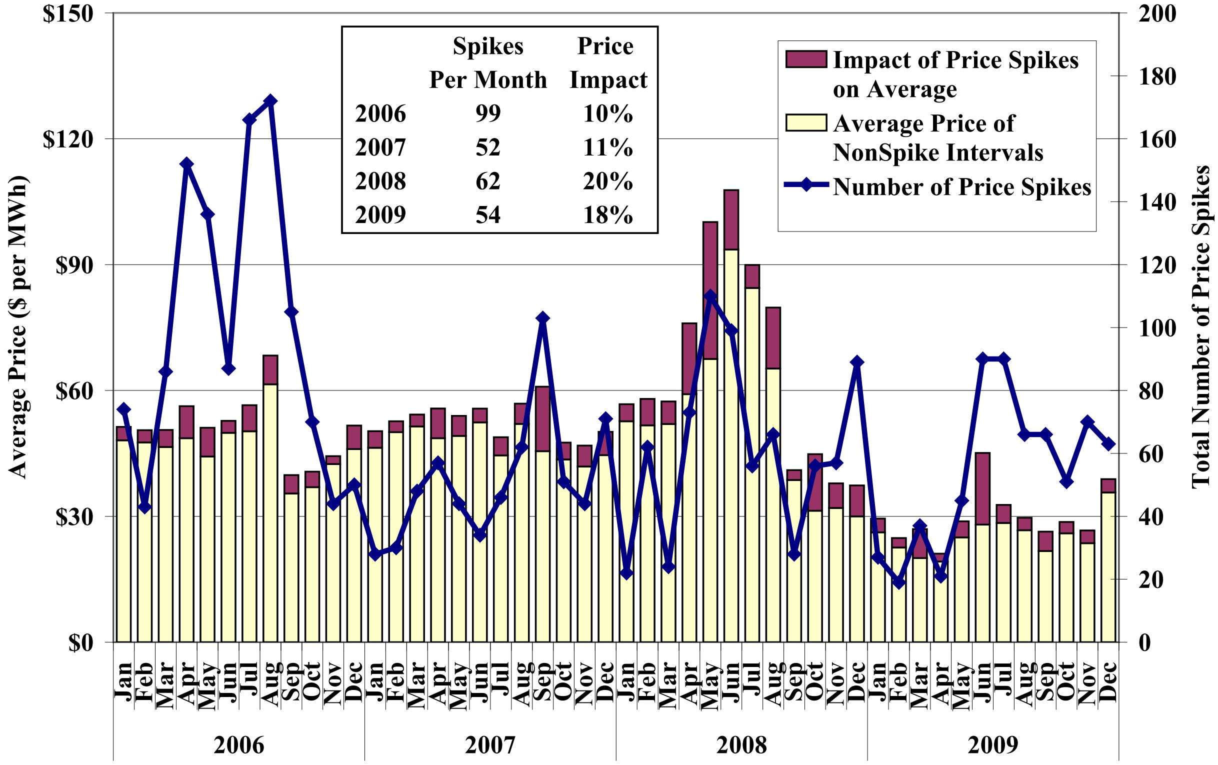


	Frequency of Prices				
	< $0	$0-$50	$50-$100	$100-$200	> $200
Houston	26	8290	289	94	61
North	30	8303	292	99	36
South	26	8293	289	93	59
West	704	7637	282	100	37

Figure 5 shows the hourly average price duration curve for each of the four ERCOT zones in 2009 and that the Houston, North and South Zones had similar prices over the majority of hours in 2009. The price duration curve for the West Zone is generally lower than all other zones, with over 700 hours when the average hourly price was less than zero. These zonal price differences are caused by zonal transmission congestion, as discussed in more detail in Section III.

Other market factors that affect balancing energy prices occur in a subset of intervals, such as the extreme demand conditions that occur during the summer or when there is significant transmission congestion. Figure 4 shows that there were differences in balancing energy market prices between 2006 and 2009 at the highest price levels. For example, 2008 experienced considerably more occasions when prices spiked to greater than $300 per MWh than previous years. To better observe the effect of the highest-priced hours, the following analysis focuses on the frequency of price spikes in the balancing energy market from 2006 to 2009. Figure 6 shows average prices and the number of price spikes in each month of 2006 to 2009. In this case, price spikes are defined as intervals where the load-weighted average Market Clearing Price of Energy ("MCPE") in ERCOT is greater than 18 MMbtu per MWh times the prevailing natural gas price (a level that should exceed the marginal costs of virtually all of the on-line generators in ERCOT).

Figure 6: Average Balancing Energy Prices and Number of Price Spikes



The number of price spike intervals was 62 per month during 2008. The number decreased in 2009 to 54 per month. The highest frequency of price spikes occurred in June and July during 2008, caused by significant transmission congestion that ERCOT was inefficiently attempting to resolve by using zonal congestion management techniques.[11] The high number of price spikes during June 2009 was also the result of zonal congestion management actions, although for reasons different than in 2008, as discussed in Section III. Other months with a higher frequency of price spikes in 2009 – particularly in the months after May 2009 – can be attributed to the more frequent deployment of off-line, quick start gas turbines in the balancing energy market as a result of the implementation of PRR 776 in May 2009, as discussed in Section II. Off-line, quick start gas turbines typically have a marginal cost that is greater than the 18 MMBtu per MWh threshold used in Figure 6.

[11] *See* 2008 ERCOT SOM Report, at 81-87.

To measure the impact of these price spikes on average price levels, the figure also shows the average prices with and without the price spike intervals. The top portions of the stacked bars show the impact of price spikes on monthly average price levels. The impact grows with the frequency of the price spikes, averaging $4.68, $5.30, $10.71 and $4.67 per MWh during 2006, 2007, 2008 and 2009, respectively. Even though price spikes account for a small portion of the total intervals, they have a significant impact on overall price levels.

Although fuel price fluctuations are the dominant factor driving electricity prices in the ERCOT wholesale market, fuel prices alone do not explain all of the price outcomes. Several other factors provided a meaningful contribution to price outcomes in 2009. These factors include (1) changes in peak demand and average energy consumption levels, as discussed in Section II; (2) changes in the frequency and magnitude of transmission congestion, as discussed in Section III; (3) the increased penetration of wind resources, as discussed in Sections II and III; (4) the effectiveness of the scarcity pricing mechanism, as discussed in Section II; and (5) the competitive performance of the wholesale market, as discussed in Section IV. Analyses in the next subsection adjust for natural gas price fluctuations to better highlight variations in electricity prices not related to fuel costs.

2. Balancing Energy Prices Adjusted for Fuel Price Changes

The pricing patterns shown in the prior subsection are driven to a large extent by changes in fuel prices, natural gas prices in particular. However, prices are influenced by a number of other factors as well. To clearly identify changes in electricity prices that are not driven by changes in natural gas prices, Figure 7 and Figure 8 show balancing energy prices adjusted to remove the effect of natural gas price fluctuations. The first chart shows a duration curve where the balancing energy price is replaced by the marginal heat rate that would be implied if natural gas were always on the margin. The *Implied Marginal Heat Rate* equals the *Balancing Energy Price* divided by the *Natural Gas Price*.[12] The second chart shows the same duration curves for the five percent of hours in each year with the highest implied heat rate. Both figures show duration curves for the implied marginal heat rate for 2006 to 2009.

12 This methodology implicitly assumes that electricity prices move in direct proportion to changes in natural gas prices.

In contrast to Figure 4, Figure 7 shows that the implied marginal heat rates were relatively consistent across the majority of hours from 2006 to 2009. The drop in energy prices from 2008 to 2009 is much less dramatic when the effect of fuel price changes is removed, which confirms that the increase in prices in most hours is primarily due to the rise in natural gas prices. However, the price differences that were apparent from Figure 4 in the highest-priced hours persist even after the adjustment for natural gas prices. For example, the number of hours when the implied heat rate was greater than 30 MMBtu per MWh was 73, 103, 145 and 146 in 2006, 2007, 2008 and 2009, respectively. This indicates that there are price differences that are due to factors other than changes in natural gas prices. The increase in the number of hours when the implied heat rate was greater than 30 MMBtu per MWh in 2008 compared to 2006 and 2007 is primarily attributable to chronic and severe congestion on the North to Houston and North to South constraints in April through June 2008. In contrast, although a portion of the 146 hours with an implied heat rate greater than 30 MMBtu per MWh in 2009 is associated with significant congestion on the North to South constraint in late June 2009, many of these hours in 2009 are associated with the implementation of PRR 776 that increased the frequency of the deployment of off-line, quick start gas turbines in the balancing energy market, as discussed in Section II. Figure 8 shows the implied marginal heat rates for the top five percent of hours in 2006 through 2009 and highlights the increase in the number of with an implied marginal heat rate greater than 30 MMBtu per MWh in 2008 and 2009 compared to 2006 and 2007.

Figure 7: Implied Marginal Heat Rate Duration Curve – All Hours

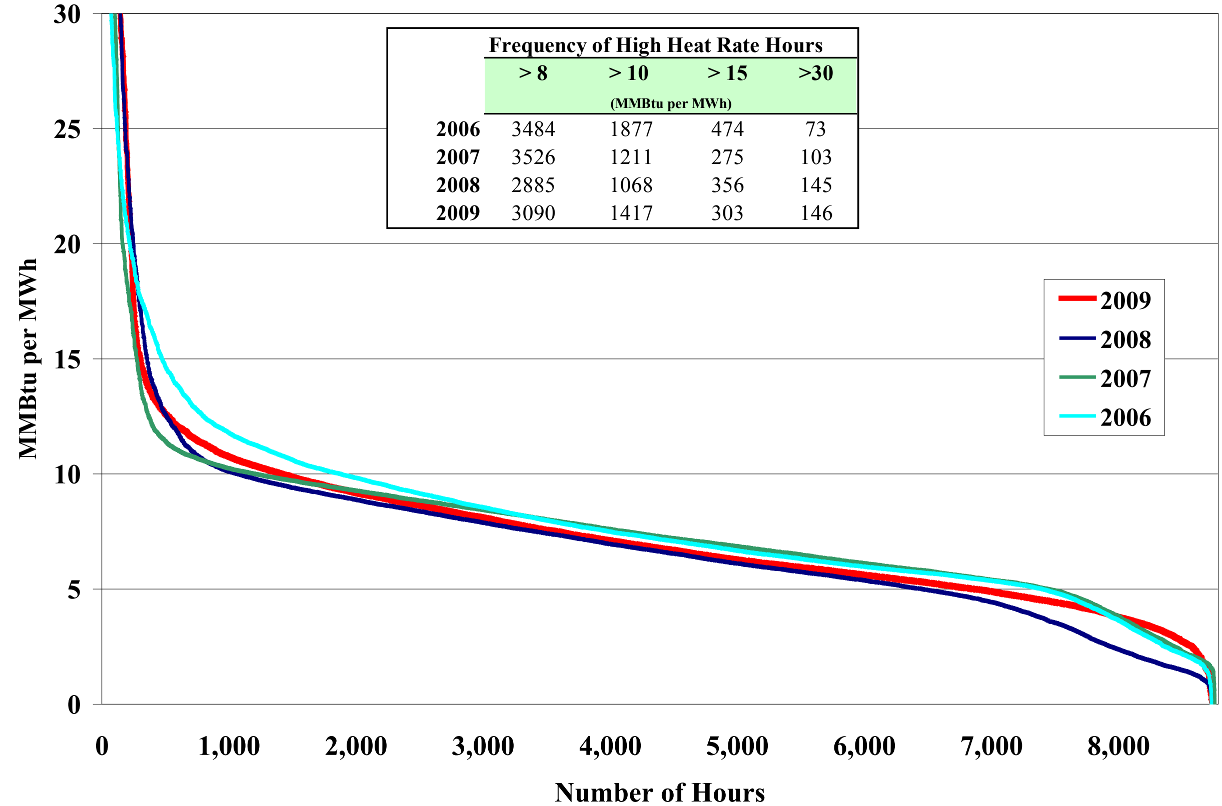
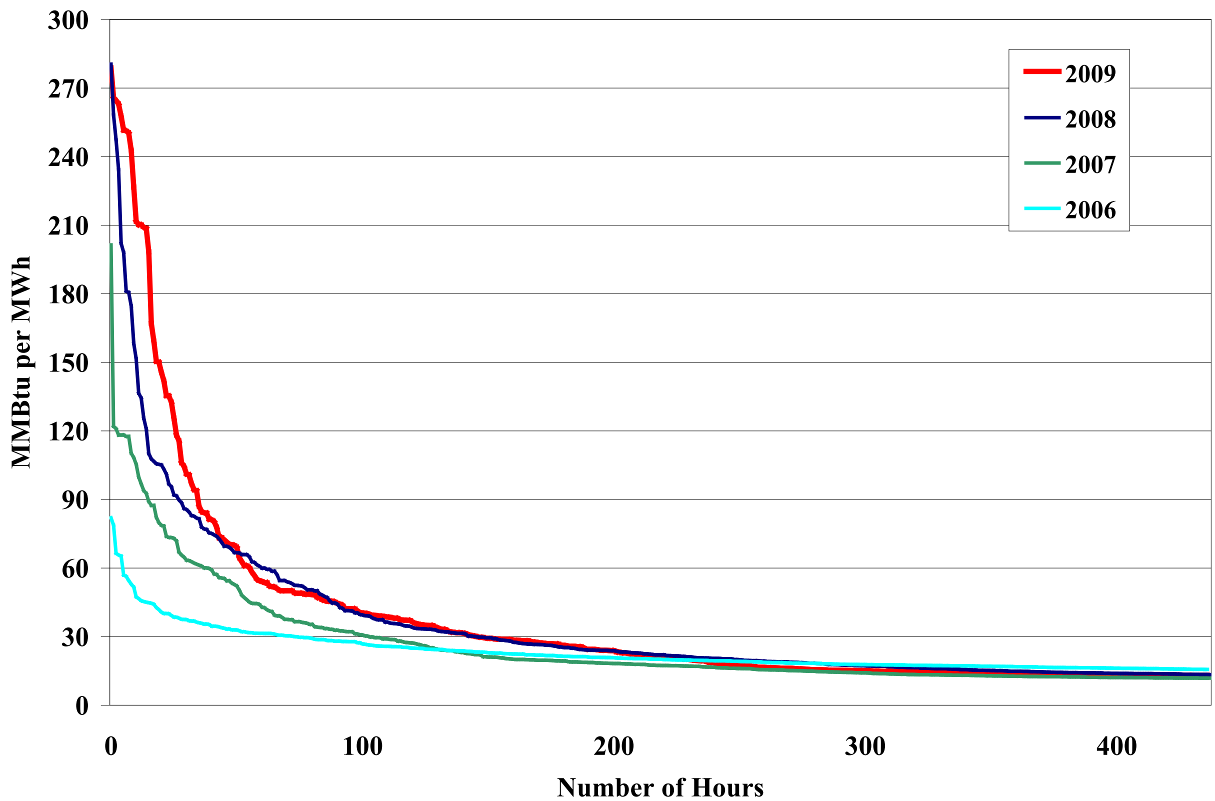
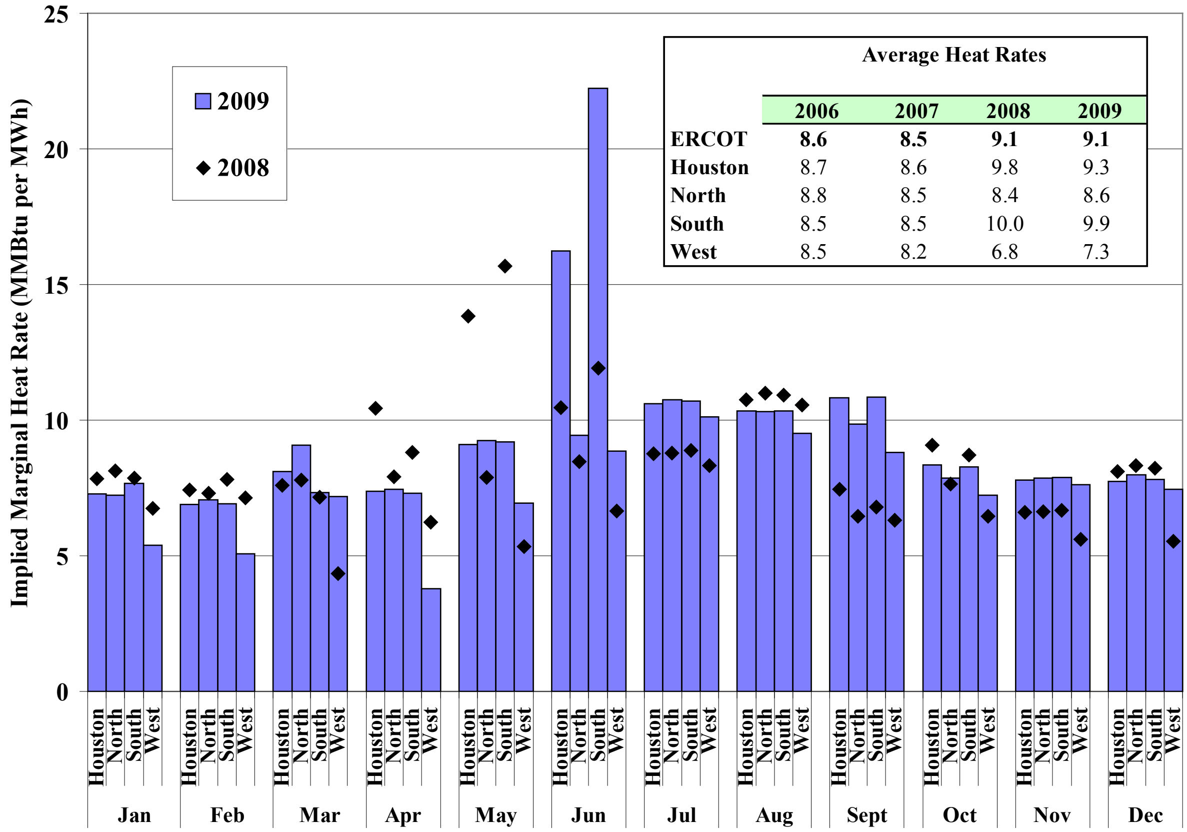


Figure 8: Implied Marginal Heat Rate Duration Curve – Top 5% of Hours



To better illustrate these differences, the next figure shows the implied marginal heat rates on a monthly basis in each of the ERCOT zones in 2008 and 2009, with annual average heat rate data for 2006 through 2009. This figure is the fuel price-adjusted version of Figure 1 in the prior sub-section. Adjusting for gas price influence, Figure 9 shows that average implied heat rate for all hours of the year was comparable in 2009 to 2008.

Figure 9: Monthly Average Implied Marginal Heat Rates



The average implied heat rate was significantly higher in 2008 than in 2009 during the months of April and May due to significant zonal congestion on the North to South and North to Houston interfaces that materialized in these months in 2008. Similarly, the magnitude of zonal congestion on the North to South interface increased significantly in late June 2009, causing the implied heat rate in June to be significantly higher in 2009 than in 2008. The implied heat rate in July was higher in 2009 than in 2008, primarily because of a stretch of extremely high temperatures and load levels, including the setting of a new record peak demand of 63,400 MW on July 13, 2009. Finally, the implied heat rate in September was much lower in 2008 than in

2009 because of the landfall of Hurricane Ike in September 2008 that resulted in widespread and prolonged loss of load in the Houston area.

3. Price Convergence

One indicator of market performance is the extent to which forward and real-time spot prices converge over time. In ERCOT, there is no centralized day-ahead market so prices are formed in the day-ahead bilateral contract market. The real-time spot prices are formed in the balancing energy market. 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.

These two conditions are largely satisfied in the current ERCOT market. Relaxed balanced schedules allow QSEs to increase and decrease their purchases in the balancing energy market. This flexibility should better enable them to arbitrage forward and real-time energy prices. While this should result in better price convergence, it should also reduce QSEs' total energy costs by allowing them to increase their energy purchases in the lower-priced market. However, volatility in balancing energy prices can create risks that affect convergence between forward prices and balancing energy prices. For example, risk-averse buyers are willing to pay a premium to purchase energy in the bilateral market thereby locking in their energy costs and avoiding the more volatile costs of the balancing energy market.

In this section, we measure two aspects of price convergence between forward and real-time markets. The first analysis investigates whether there are significant differences in prices between forward markets and the real-time market. The second tests whether there is a large spread between real-time and forward prices on a daily basis.

To determine whether there are significant differences between forward and real-time prices, we examine the difference between the average forward price and the average balancing energy price in each month between 2006 and 2009.[13] This analysis reveals whether persistent and predictable differences exist between forward and real-time prices, which participants should arbitrage over the long-term.

To measure the short-term deviations between real-time and forward prices, we also calculate the average of the absolute value of the difference between the forward and real-time price on a daily basis during peak hours. It is calculated by taking the absolute value of the difference between a) the average daily peak period price from the balancing energy market (*i.e.*, the average of the 16 peak hours during weekdays) and b) the day-ahead peak hour bilateral price. This measure captures the volatility of the daily price differences, which may be large even if the forward and balancing energy prices are the same on average. For instance, if forward 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 price difference between the forward 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. These two statistics are shown in Figure 10 for each month between 2006 and 2009.

13 Day-ahead bilateral prices as reported by Megawatt Daily are used to represent forward prices. For 2005-2007, we use the ERCOT Seller's Choice product. For 2008 and 2009, we use the average of the North, South and Houston Zone products.

Figure 10: Convergence between Forward and Real-Time Energy Prices

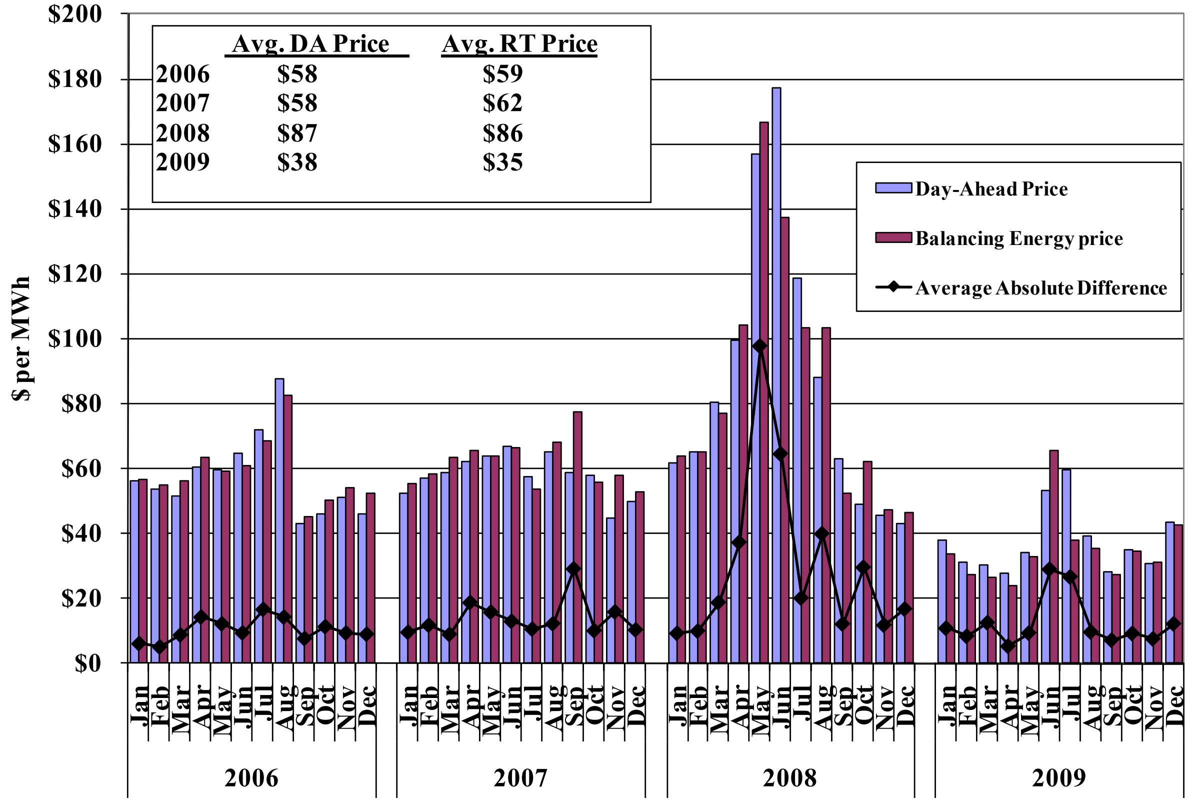


Figure 10 shows price convergence during peak periods (*i.e.*, weekdays between 6 AM and 10 PM). Day-ahead prices averaged $38 per MWh in 2009 compared to an average of $35 per MWh for real-time prices. Although the day-ahead and real-time prices exhibit relatively good average convergence in 2009, Figure 10 also shows that the average absolute price difference increased during the months of June and July 2009.

The average absolute difference was $10 in 2006, $14 in 2007, $31 in 2008 and $12 in 2009. As noted above, the average absolute difference measures the volatility of the price differences. Similar to the months of April, May and June 2008, the price volatility in June 2009 was due in large part to the significant and unpredictable transmission congestion experienced in that timeframe that caused average real-time prices to exceed day-ahead prices in June 2009. In contrast, average day-ahead prices were significantly higher than real-time prices in July 2009, which may be associated with transmission congestion expectations based on the experience in the prior month, as well as real-time pricing expectations associated with the extremely high temperatures and loads experienced during July 2009.

4. Volume of Energy Traded in the Balancing Energy Market

The primary purpose of the balancing energy market is to match supply and demand in real-time and to manage zonal congestion. In addition to fulfilling this purpose, the balancing energy market signals the value of power for market participants entering into forward contracts and plays a role in governing real-time dispatch. This section examines the volume of activity in the balancing energy market.

The average amount of energy traded in ERCOT's balancing energy market is small relative to overall energy consumption, although the balancing energy market can at times represent well over ten percent of total demand. Most energy is purchased and sold through forward contracts that insulate participants from volatile spot prices. Because forward contracting does not precisely match generation with real-time load, there will be residual amounts of energy bought and sold in the balancing energy market. Moreover, the balancing energy market enables market participants to make efficient changes from their forward positions, such as replacing relatively expensive generation with lower-priced energy from the balancing energy market.

Hence, the balancing energy market will improve the economic efficiency of the dispatch of generation to the extent that market participants make their resources available in the balancing energy market. In the limit, if all available resources were offered competitively in the balancing energy market (to balance up or down), prices in ERCOT's current market would be identical to prices obtained by clearing all power through a centralized spot market, even though most of the commodity currently settles bilaterally. It is rational for suppliers to offer resources in the balancing energy market even when they are fully contracted bilaterally because they may be able to increase their profit by reducing the output from their resources and support the bilateral sale with balancing energy purchases. Therefore the balancing energy market should govern the output of all resources, even though only a small portion of the energy is settled through the balancing energy market.

In addition to their role in governing real-time dispatch, balancing energy prices also provide a vital signal of the value of power for market participants entering into forward contracts. As discussed above, the spot prices emerging from the balancing energy market should directly affect forward contract prices, assuming that the market conditions and market rules allow the two markets to converge efficiently.

This section summarizes the volume of activity in the balancing energy market. Figure 11 shows the average quantities of up balancing and down balancing energy sold by suppliers in each month, along with the net purchases or sales (*i.e.*, up balancing energy minus down balancing energy).

Figure 11: Average Quantities Cleared in the Balancing Energy Market

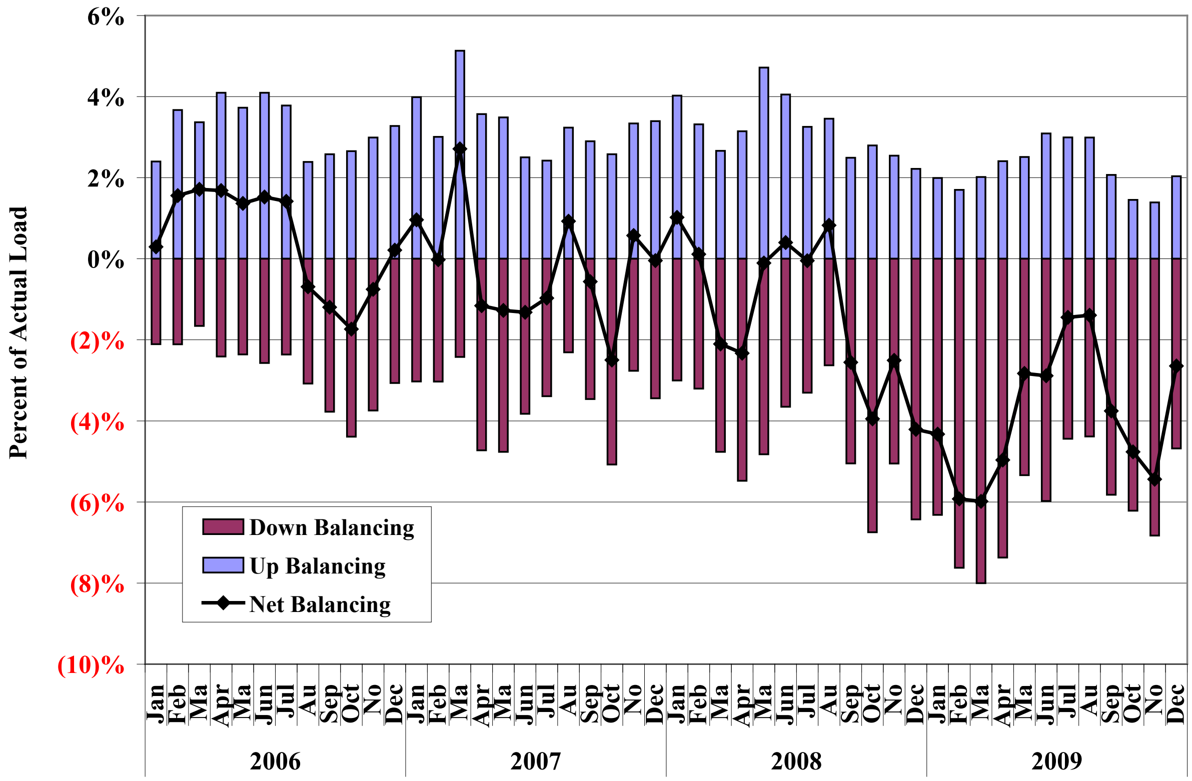
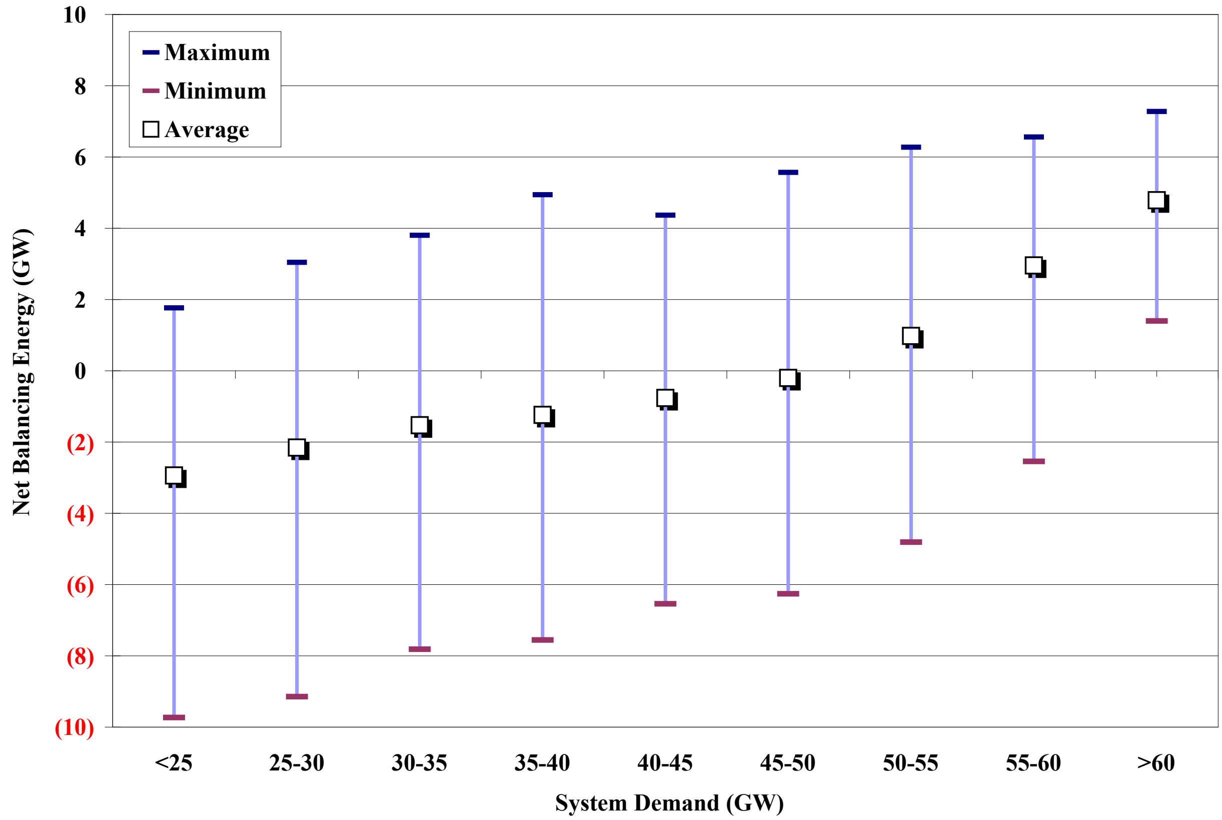


Figure 11 shows that the total volume of up balancing and down balancing energy as a share of actual load increased from an average of 7.7 percent in 2008 to 8.3 percent in 2009. Starting in August 2006, the average volume of down balancing energy began to increase. In 2008, for the first time the average amount of down balancing energy was greater than up balancing energy. This trend continued through 2009. The net quantity of balancing energy for every month in 2009 was negative, meaning that the average quantity of down balancing energy was greater than the quantity of up balancing energy. As discussed in Section II, this trend is related to the large increase in wind generation capacity added to the ERCOT region since the fall of 2008 and the associated scheduling patterns of these resources.

Figure 12 provides additional perspective to the monthly average net balancing energy deployments shown in Figure 11 by showing the net balancing energy deployments by load level for all intervals in 2009.

Figure 12: 2009 Net Balancing Energy by Load Level



While Figure 11 shows average net down balancing energy deployments in 2009, Figure 12 shows that this relationship is quite different when viewed as a function of the ERCOT system demand. Figure 12 shows average net down balancing deployments at load levels less than 50 GW, and average net up balancing deployments for load levels greater than 50 GW. Further, maximum net up balancing deployments exceeded 10 percent of demand at all system load levels in excess of 25 GW, except for levels exceeding 60 GW when net balancing deployments were exclusively in the upward direction.

Relaxed balanced schedules allow market participants to intentionally schedule more or less than their anticipated load, buying or selling in the balancing energy market to satisfy their actual load obligations. This scheduling flexibility allows the balancing energy market to operate as a centralized energy spot market. Although convergence between forward prices and spot prices has not been good on a consistent basis, the centralized nature of the balancing energy market facilitates participation in the spot market and improves the efficiency of the market results.

Aside from the introduction of relaxed balanced schedules, another reason for significant balancing energy quantities is that large quantities of up balancing and down balancing energy are often deployed simultaneously to clear "overlapping" balancing energy offers. Deployment of overlapping offers improves efficiency because it displaces higher-cost energy with lower-cost energy, lowering the overall costs of serving load and allowing the balancing energy price to more accurately reflect the marginal value of energy.

When large quantities of net up balancing or net down balancing energy are scheduled, it indicates that Qualified Scheduling Entities (QSEs) are systematically under-scheduling or over-scheduling load relative to real-time needs. If large hourly under-scheduling or over-scheduling occurs suddenly, the balancing energy market can lack the ramping capability (*i.e.*, how quickly on-line generation can increase or decrease its output) and sometimes the volume of energy offers necessary to achieve an efficient outcome. In these cases, large net balancing energy purchases can lead to transient price spikes when capacity exists to supply the need, but is not available in the 15-minute timeframe of the balancing energy market. The remainder of this sub-section and the next section will examine in detail the patterns of over-scheduling and under-scheduling that has occurred in the ERCOT market, and the effects that these scheduling patterns have had on balancing energy prices.

To provide a better indication of the frequency with which net purchases and sales of varying quantities are made from the balancing energy market, Figure 13 presents a distribution of the hourly net balancing energy. The distribution is shown on an hourly basis rather than by interval to minimize the effect of short-term ramp constraints and to highlight the market impact of persistent under- and over-scheduling. Each of the bars in Figure 13 shows the portion of the hours during the year when balancing energy purchases or sales were in the range shown on the x-axis. For example, the figure shows that the quantity of net balancing energy traded was between zero and positive 0.5 gigawatts (*i.e.*, loads were under-scheduled on average) in approximately 7 percent of the hours in 2009.

Figure 13: Magnitude of Net Balancing Energy and Corresponding Price

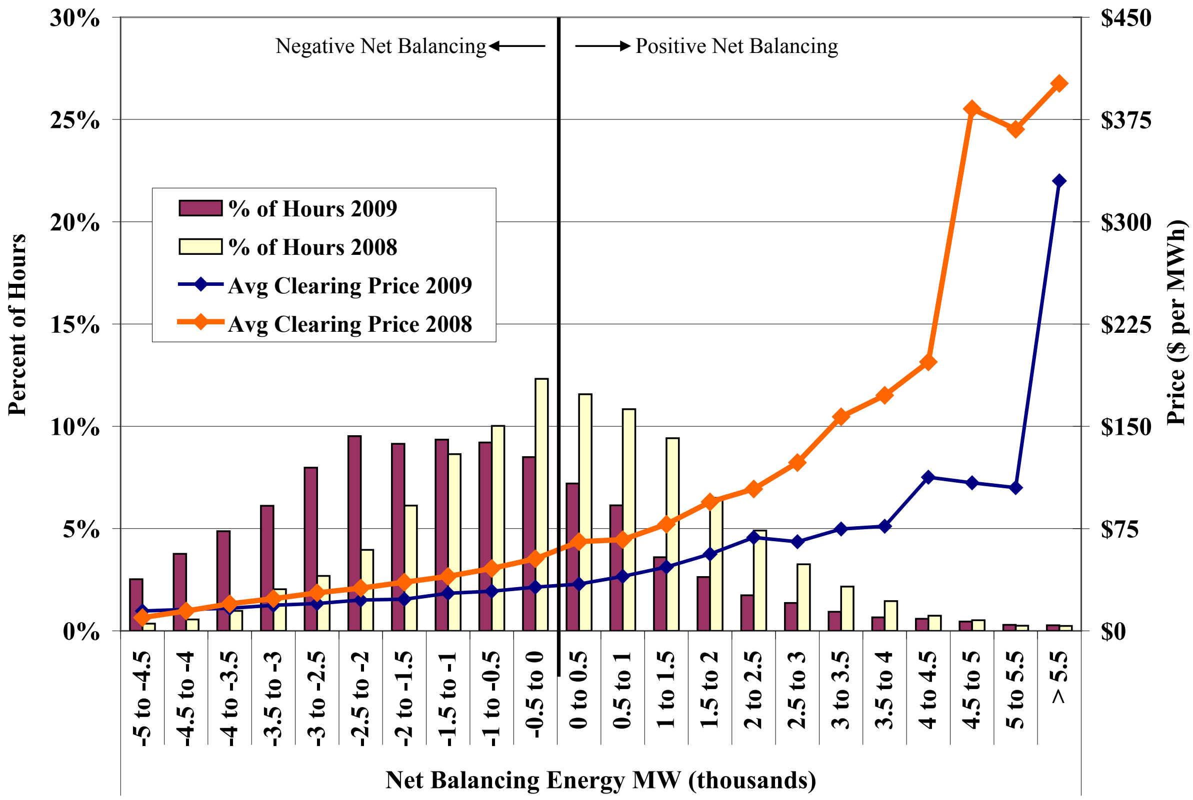


Figure 13 shows that the distribution of net balancing energy deployments in 2009 is shifted well to the left of zero, meaning that more down balancing energy was deployed than up balancing energy. This change in 2009 is consistent with the data shown in Figure 11, and is discussed in more detail in Section II. The lines plotted in Figure 13 show the average balancing energy prices corresponding to each level of balancing energy volumes for 2008 and 2009. In an efficiently functioning spot market, there should be little relationship between the balancing energy prices and the net purchases or sales. Instead, one should expect that prices would be primarily determined by more fundamental factors, such as actual load levels and fuel prices. However, this figure clearly indicates that balancing energy prices increase as net balancing energy volumes increase. This relationship is explained in part by the fact that net balancing energy deployments tend to be positively correlated with the level of demand as shown in Figure 12. However, scheduling practices and ramping issues contribute significantly to the observed pattern. We analyze this relationship more closely in the next subsections.

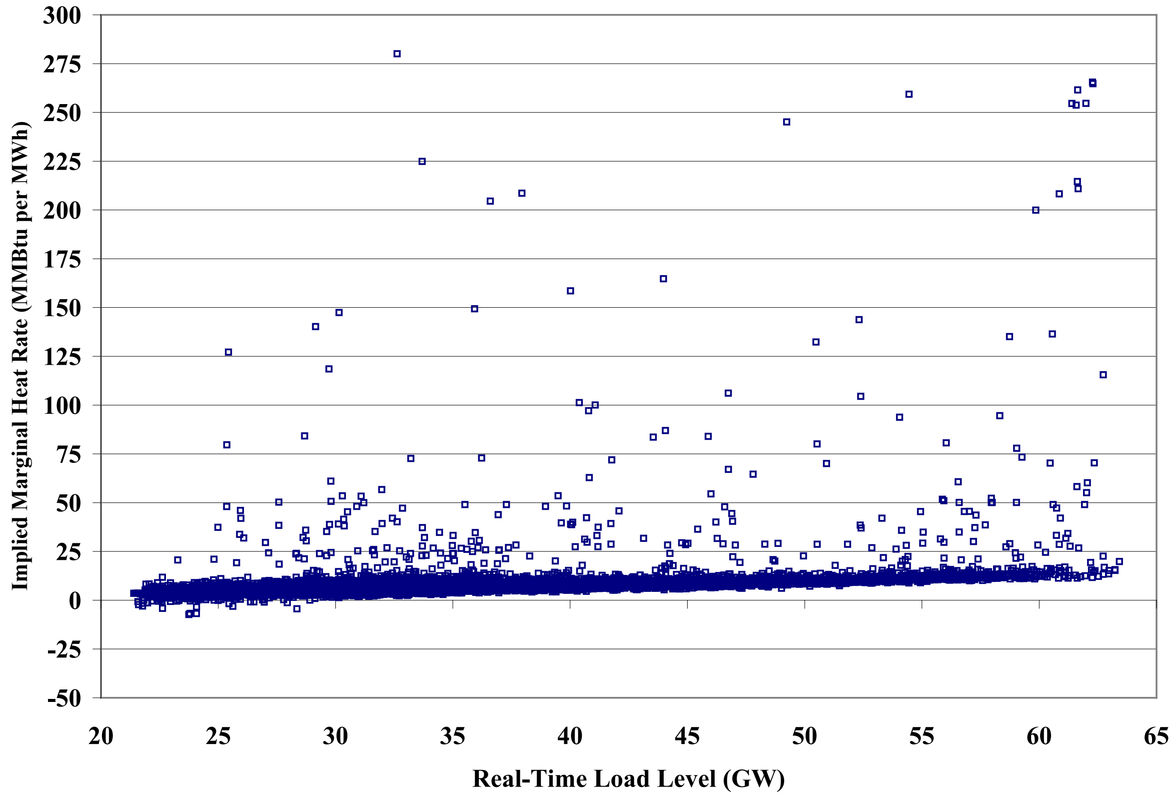
5. Determinants of Balancing Energy Prices

The prior section shows that the level of net sales in the balancing energy market appears to play a significant role in explaining the balancing energy prices. In this section, we examine this relationship in more detail, as well as the role of more fundamental determinants of balancing energy prices, such as the ERCOT load and fuel prices.

In an efficient market, we expect peak prices to occur under extreme demand conditions or as a result of unforeseen conditions that cause brief shortages, such as the loss of a large generator or an unanticipated rise in load. In ERCOT, prices in the balancing market can reach extremely high levels even when demand is not particularly high and absent such unforeseen operating conditions. This is primarily due to structural inefficiencies in the balancing energy market that are inherent to the zonal market model and the lack of a centralized unit commitment.

To further examine the relationship between actual load in ERCOT and balancing energy prices, Figure 14 shows the hourly average gas price-adjusted balancing energy prices versus the hourly average loads in ERCOT irrespective of time. This type of analysis shows more directly the relationship between balancing energy prices adjusted for natural gas prices and actual load. In a well-performing market, one should expect a clear positive relationship between these variables since resources with higher marginal costs must be dispatched to serve rising load.

Figure 14: Hourly Gas Price-Adjusted Balancing Energy Price vs. Real-Time Load



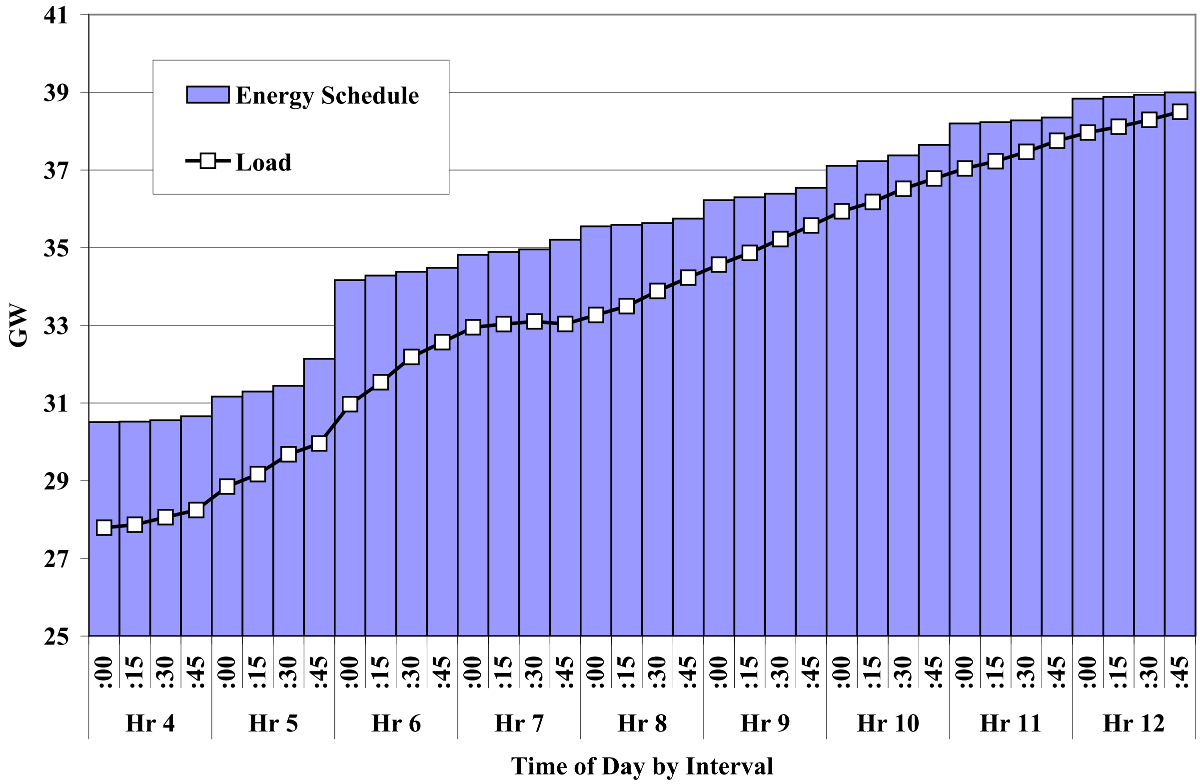
The figure indicates a positive correlation between real-time load and the clearing price in the balancing market. Although prices were generally higher at higher load levels, the data in Figure 13 indicate that the net volume of energy purchased in the balancing energy market is often a stronger determinant of price spikes than the level of demand.

6. Balancing Energy Market Scheduling

In the previous subsection, we analyzed balancing energy prices adjusted for fuel and load and found that while balancing energy prices are correlated to real-time load levels, other factors also have substantial effects on balancing energy levels. In this subsection, we investigate whether balancing energy prices are influenced by market participants' scheduling practices that tend to intensify the demand for balancing energy during hours when load is ramping.

We begin our analysis by examining factors that determine the demand for balancing energy during periods when load is ramping up and periods when it is ramping down. Figure 15 shows average energy schedules and actual load for each interval from 4 AM to 1 PM during 2009.

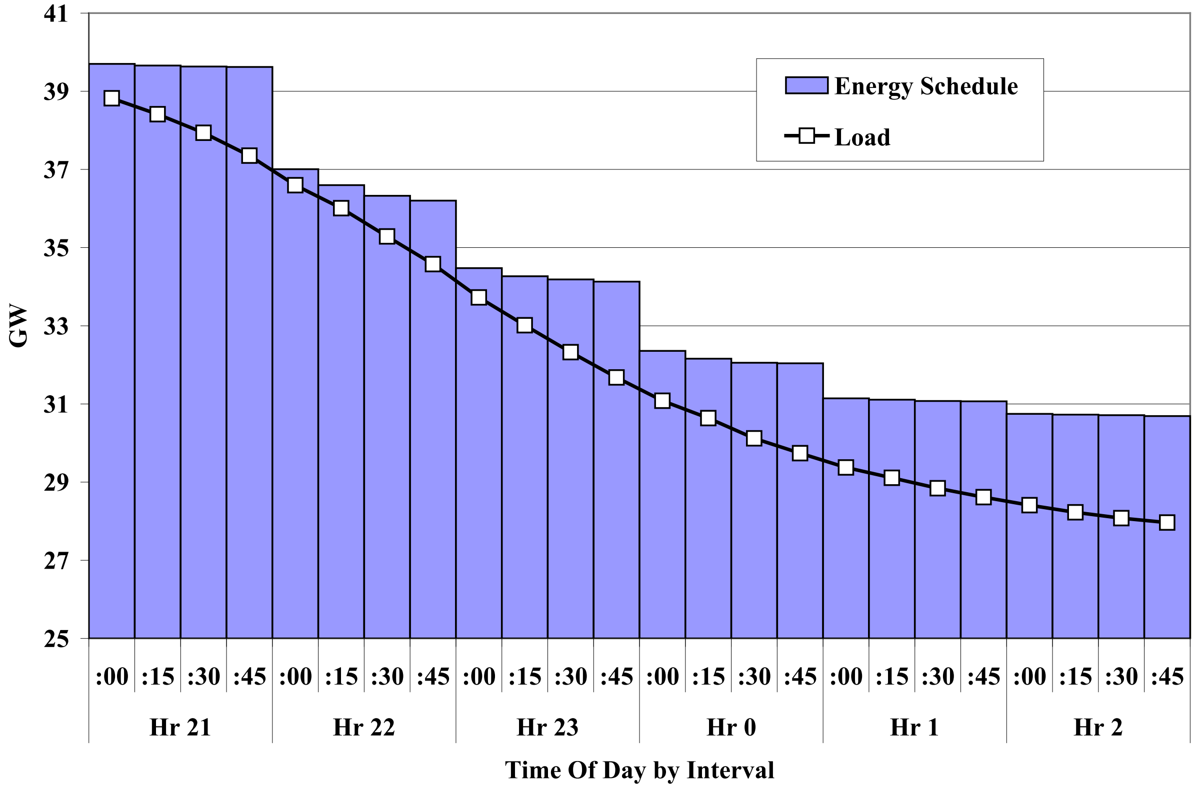
Figure 15: Final Energy Schedules during Ramping Up Hours



For ERCOT as a whole, energy schedules that are less than the actual load result in balancing energy purchases while energy schedules higher than actual load result in balancing energy sales. On average, load increases from approximately 28 GW to almost 39 GW in the nine hours shown in Figure 15, resulting in an average increase per 15-minute interval of approximately 330 MW.

The increase in load during ramping up hours is steady relative to the increase in energy schedules. Energy schedules rise less smoothly, with small increases from the first to fourth interval in each hour and larger increases from the fourth interval to the first interval of the next hour. For instance, the average energy schedule increases by more than 2.7 GW from the last interval of the hour ending 6 AM to the interval beginning at 6 AM, while the average energy schedule increases by only 160 megawatts in the subsequent three intervals. The same scheduling patterns exist in the ramping down hours. Figure 16 shows average energy schedules and load for each interval from 9 PM to 3 AM during 2009.

Figure 16: Final Energy Schedules during Ramping Down Hours



On average, load drops from approximately 39 GW to less than 29 GW in the six hours shown in Figure 16. The average decrease per 15-minute interval is 417 MW, although the rate of decrease is greatest from 9:45 PM to midnight. The progression of load during ramping down hours is steady relative to the progression of energy schedules. As was the case during ramping up hours, energy schedules change (decrease) in relatively large steps at the beginning of each hour. For example, the average energy schedule drops nearly 3.7 GW from the last interval before 10 PM to the interval beginning at 10 PM.

The sudden changes in energy schedules that occur at the beginning of each hour during ramping up hours and at the end of each hour during ramping down hours arise from the fact that much of the generation in ERCOT is scheduled by QSEs that submit energy schedules that change hourly. In addition, as indicated in Figure 15 and Figure 16, a number of schedules are based on bilateral contracts for 16-hour service, beginning as 6 AM and ending at 10 PM. Differences between energy schedules submitted by QSEs and load forecasted by ERCOT will result in purchases or sales in the balancing energy market. Specifically, the amount of net up balancing energy is equal to ERCOT's load forecast minus scheduled energy.

To evaluate the effects of systematic over- and under-scheduling more closely, we analyzed balancing energy prices and deployments in each interval during the ramping up period and ramping down period (consistent with the periods shown in Figure 15 and Figure 16). This analysis is similar to that shown in Figure 11 and Figure 12, except instead of showing balancing energy prices relative to load, we show balancing energy prices relative to net balancing energy deployments. Figure 17 shows the analysis for ramping up hours.

Figure 17: Balancing Energy Prices and Volumes Ramping Up Hours

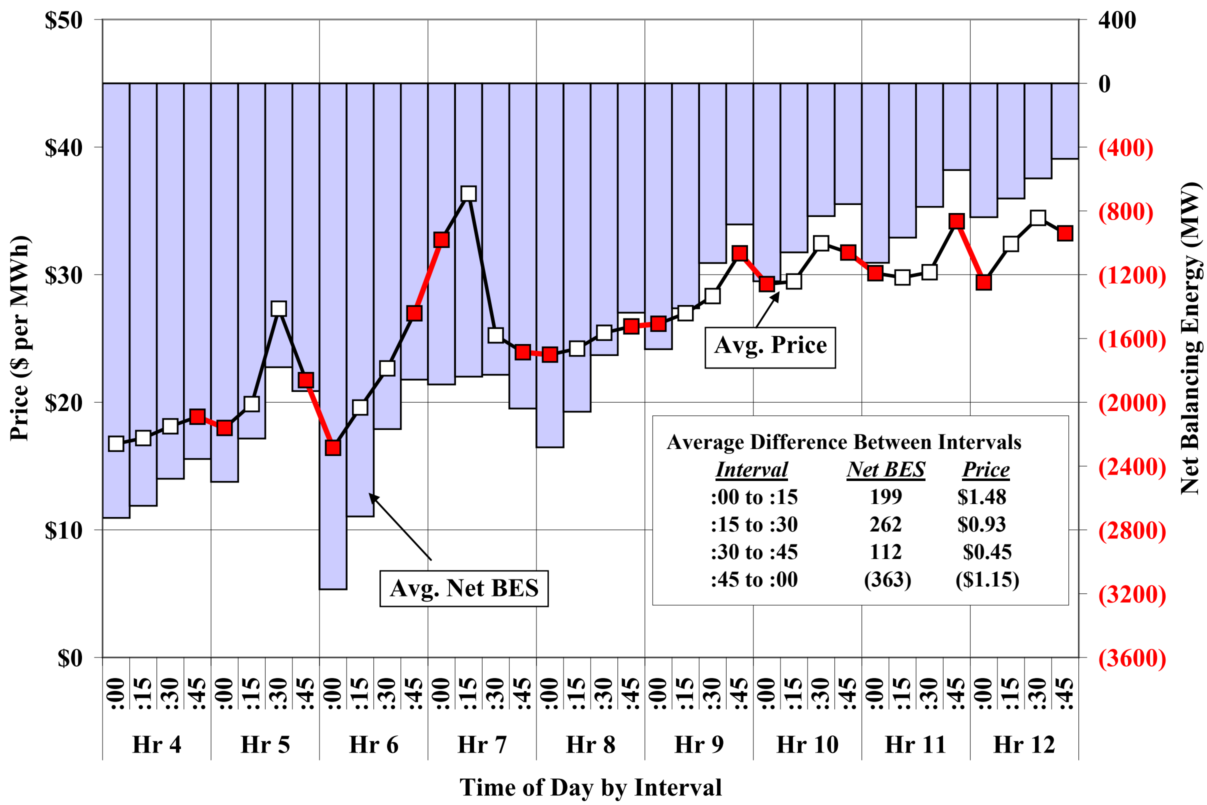


Figure 17 reveals two key aspects of the balancing energy market. First, as discussed above, balancing energy prices are highly correlated with balancing energy deployments. Second, with the exception of hour 7, there is a distinct pattern of increasing net balancing energy deployments during the hour. This is consistent with the notion that hourly schedules are established at a level that corresponds to an average expected load for the hour. The scheduling patterns that create these balancing deployments result in inefficient prices that are relatively volatile and could result in erratic dispatch signals to the generators.

Figure 18: Balancing Energy Prices and Volumes Ramping Down Hours

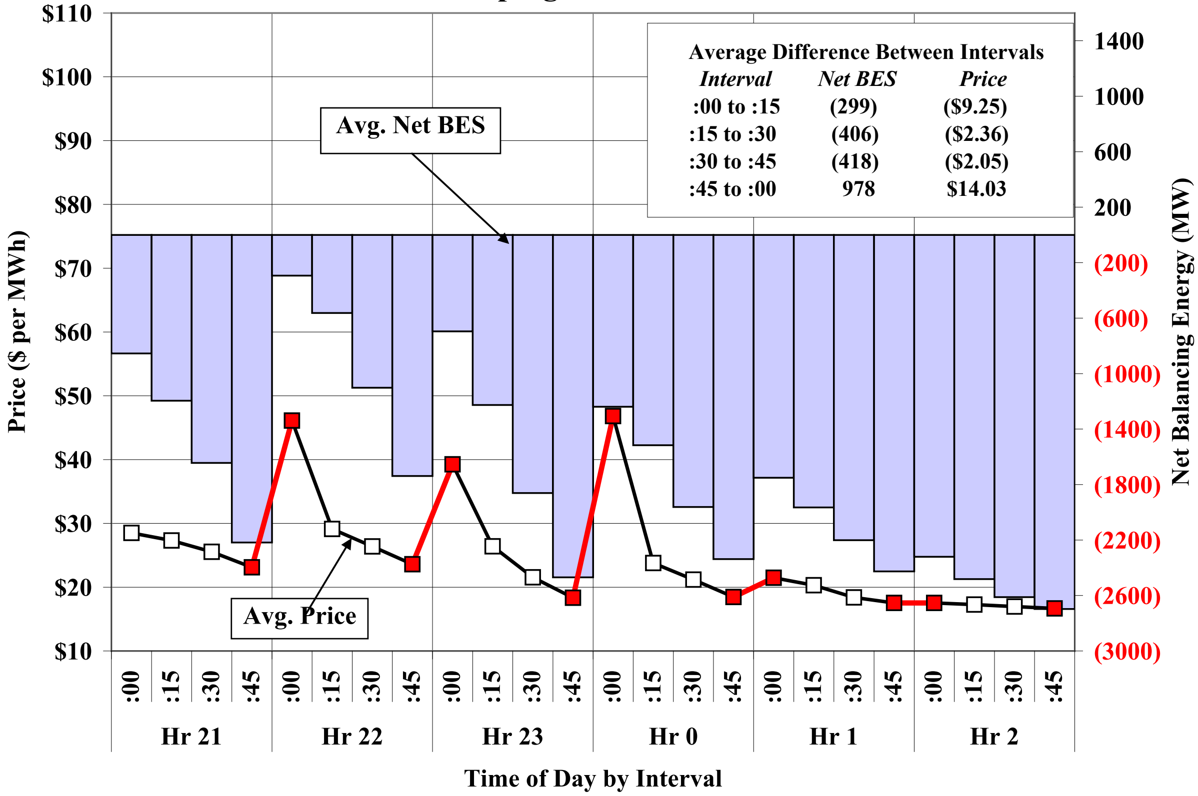


Figure 18 shows the same analysis for the ramping down hours. During ramping down hours, at the beginning of the hour, actual load tends to be higher than energy schedules, resulting in substantial balancing energy purchases. At the end of the hour actual load tends to be lower relative to the energy schedules, resulting in lower balancing energy demand.

To further examine how balancing energy prices relate to actual load levels, the final analysis in this subsection shows the average balancing energy prices by interval during the hours each day when load is increasing or decreasing rapidly (*i.e.*, when load is ramping up and ramping down). ERCOT load increases during the day from an average of almost 28 GW at 4 AM to 39 GW at 1 PM. Thus, the change in load averages 1,290 MW per hour (322 MW per 15-minute interval) during the morning and early afternoon. Figure 19 shows the average load and balancing energy price in each interval from 4 AM through 1 PM during 2009.

Figure 19: Average Balancing Energy Prices and Load by Time of Day Ramping Up Hours

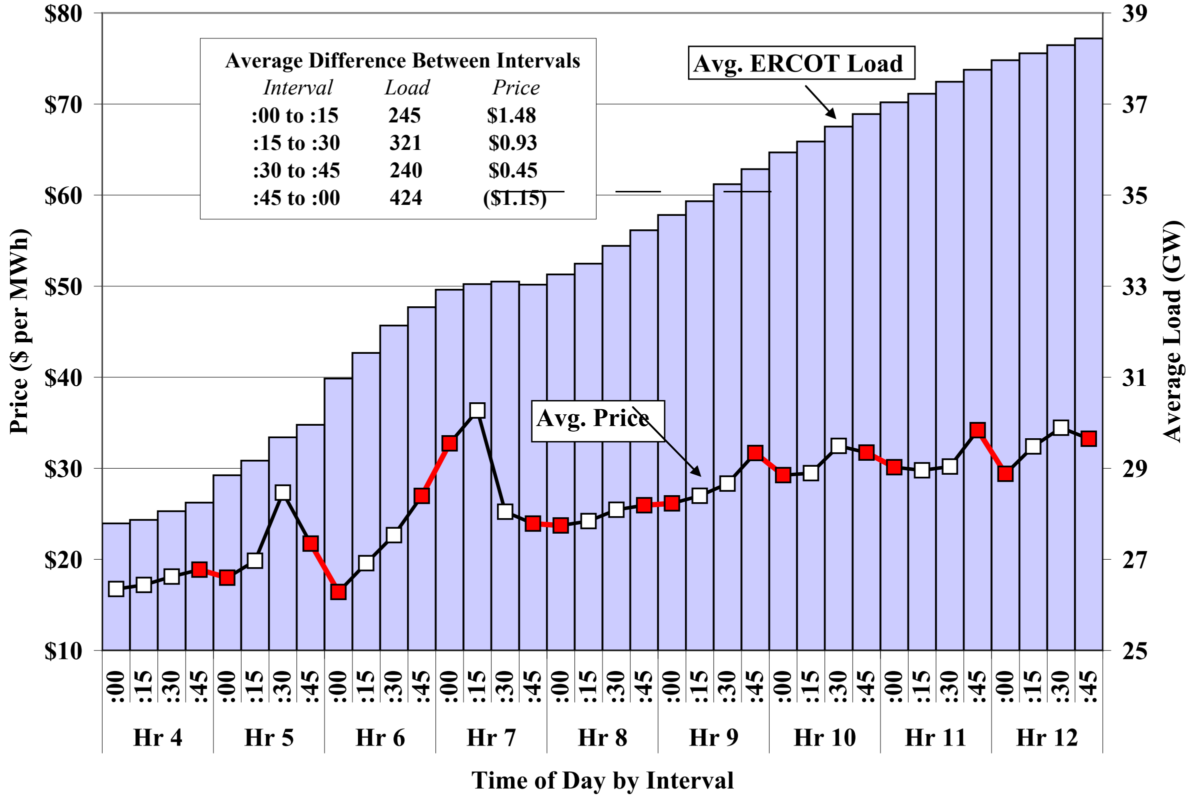


Figure 19 shows that, with the exception of hour 7, load steadily increases in every interval and prices generally move upward from an average of $18 per MWh at 4:00 AM to $32 per MWh at 12:45 PM. If actual load were the primary determinant of energy prices, the balancing energy prices would rise gradually as the actual load rises. However, Figure 19 shows this is not the case. In most hours the balancing energy price rises throughout each hour and drops substantially in the first interval of the next hour. In the figure, the red lines highlight the transition from one hour to the next hour. The average price change from the last interval of one hour to the first interval of the next hour is -$1.15 per MWh. This occurs because participants tend to change their schedules once per hour, bringing on additional substantial quantities of generation at the beginning of the hour which reduces the balancing energy prices.

A similar pattern is observed at the end of the day when load is decreasing. In ERCOT, load tends to decrease in the evening more quickly than it increases early in the day. Most of the decrease occurs over a six hour period, averaging a decrease of 1,891 MW per hour (473 MW per 15-minute interval) during the late evening. Figure 20 shows this decrease in load by interval, together with the average balancing energy prices for the intervals from 9 PM to 3 AM.

Figure 20: Average Balancing Energy Prices and Load by Time of Day Ramping Down Hours

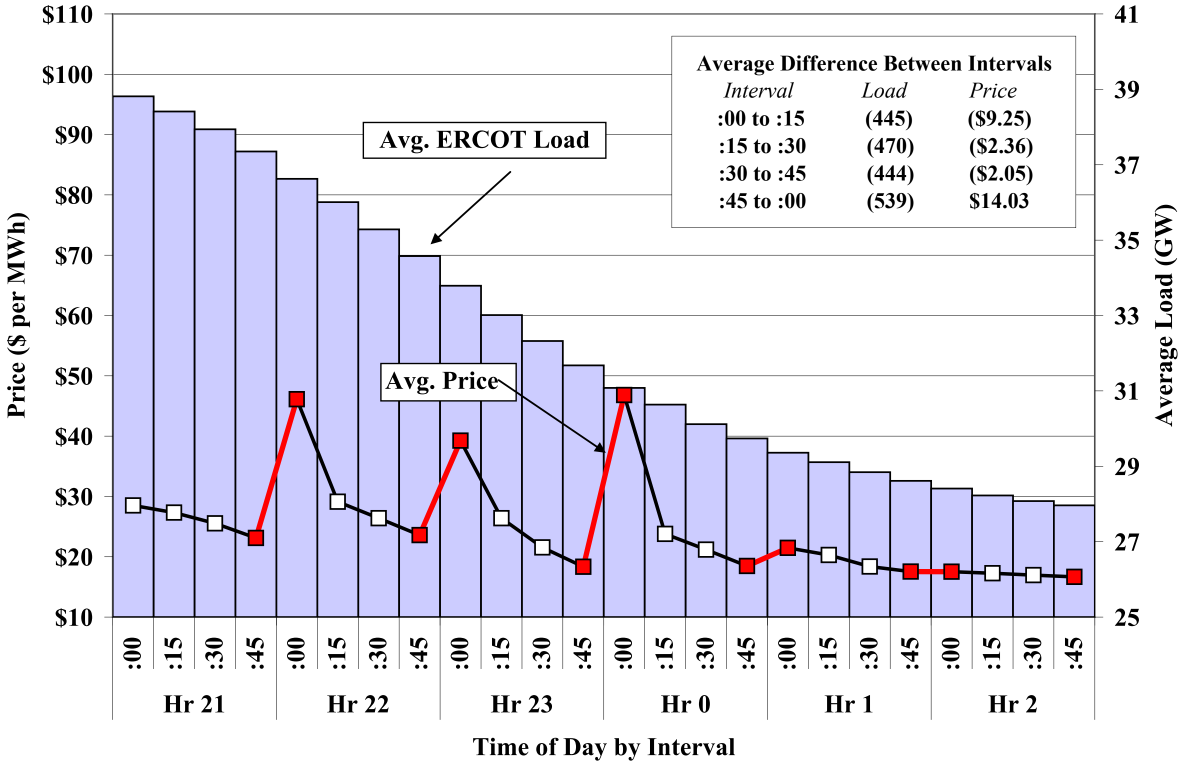


Figure 20 shows that while balancing energy prices decrease over these intervals, the pattern is similar to that exhibited in the ramping up hours. The balancing energy price decreases in each interval of the hour before rising substantially in the first interval of the following hour. The balancing energy price increases by an average of $14.03 per MWh from the last interval of one hour to the first interval of the next hour during this period. This occurs because participants tend to change their schedules once per hour, de-committing generating resources at the beginning of the hour. Because the supply decreases at the beginning of these hours by much more than load decreases, the balancing energy prices generally increase. This is consistent with the patterns of energy schedules and balancing prices in 2006 through 2008.[14]

[14] See 2006, 2007 and 2008 SOM Reports.

Collectively, these figures show that this pattern of balancing energy prices by interval is not explained by changes in actual load. Rather, changes in balancing energy deployments by interval underlie this pricing pattern. Sizable changes in balancing energy deployments occur between intervals, particularly in the first interval of the hour. These changes are associated with large hourly changes in energy schedules.

While QSEs have the option to submit schedules that change for every 15 minute interval, many QSEs schedule only on an hourly basis, making little or no changes on a 15-minute basis. It is primarily the scheduling patterns by the QSEs that schedule on an hourly basis that result in the balancing energy deployments and prices shown in Figure 17 and Figure 18.

The analysis in this section shows that one of the significant issues in the current ERCOT market is the tendency of most QSEs to alter their energy schedules hourly. This tendency may be related to the fact that balancing energy bids and offers are submitted hourly and are made relative to the energy schedule. For example, if a QSE schedules 200 MW from a 300 MW resource, it may offer the remaining 100 MW in the balancing energy market. If it schedules 230 MW, it may offer 70 MW. However, if the energy schedule changes on a 15-minute basis, it may be difficult to reconcile the schedule with the hourly balancing energy offer, leading most QSEs to simply submit hourly schedules. This places a burden on the balancing energy market to reconcile the differences between the hourly schedules and the 15-minute actual load levels, which can result in inefficient price fluctuations. This issue should not continue to be a problem under the nodal market design since resource-specific offers will not be interpreted as a deviation from an energy schedule.

As discussed in this subsection, a significant portion of the volatility of the balancing energy prices in each interval is related to the energy scheduling patterns. This volatility can be exacerbated when portfolio ramp rates are binding. Portfolio ramp rates are constraints QSEs submit with their balancing energy offers to limit the quantity of up balancing or down balancing energy that may be deployed in one interval. These ramp rates are important because they prevent a QSE from receiving deployment instructions that it cannot meet physically. Large changes in balancing energy deployments from interval to interval can cause the ramp rate constraints to bind, preventing the deployment of lower-cost offers and compelling the deployment of higher-cost offers from other QSEs. Ramp rate constraints can also be limiting when resources are instructed to ramp down quickly, although this is less common.

In many cases, the lack of ramp capable resources offered to the balancing energy market results in inefficient price spikes.[15] The efficiency implications associated with these issues continued in 2009 and will likely continue until the current zonal market design is replaced. However, ERCOT implemented 14 minute ramp rates in late October 2009 that are expected to help make more balancing energy ramping capability available, which in turn is expected to reduce the frequency and magnitude of price spikes associated with large schedule changes.[16]

B. Ancillary Services Market

The primary ancillary services are up regulation, down regulation, and responsive reserves. Market participants may self-schedule ancillary services or purchase their required ancillary services through the ERCOT markets. Historically, ERCOT has also procured non-spinning reserves as needed during periods of increased supply and demand uncertainty. However, beginning in November 2008, ERCOT began procuring non-spinning reserves across all hours based on its assessment of "net load" error, where "net load" is equal to demand minus wind production. This section reviews the results of the ancillary services markets in 2009.

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 at least 2,300 MW of responsive reserves to ensure adequate protection against the loss of the two largest units. Non-spinning reserves are procured as a means for ERCOT to implement supplemental generator commitments to increase the supply of energy in the balancing energy market if needed. The balancing energy market deployments that occur in the 15-minute timeframe and regulation deployments that occur in the 4-second timeframe are the primary means for meeting load fluctuations across and within each 15-minute interval.

---

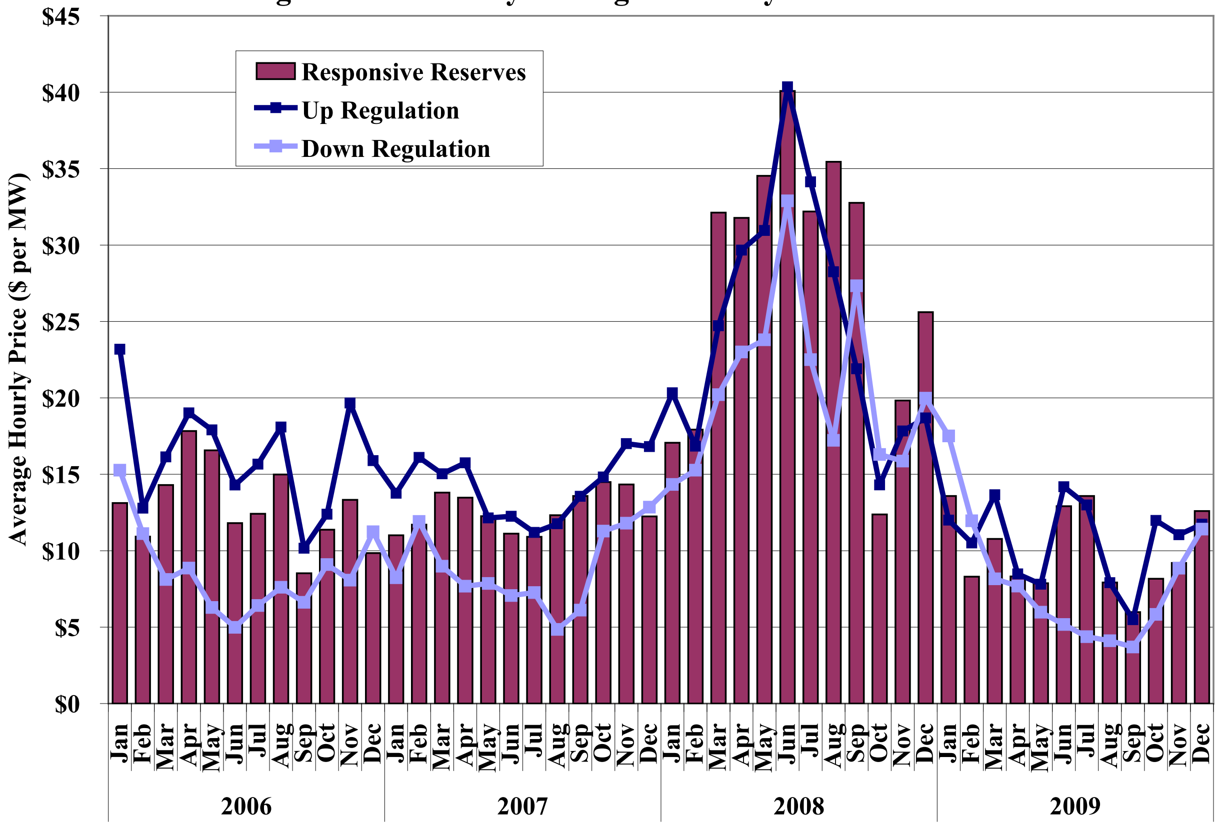
15 2005 SOM Report at 68-76.

16 There are insufficient data to perform an assessment of the effects of the 14-minute ramp implementation in 2009.

1. Reserves and Regulation Prices

Our first analysis in this section provides a summary of the ancillary services prices over the past four years. Figure 21 shows the monthly average ancillary services prices between 2006 and 2009. Average prices for each ancillary service are weighted by the quantities required in each hour.

Figure 21: Monthly Average Ancillary Service Prices



This figure shows that ancillary service capacity prices generally returned to levels seen in 2006 and 2007 after reaching significantly higher levels in 2008. These price movements can be primarily attributed to the variations in energy prices that occurred over the same timeframe.

Because ancillary services markets are conducted prior to the balancing energy market, participants must include their expected value of foregone sales in the balancing energy market in their offers for responsive reserves and regulation. Providers of both responsive reserves and up regulation can incur such opportunity costs if they reduce the output from economic units to make the capability available to provide these services. Likewise, providers of down regulation can incur opportunity costs in real-time if they receive instructions to reduce their output below the most profitable operating level. Further, because generators must be online to provide regulation and responsive reserves, there is an economic risk during low-price periods of operating uneconomically at minimum output levels (or having to operate above minimum output levels if providing down regulation).

Figure 21 shows that average down regulation prices have been lower than prices for up regulation service over the last four years, indicating that the opportunity costs were greater for providers of up regulation, with the exception of September 2008 through February 2009 when the average down regulation price was slightly higher than the average up regulation price.

Figure 21 also shows that, on average, the price of up regulation is slightly higher than the price of responsive reserves from 2006 through 2009. This is consistent with expectations because a supplier incurs opportunity costs to provide either service, while providing up regulation can generate additional costs. These additional costs include (a) the costs of frequently changing resource output levels, and (b) the risk of having to produce output when regulating at balancing energy prices that are less than the unit's variable production costs. However, during periods of persistent high prices, up regulation providers may have lower opportunity costs than responsive reserves providers to the extent that they are dispatched up to provide regulation. This factor explains in part the reversal in the relationship between responsive reserve and up regulation prices in 2008 when average responsive reserve prices were greater than or equal to average up regulation prices in seven out of twelve months.

One way to evaluate the rationality of prices in the ancillary services markets is to compare the prices for different services to determine whether they exhibit a pattern that is reasonable relative to each other. Table 1 compares the average prices for responsive reserves and non-spinning reserves over the past four years in those hours when ERCOT procured non-spinning reserves. Non-spinning reserves were purchased in approximately 20 and 14 percent of hours in 2006 and 2007, respectively, but increased to 51 percent of the hours in 2008. ERCOT began procuring non-spinning reserves in every hour beginning in November 2008, primarily to address the increasing uncertainty in net load associated with increasing levels of intermittent generation resources.

Table 1: Average Hourly Responsive Reserves and Non-Spinning Reserves Prices During Hours When Non-Spinning Reserves Were Procured

	2006	2007	2008	2009
Non-Spin Reserve Price	\$21.75	\$6.07	\$7.97	\$3.08
Responsive Reserve Price	\$25.55	\$16.74	\$36.39	\$9.68

Table 1 shows that responsive reserves prices are higher on average than non-spinning reserves prices during hours when non-spinning reserves were procured. It is reasonable that responsive reserves prices would generally be higher since responsive reserves are a higher quality product that must be delivered in 10 minutes from on-line resources while non-spinning reserves must be delivered in 30 minutes. Further, the significant reduction in the price of non-spinning reserves relative to responsive reserves beginning in 2007 was associated with the implementation of Protocol Revision Request ("PRR") 650, which significantly reduced the risk of uneconomic deployments for providers of non-spinning reserves, thereby reducing the capacity price for the provision of this service.

In contrast to the previous data that show the individual ancillary service capacity prices, Figure 22 shows the monthly total ancillary service costs per MWh of ERCOT load and the average balancing energy price for 2006 through 2009.

Figure 22: Ancillary Service Costs per MWh of Load

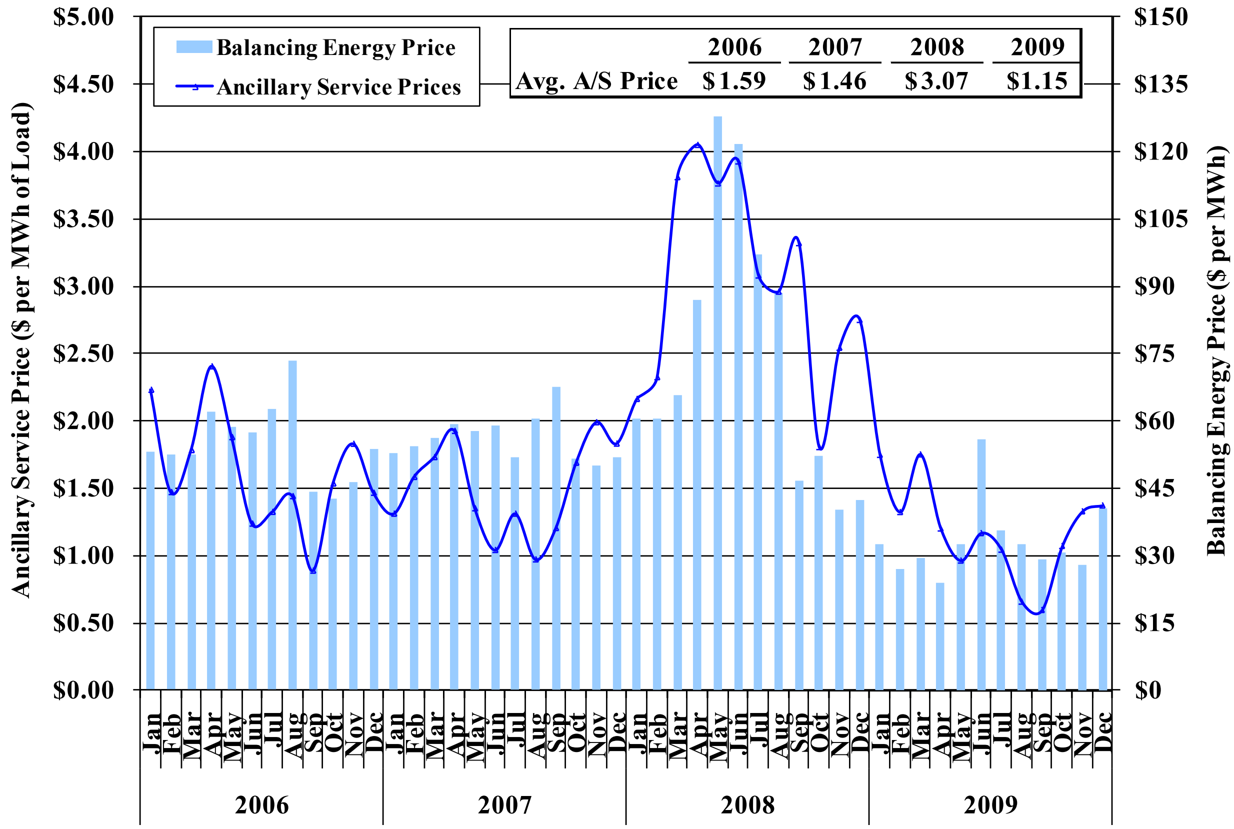


Figure 22 shows that total ancillary service costs are generally correlated with balancing energy price movements which, as previously discussed, are highly correlated with natural gas price movements. The average ancillary service cost per MWh of load decreased to $1.15 per MWh in 2009 compared to $3.07 per MWh in 2008, a decrease of more than 63 percent. Ancillary service costs were equal to 4.0 and 3.5 percent of the load-weighted average energy price in 2008 and 2009, respectively.

Our next analysis evaluates the variations in regulation prices. Regulation providers continuously vary their output levels to keep ERCOT-wide load and generation continually in balance during the time between SPD instructions, which are issued every fifteen minutes. When load and generation fluctuate by larger amounts, additional regulation resources are needed to keep the system in balance. This is particularly important in ERCOT due to the limited interconnections with adjacent areas, which results in much greater variations in frequency when generation does not precisely match load. Movements in load and generation are greatest when the system is ramping, thus ERCOT needs substantially more regulating capacity during ramping hours

Figure 23 shows the relationship between the quantities of regulation required by ERCOT and regulation price levels. This figure compares regulation prices to the average regulation quantity (both up and down regulation) procured, shown for each hour of the day. Regulation prices are weighted by the quantities of each service procured.

The figure shows that ERCOT requires approximately 1,350 MW of regulation capability prior to the initial ramping period (beginning at 6 AM). The requirement then increases to more than 1,900 MW during the steepest ramping hours from 6 AM to 9 AM. The requirement declines to about 1,400 MW during the late morning and afternoon hours when system load is relatively steady. From 6 PM until midnight, the system is ramping down rapidly and demand for regulation averages approximately 1,800 MW.

Figure 23: Regulation Prices and Requirements by Hour of Day

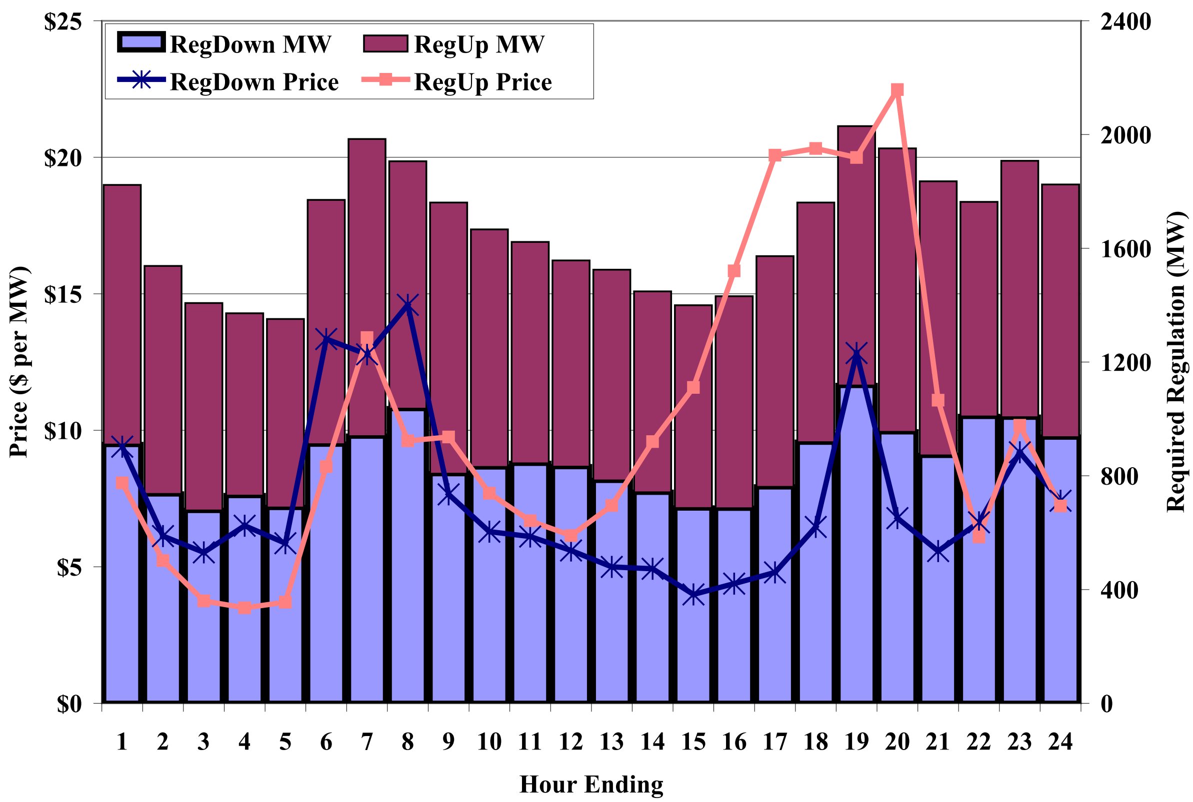


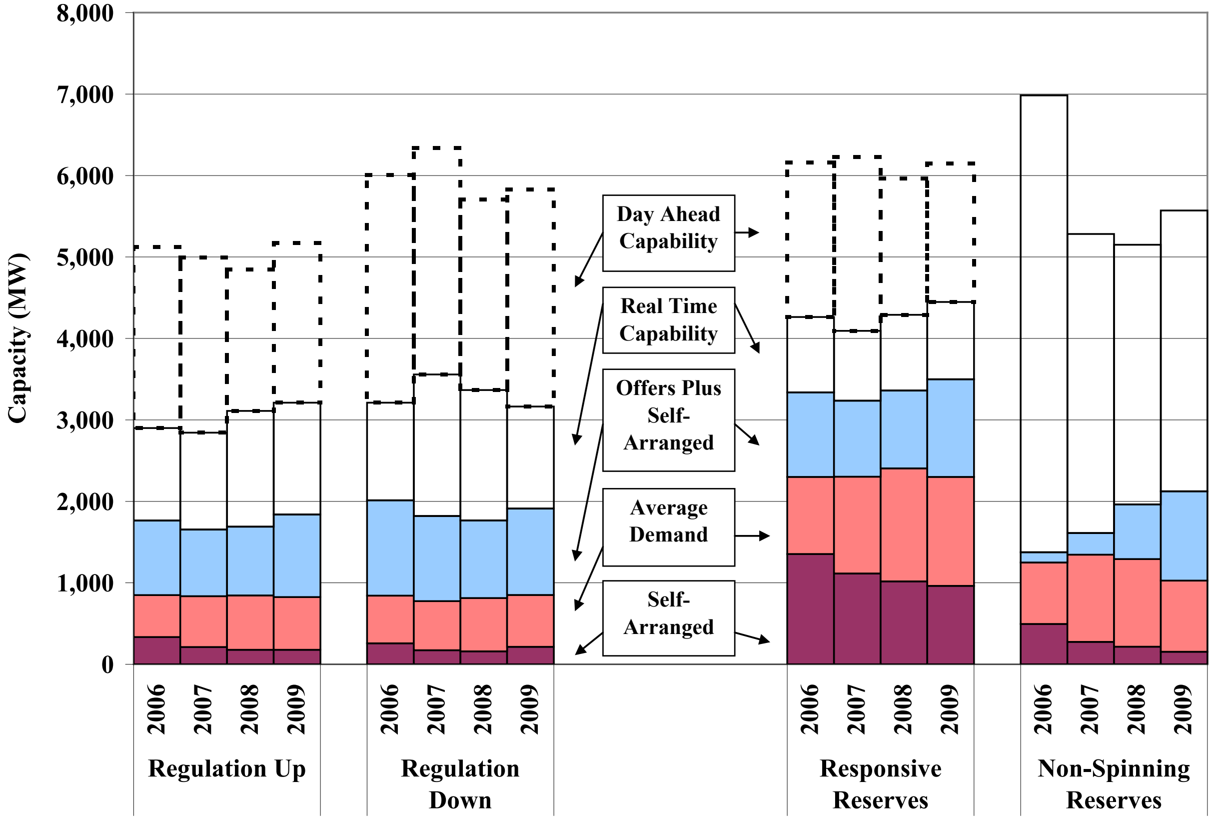
Figure 23 indicates that average regulation prices are generally correlated with the regulation quantity purchased and the typical load pattern in ERCOT. During non-ramping hours, such as overnight and late morning, up and down regulation prices are at their lowest levels. During the ramping hours in early morning average up and down regulation prices reached approximately \$11 per MW. During evening ramping hours, down regulation prices also reached \$7 per MW, while up regulation prices topped out at almost \$14 per MW. Up regulation prices are higher on average in the late afternoon hours because load levels and balancing energy prices are typically higher in these hours and the amount of capacity available to supply up regulation is lower than in other hours.

2. Provision of Ancillary Services

To better understand the reserve prices and evaluate the performance of the ancillary services markets, we analyze the capability and offers of ancillary services in this section. The analysis is shown in Figure 24. This figure summarizes the quantities of ancillary services offered and self-arranged relative to the total capability and the typical demand for each service. The bottom segment of each bar in Figure 24 is the average quantity of ancillary services self-arranged by owners of resources or through bilateral contracts. The second segment of each bar is the average amount offered and cleared in the ancillary services market. Hence, the sum of the first two segments is the average demand for the service.

The third segment of each bar is the quantity offered into the auction market that is not cleared. Therefore, the sum of the second and third segments is the total quantities offered in each ancillary services auction on average, including the quantities cleared and not-cleared. The empty segments correspond to the ancillary services capability that is not scheduled or offered in the ERCOT markets. The lower part of the empty segments correspond to the amount of real-time capability that is not offered while the top part of the empty segments correspond to the additional quantity available in the day-ahead that was not offered. Capabilities are generally lower in the real-time because offline units that require significant advance notice to start-up will not be capable of providing responsive reserves or regulation in real time (only capability held on online resources is counted).

Figure 24: Reserves and Regulation Capacity, Offers, and Schedules



The capability shown in Figure 24 incorporates ERCOT's requirements and restrictions for each type of service. For regulation, the capability is calculated based on the amount a unit can ramp in five minutes for those units that have the necessary equipment to receive automatic generation control signals on a continuous basis. For responsive reserves, the capability is calculated based on the amount a unit can ramp in ten minutes. This is limited by an ERCOT requirement that no more than 20 percent of the capacity of a particular resource is allowed to provide responsive reserves. However, the responsive reserve capability shown in Figure 24 is not reduced to account for energy produced from each unit, which causes the capability on some resources to be overstated in some hours.

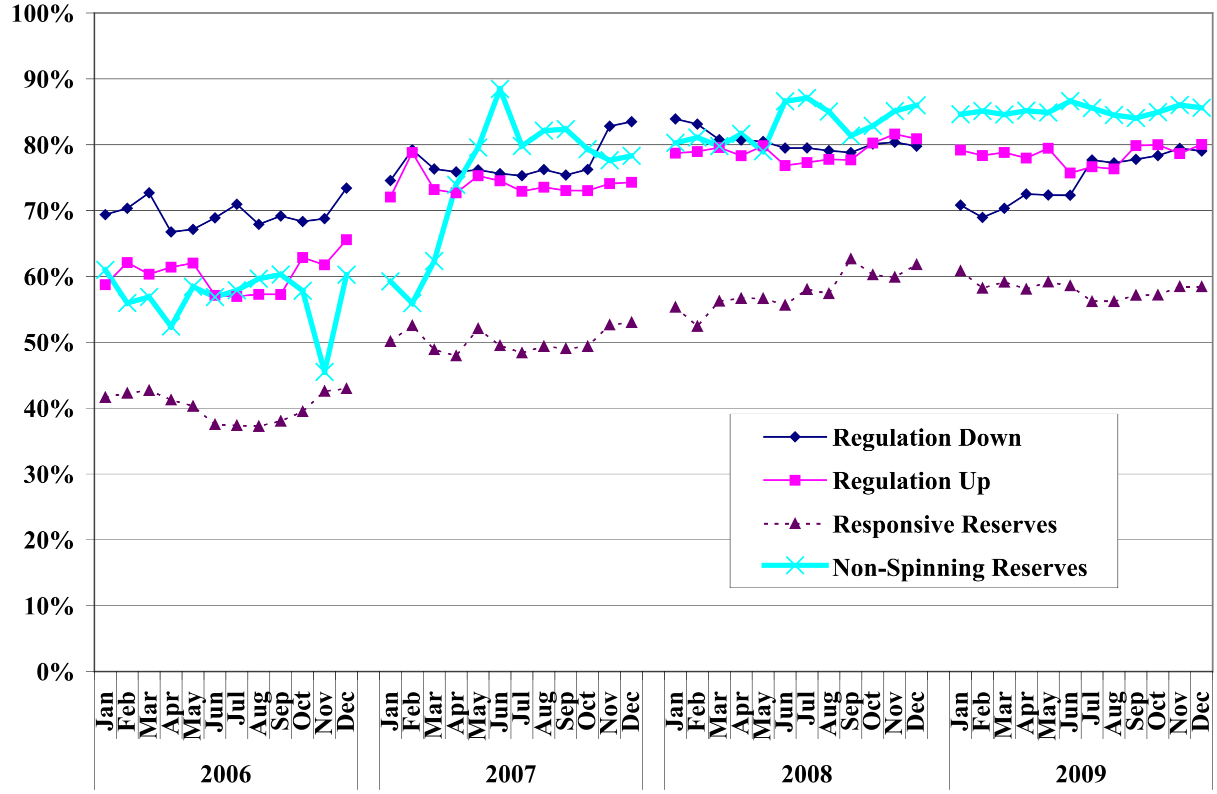
For non-spinning reserves, Figure 24 includes the capability of units that QSEs indicate are able to ramp-up in thirty minutes and able to start-up on short notice. The total capability shown in this figure does not account for capacity of online resources. However, it should be noted that any on-line resource with available capacity can provide non-spinning reserves, so the actual capability is larger than shown in the figure.

Figure 24 shows that except for responsive reserves, for which approximately 55 percent of available responsive reserve capacity was offered, less than one-half of each type of ancillary services capability was offered during the year from 2006 to 2009. One explanation for these levels of offers is that the ancillary services markets are conducted ahead of real time so participants may not offer resources that they expect to dispatch to serve their load or to support sales in the balancing energy market. In other words, some of the available reserves and regulation capability becomes unavailable in real time because the resources are dispatched to provide energy. The current market design creates risk and uncertainty for suppliers who must predict one day in advance whether their resources will be more valuable as energy or as ancillary services.

In addition, participants may not offer the capability of resources they do not expect to commit for the following day. Suppliers could submit offer prices high enough to ensure that their costs of committing additional resources to support the ancillary services offers are covered. However, under the current market design, ancillary services are procured independently for each hour and not optimized over the entire day (e.g., including minimum run times and minimum quantities), which greatly increases the risk for generators. The nodal market will include co-optimized procurement of energy and reserves over the entire operating day, which should enhance the efficiency of the procurement of reserves.

These services can be self-supplied from owned resources or from resources purchased bilaterally. To evaluate the quantities of ancillary services that are not self-supplied more closely, Figure 25 shows the share of each type of ancillary service that is purchased through the ERCOT market.

Figure 25: Portion of Reserves and Regulation Procured Through ERCOT

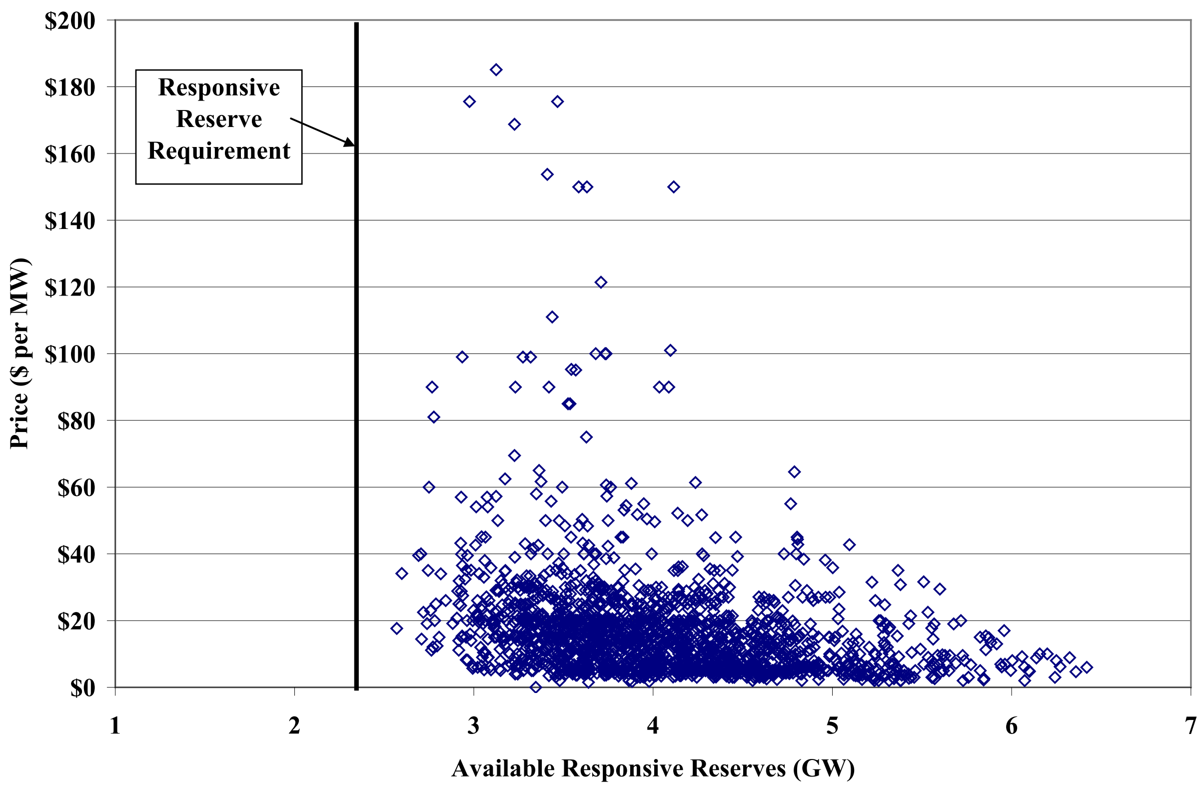


As market participants have gained more experience with the ERCOT markets, larger portions of the available reserves and regulation capability have been offered into the market, thereby increasing the market's liquidity. Nevertheless, Figure 25 shows that a fair share of these services is still self-supplied, particularly responsive reserves.

Prices in the ERCOT responsive reserve market tend to be somewhat higher than in other markets that co-optimize the procurement and dispatch of energy and responsive reserves. Responsive reserve prices in the ERCOT market are also affected by relatively higher requirements than other markets, as well as reliability restrictions that limit the quantity of responsive reserves that can be provided by each generating unit. Lower prices occur in co-optimized markets because the procurement is optimized with energy over the entire operating day and in most hours there is substantial excess online capacity that can provide responsive reserves at very low incremental costs. For example, a steam unit that is not economic to operate at its full output in all hours will have output segments that can provide responsive reserves at very low incremental costs. If the surplus responsive reserves capability from online resources is relatively large in some hours, one can gauge the efficiency of the ERCOT reserves market by evaluating the prices in these hours.

Figure 26 plots the hourly real-time responsive reserves capability against the responsive reserves prices during the peak afternoon hours of 2 PM to 6 PM. The capability calculated for this analysis reflects the actual energy output of each generating unit and the actual dispatch point for LaaRs. Hence, units producing energy at their maximum capability will have no available responsive reserves capability and, consistent with ERCOT rules, the responsive reserve that can be provided by each generating unit is limited to 20 percent of the unit's maximum capability. The figure also shows the responsive reserves requirement of 2,300 MW in 2009 to show the amount of the surplus in each hour.

Figure 26: Hourly Responsive Reserves Capability vs. Market Clearing Price Afternoon Peak Hours



In a well functioning-market for responsive reserves, we would expect excess capacity to be negatively correlated with the clearing prices. The data in this figure indicate only a weak negative correlation. Particularly surprising is the frequency with which price exceeds $20 per MW when the responsive reserve capability is more than 2,000 MW higher than the requirement.

In these hours the marginal costs of supplying responsive reserves should be very low. These results reinforce the potential benefits which should result from jointly optimizing the operating reserves and energy markets. The upcoming nodal market implementation will include day ahead co-optimization, but not real-time.

II. Demand and Resource Adequacy

The first section of this report reviewed the market outcomes and provided analyses of a variety of factors that have influenced the market outcomes. This section reviews and analyzes the load patterns during 2009 and the existing generating capacity available to satisfy the load and operating reserve requirements.

A. ERCOT Loads in 2009

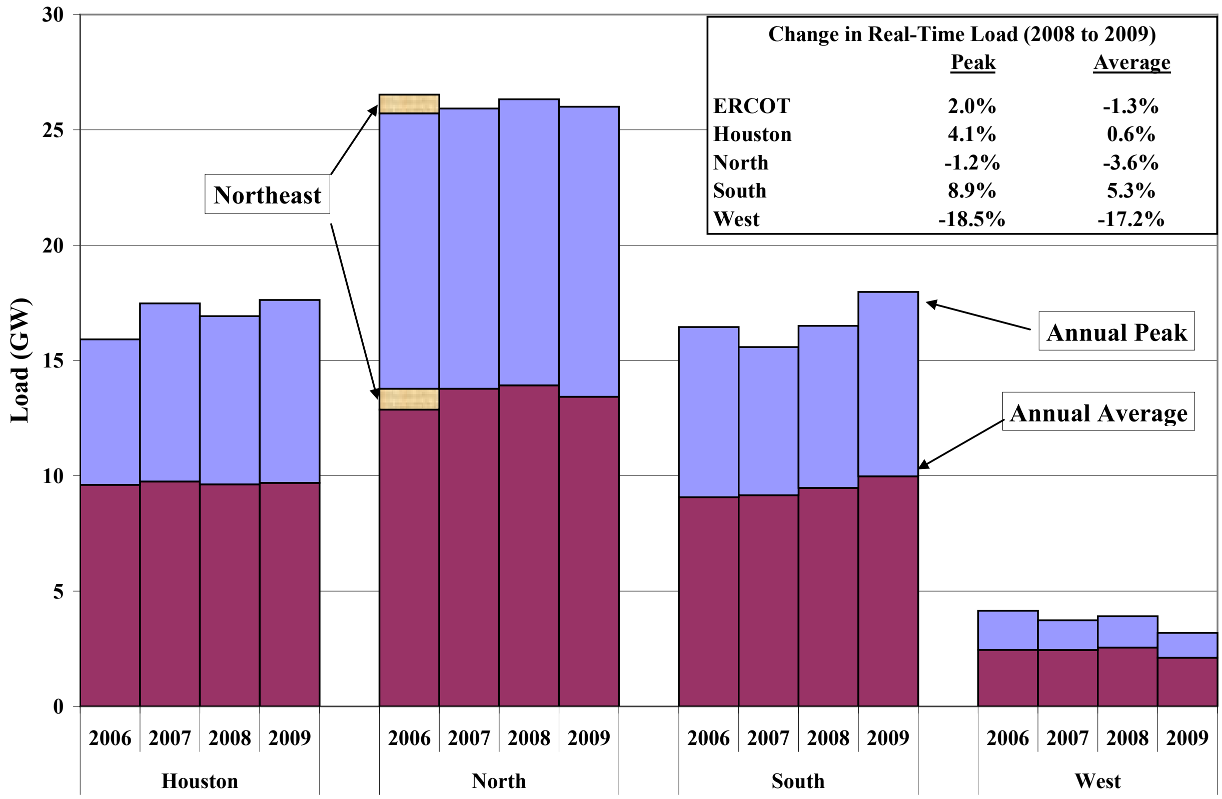
There are two important dimensions of load that should be evaluated separately. First, the changes in overall load levels from year to year can be shown by tracking the changes in average load levels. This metric will tend to capture changes in load over a large portion of the hours during the year. Second, it is important to separately evaluate the changes in the load during the highest-demand hours of the year. Significant changes in these peak demand levels have historically been very important and played a major role in assessing the need for new resources. The expectation in a regulated environment was that adequate resources would be acquired to serve all firm load, and this expectation remains in the competitive market. The expectation of resource adequacy is based on the value of electric service to customers and the damage and inconvenience to customers that can result from interruptions to that service. Additionally, significant changes in peak demand levels affect the probability and frequency of shortage conditions (*i.e.*, conditions where firm load is served but required operating reserves are not maintained). Hence, both of these dimensions of load during 2009 are examined in this subsection and summarized in Figure 27.

This figure shows peak load and average load in each of the ERCOT zones from 2006 to 2009. It indicates that in each zone, as in most electrical systems, peak demand significantly exceeds average demand. The North Zone is the largest zone (about 38 percent of the total ERCOT load);[17] the South and Houston Zones are comparable (with about 28 percent) while the West Zone is the smallest (with about 6 percent of the total ERCOT load). Figure 27 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

[17] The Northeast Zone was integrated into the North Zone in 2007.

zones. As a result, the sum of the non-coincident peaks for the zones was greater than the annual ERCOT peak load.

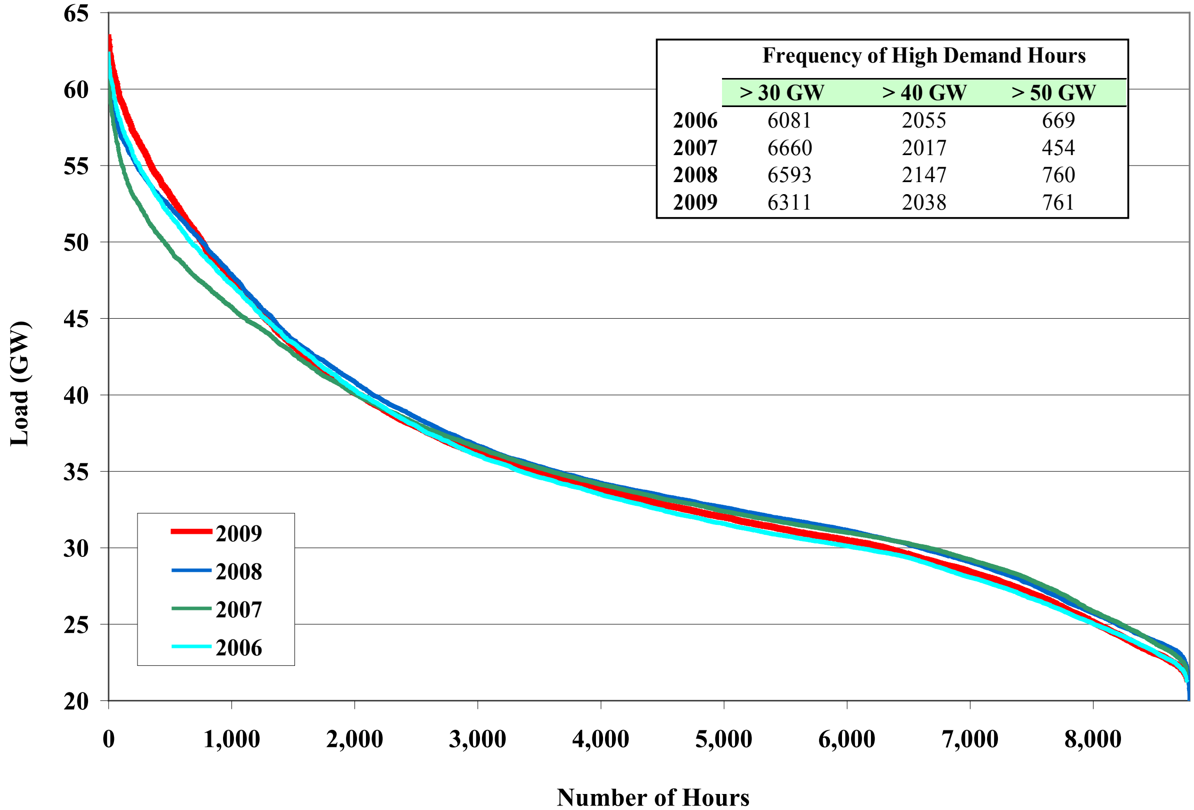
Figure 27: Annual Load Statistics by Zone



Some of the changes in zonal peak and average loads from 2008 to 2009 can be attributed to changes to the zonal definitions that resulted in some loads moving to a different zone in 2009. Overall, the ERCOT average load decreased from 312,401 GWh in 2008 to 308,278 GWh in 2009, a decrease of 1.3 percent. In contrast, the ERCOT coincident peak demand increased from 62,174 MW in 2008 to 63,400 MW in 2009, an increase of 2.0 percent.

To provide a more detailed analysis of load at the hourly level, Figure 28 compares load duration curves for each year from 2006 to 2009. 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, as most hours exhibit low to moderate electricity demand, with peak demand usually occurring during the afternoon and early evening hours of days with exceptionally high temperatures.

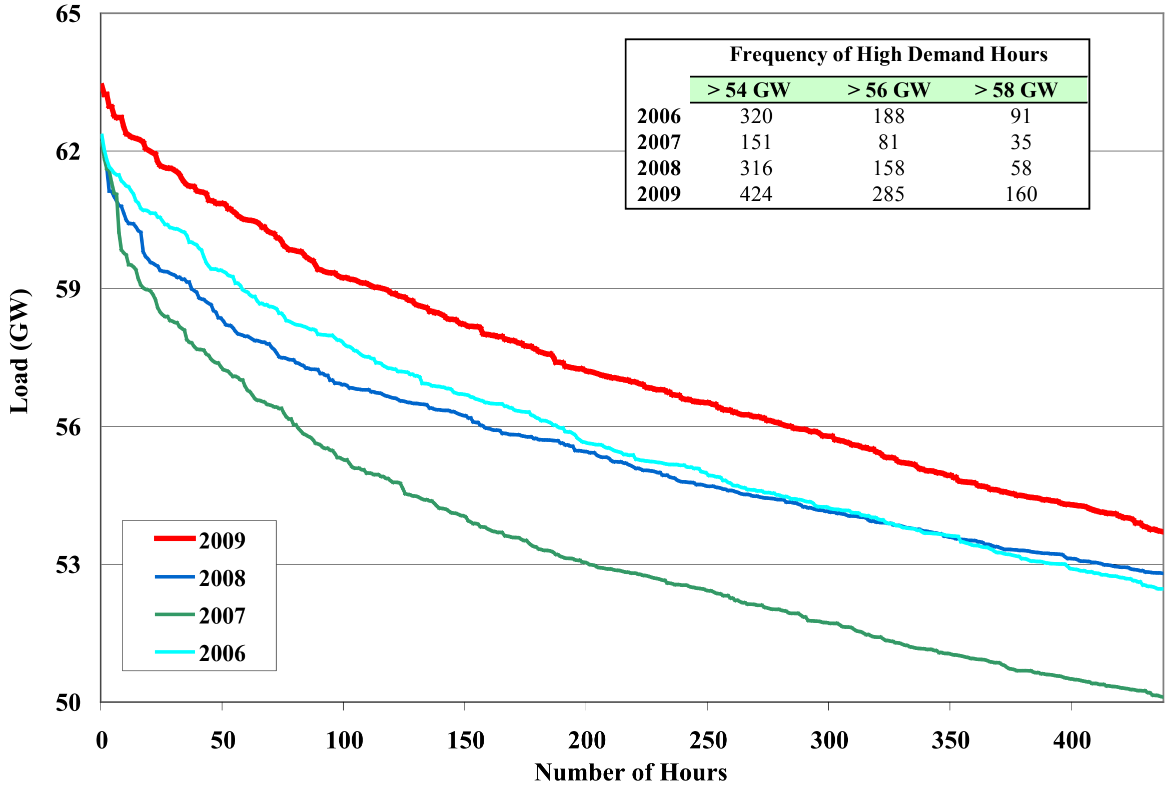
Figure 28: ERCOT Load Duration Curve – All Hours



As shown in Figure 28, the load duration curve for 2009 is slightly lower than in 2008 at load levels less than 45 GW, which accounts for approximately 85 percent of the hours in 2009 and is consistent with the load reduction of 1.3 percent from 2008 to 2009. However, the number of high demand hours (more than 50 GW) in 2008 and 2009 are at comparable levels (760 and 761 hours respectively).

To better show the differences in the highest-demand periods between years, Figure 29 shows the load duration curve for the five percent of hours with the highest loads. This figure shows that while average load increased in each year from 2006 to 2008 and decreased in 2009, the frequency of high-demand hours in 2009 increased compared with year 2008. Load exceeded 58 GW in 160 hours in 2009, more than double the hours in 2008.

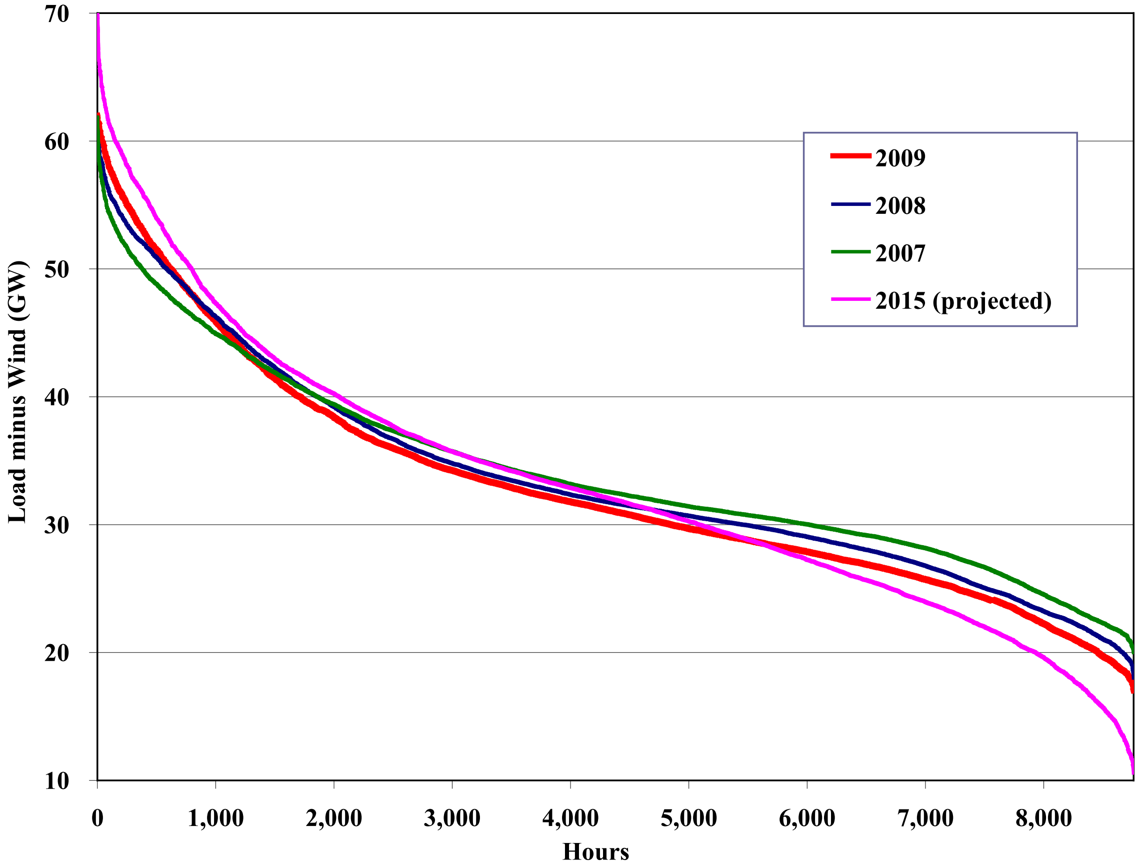
Figure 29: ERCOT Load Duration Curve – Top 5% of Hours



This figure also shows that the peak load in each year is significantly greater than the load at the 95th percentile of hourly load. From 2006 to 2009, the peak load value averaged 19.7 percent greater than the load at the 95th percentile. These load characteristics imply that a substantial amount of capacity – over 10 GW – is needed to supply energy in less than 5 percent of the hours. Additionally, another 8 GW of capacity is required to meet the ERCOT planning reserve requirement of expected peak demand plus 12.5 percent. These factors serve to emphasize the importance of efficient energy pricing during peak demand conditions and other times of system stress that send accurate economic signals for the investment in and retention of the resources required to meet these real-time system demands as well as achieving long-term resource adequacy requirements.

Increasing levels of wind resource in ERCOT also has important implications for the net load duration curve faced by the non-wind fleet of resources. Net load is defined as the system load minus wind production. Figure 30 shows the net load duration curves for 2007 through 2009, with projected values for 2015 based on ERCOT data from its Competitive Renewable Energy Zones assessment.

Figure 30: Net Load Duration Curves



The data in Figure 30 show that while the peak net load has increased from 2007 to 2009, the remainder of the net load duration curve has been reduced. This is due in part in to the 1.3 percent decline in energy consumption in 2009, but is largely associated with the increase in wind production in the ERCOT region over this time period. Over 90 percent of the wind resources in the ERCOT region are located in West Texas, and the wind profiles in this area are such that most of the wind production occurs during off-peak hours or other times of relatively low system demand. This profile results in only modest reductions of the net load relative to the actual load during the hours of highest demand, but much more significant reductions in the net load relative to the actual load in the other hours of the year. The projection for 2015 indicates that the trend shown from 2007 to 2009 is expected to continue and amplify with the addition of significant new wind resources and the reduction in the curtailment of existing wind resources. Focusing on the left side of the net load duration curve, the average difference between peak net load and the 95th percentile of net load was 10.7 GW in 2007 to 2009, but this differential is projected to increase to over 15 GW by 2015. With an additional capacity requirement of approximately 9 GW to meet the 12.5 percent reserve margin requirement, this means that over 24 GW of non-wind capacity will be required to exist on the system with an expectation of operating five percent of the hours in a year or less. On the right side of the net load duration curve, the minimum net load was 17 GW in 2007 to 2009, but the minimum is projected to decrease to less than 11 GW by 2015.

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

B. Load Scheduling

In this subsection, we evaluate load scheduling patterns by comparing load schedules to actual real-time load. Under the ERCOT Protocols, scheduled load must be balanced with scheduled resources for each QSE for each settlement interval; however, there is no requirement that the scheduled load be consistent with the actual load of a QSE. Additionally, a QSE may balance its scheduled load with resources scheduled from ERCOT. Because the financial effect of scheduling resources from ERCOT to balance a load schedule is the same as if the load were unscheduled, in this section, we adjust the load schedules by subtracting the amount that consists of resources scheduled from ERCOT.

To provide an overview of the scheduling patterns, Figure 31 shows a scatter diagram that plots the ratio of the final load schedules to the actual load level during 2009. The ratio shown in the figure will be greater than 100 percent when the final load schedule is greater than the actual load.

**Figure 31: Ratio of Final Load Schedules to Actual Load
All ERCOT**

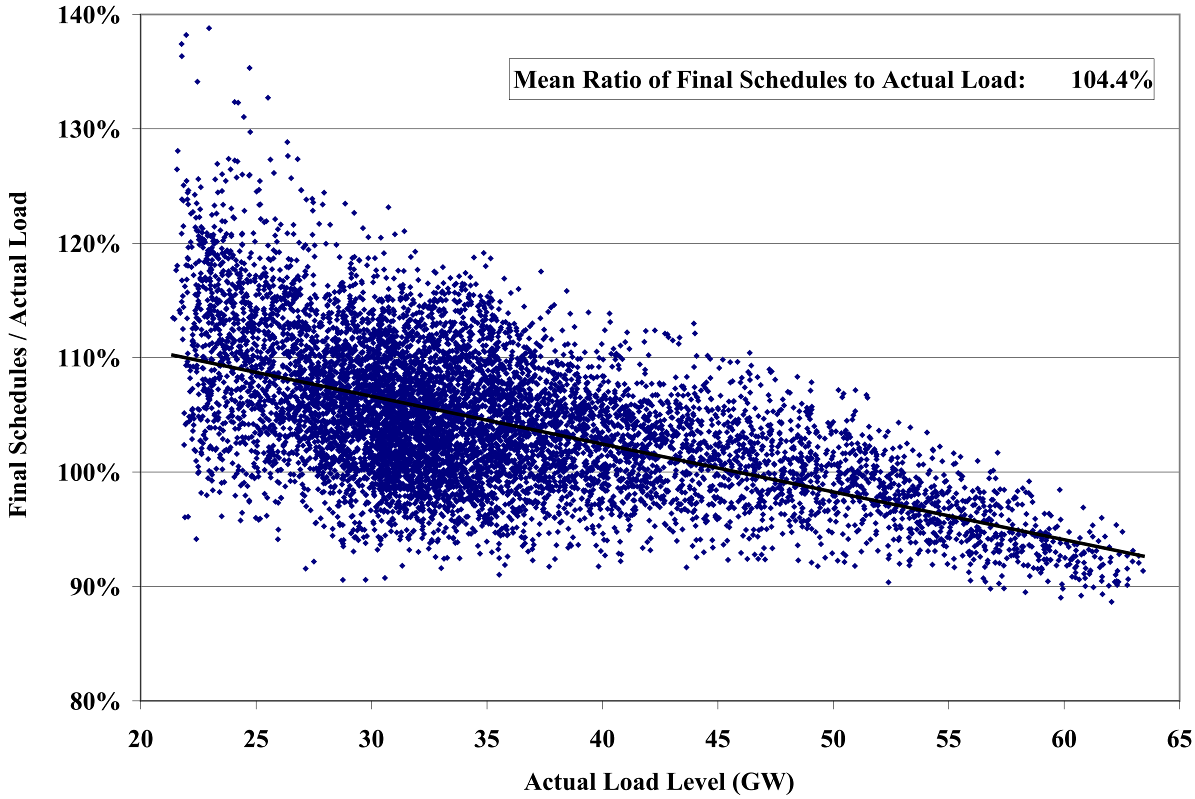


Figure 31 shows that final load schedules on average was higher than the actual load in aggregate, as indicated by an average ratio of the final load schedules to actual load of 104.4 percent. However, the figure also includes a trend line indicating that the ratio of final load schedules to actual load tends to decrease as load rises. In particular, the ratio given by the trend line is above 100 percent for loads under 45 GW and declines to 92 percent at higher load levels. The overall pattern shown in the figure above is similar to previous years, which exhibited the same downward trend in final load schedules relative to actual load.

On average, balancing energy prices are higher and more volatile at high load levels, although the previous subsection showed that spikes can occur under all load conditions. Market participants that are risk averse might be expected to schedule forward to cover a significant portion of their load during high load periods rather than reducing their forward scheduling levels during those periods. There are several explanations for the apparent under-scheduling during high load conditions. First, while the data suggest that QSEs rely more on the balancing energy market at higher load levels, doing so does not necessarily subject them to greater price risk. Financial contracts or derivatives may be in place to protect market participants from price risk in the balancing energy market, such as a contract for differences. Second, market participants who own generation can offer their expensive generation into the market to cover their load needs if balancing energy market prices are high but otherwise allow their load obligations to be met with lower-priced balancing energy. Third, some market participants may not have contracted for sufficient resources to cover their peak load and may, therefore, not be able to fully schedule their load.

Figure 32: Average Ratio of Final Load Schedules to Actual Load by Load Level

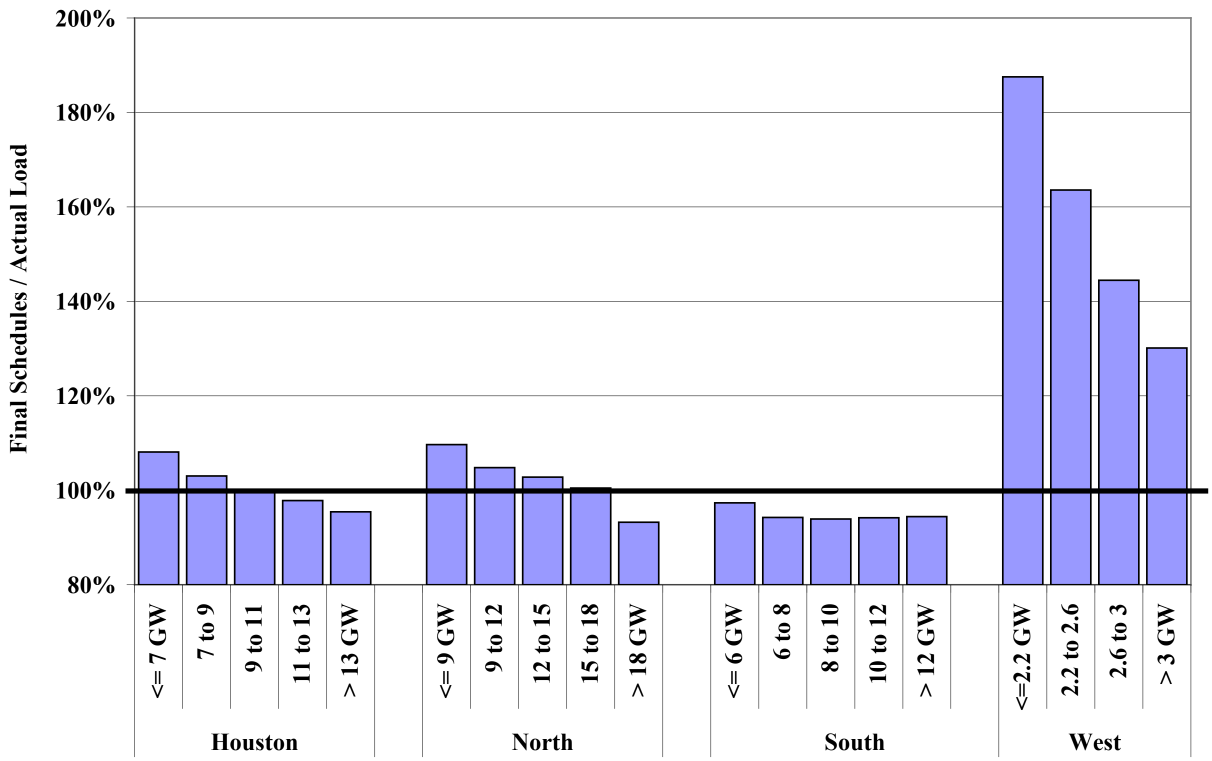


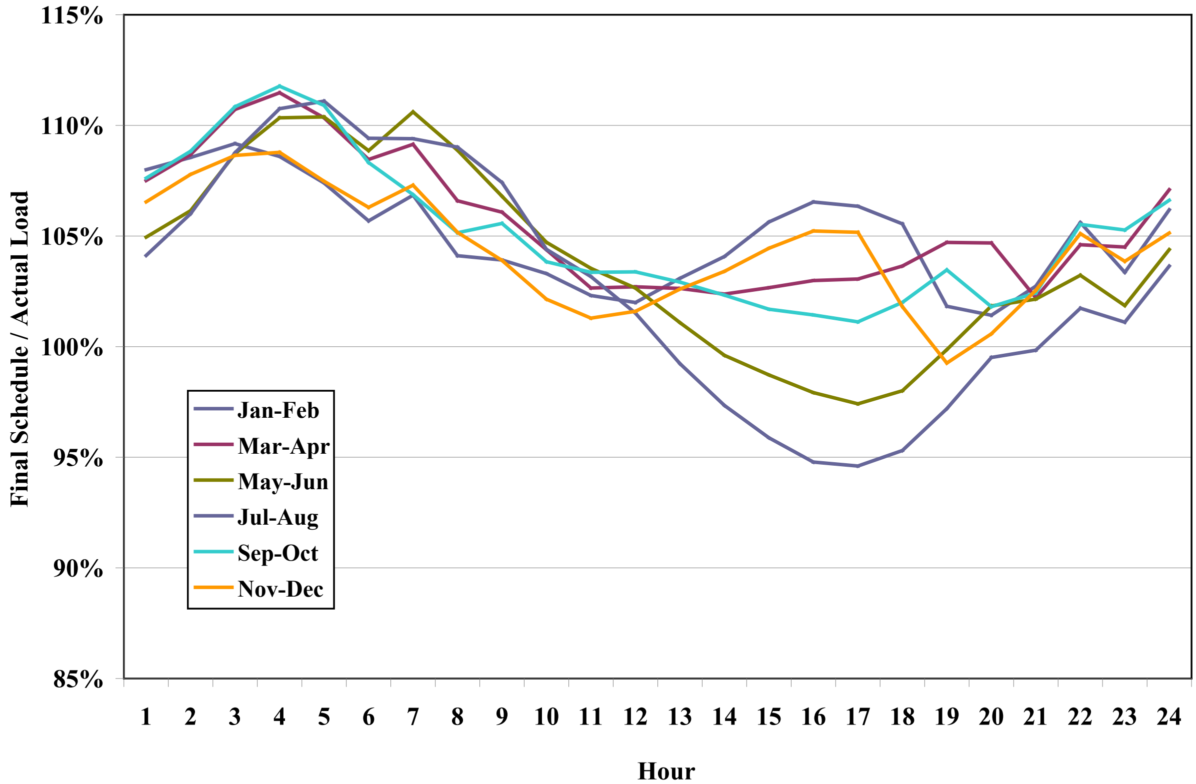
Figure 32 shows the ratio of final load schedules to actual load evaluated at five different load levels for each of the ERCOT zones. Figure 32 shows that:

- The West Zone is significantly over-scheduled, although the ratio declines as load increases.
- The Houston and North Zones are under-scheduled at the highest load levels.
- The South Zone is under-scheduled at all load levels.

It should be noted that regardless of the relationship between the aggregate scheduled load and actual load, individual QSEs may be significant net sellers or purchasers in the balancing energy market. Persistent load imbalances are not necessarily a problem. Imbalances can reflect the fact that some suppliers schedule energy from resources they expect to be economic in the balancing energy market when they have not already sold the power in a bilateral contract. Rather than selling power to the balancing energy market through deployments in the balancing energy market, they sell through load imbalances. Additionally, some load-serving entities may choose to purchase a portion of their load obligations in the balancing energy market. These approaches reflect economic decisions of wholesale buyers and sellers and generally do not present operational concerns.

To further analyze load scheduling, Figure 33 shows the ratio of final load schedules to actual load by hour in two month blocks.

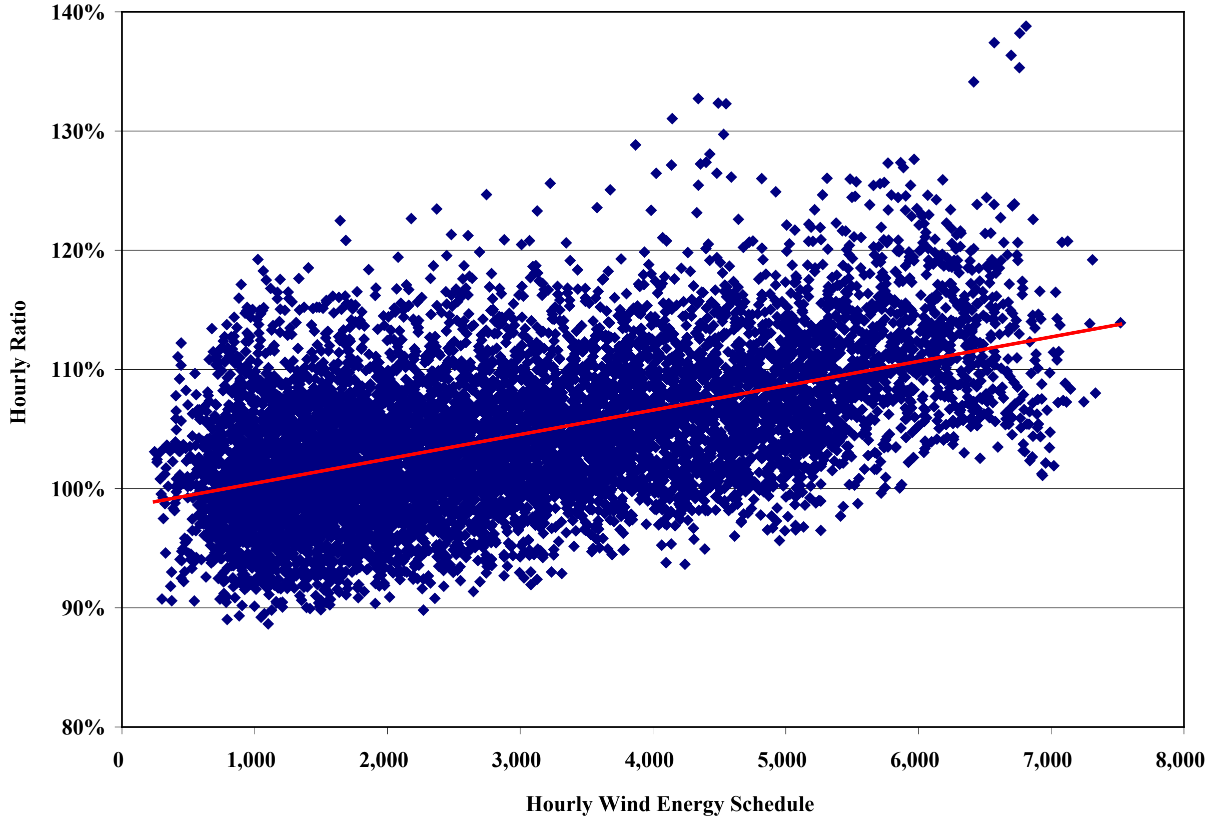
Figure 33: Average Ratio of Final Load Schedules to Actual Load by Hour



This figure shows that the final schedules exceed actual load in all months for the hours 1-12 and 21-24. Final schedules are significantly less than actual load only in the summer months of May through August during the peak demand hours in the afternoon.

A significant factor influencing the relationship between final load schedules and actual load in 2009 was the increased wind generation capacity. Figure 34 shows the load schedule as a percentage of actual load versus wind energy schedules in 2009.

Figure 34: Load Schedule/Actual Load vs. Wind Energy Schedule

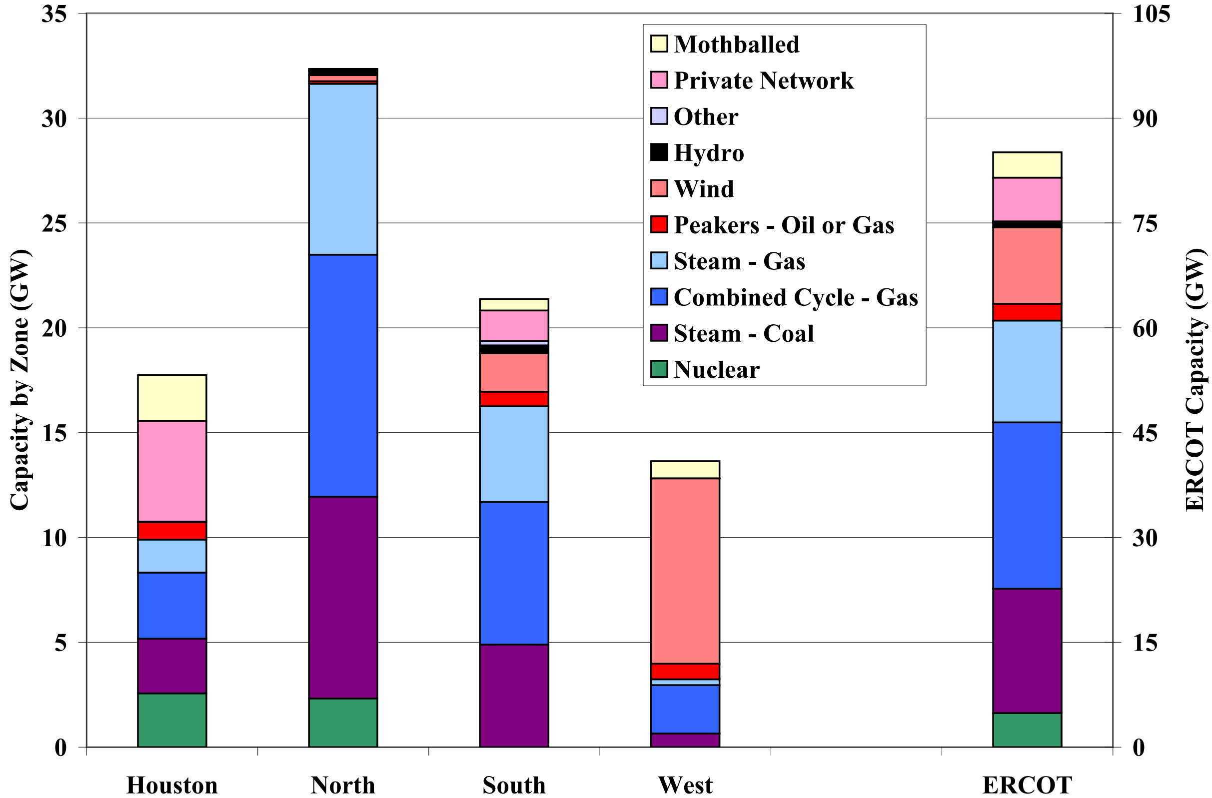


This figure shows a positive correlation between the load schedule as a percentage of actual load and the wind energy schedules. Typically, the production profile for wind resources in the ERCOT market is such that most output occurs during off-peak hours and in the non-summer months. Thus, the data in Figure 34 provide further explanation of the results in Figure 33 that shows that final load schedules exceed actual load most significantly during the off-peak hours and in the non-summer months.

C. Generation Capacity in ERCOT

In this section we evaluate the generation mix in ERCOT. With the exception of the wind resources in the West Zone and the nuclear resources in the North and Houston Zones, the mix of generating capacity is relatively uniform in ERCOT. Figure 35 shows the installed generating capacity by type in each of the ERCOT zones.

Figure 35: Installed Capacity by Technology for each Zone



The nuclear capacity is located in both the North and Houston Zones. Lignite and coal generation is also a significant contributor in ERCOT. However, the primary fuel in ERCOT is natural gas, accounting for nearly 58 percent of generation capacity in ERCOT as a whole and almost 60 percent in the South Zone. Approximately 60 percent of this natural gas-fired capacity represents relatively new combined-cycle units that have been installed throughout ERCOT over the past decade. These new installations have resulted in a small increase in the gas-fired share of installed capacity but have not changed the overall mix significantly, since the generators that have gone out of service during this period were primarily gas-fired steam turbines.

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 38 percent of capacity, the South Zone 25 percent, the Houston Zone 21 percent, and the West Zone 16 percent. The Houston Zone typically imports power, while the West and North Zones typically export power. Excluding mothballed resources and including only 8.7 percent of wind capacity as capacity available to reliably meet peak demand, the North

Zone accounts for approximately 45 percent of capacity, the South Zone 27 percent, the Houston Zone 22 percent, and the West Zone 7 percent.

While ERCOT has coal/lignite and nuclear plants that operate primarily as base load units, its reliance on natural gas resources makes it vulnerable to natural gas price spikes. There is approximately 22.6 GW of coal and nuclear generation in ERCOT. Because there are very few hours when ERCOT load drops as low as 20 GW, natural gas resources will be dispatched and set the balancing energy spot price in most hours. Hence, although coal-fired and nuclear units combined produce approximately half of the energy in ERCOT, they have historically played a much less significant role in setting spot electricity prices. However, with the significant increases in wind capacity that has a lower marginal production cost than coal and lignite, the frequency at which coal and lignite are the marginal units in ERCOT is expected to increase in the future, particularly during the off-peak hours in the spring and fall, and even more as additional transmission capacity is added that will accommodate increased levels of wind production in the West Zone.

Figure 36 and Figure 37 show the marginal fuel frequency for the Houston and West Zones, respectively, for each month from 2007 through 2009.[18] The marginal fuel frequency is the percentage of hours that a generation fuel type is marginal and setting the price at a particular location.

As shown in Figure 36, the frequency at which coal was the price setting fuel for the Houston Zone experienced a significant and sustained increase beginning in September 2008. This increase can be attributed to (1) increased wind resource production; (2) a slight reduction in demand in 2009 due to the economic downturn; and (3) periods when natural gas prices were very low thereby makinn coal and combined-cycle natural gas resources competitive from an economic dispatch standpoint. As significant additional wind, coal and potentially nuclear resources are added to the ERCOT region and transmission constraints that serve to limit existing wind production are alleviated, it is likely that the frequency of coal as the marginal fuel will increase in coming years.

18 The marginal fuel frequency for the North and South Zones are very similar to the Houston Zone.

Figure 36: Marginal Fuel Frequency (Houston Zone)

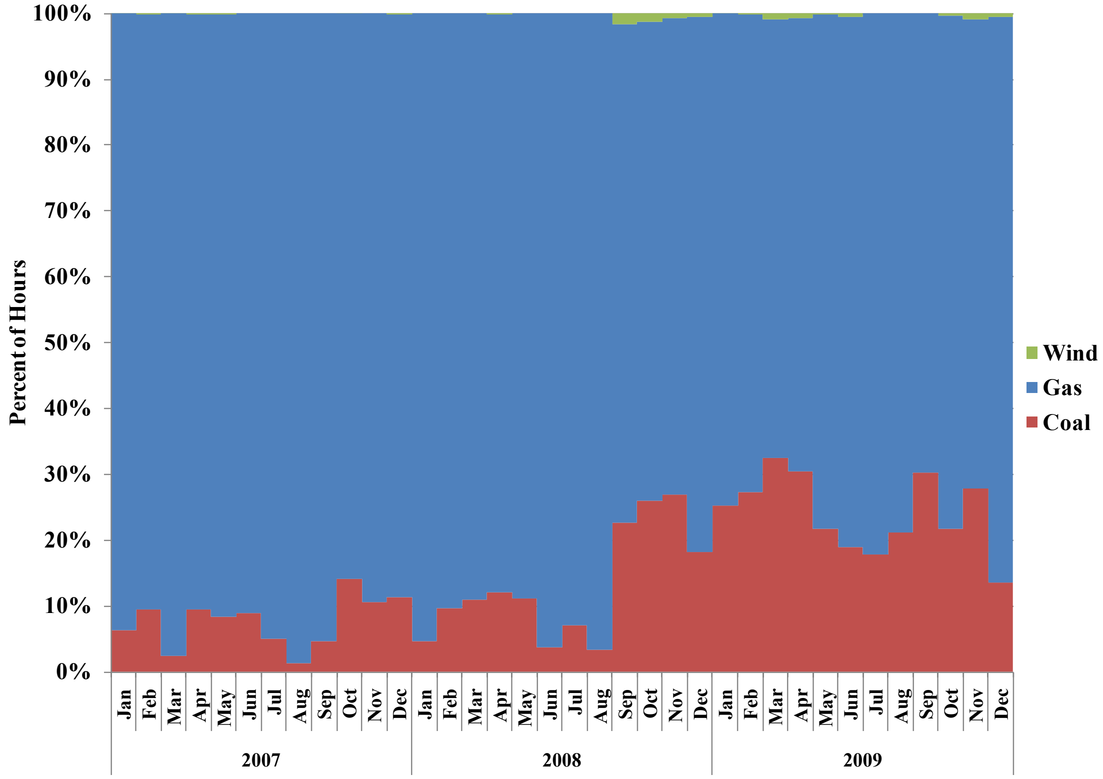


Figure 37: Marginal Fuel Frequency (West Zone)

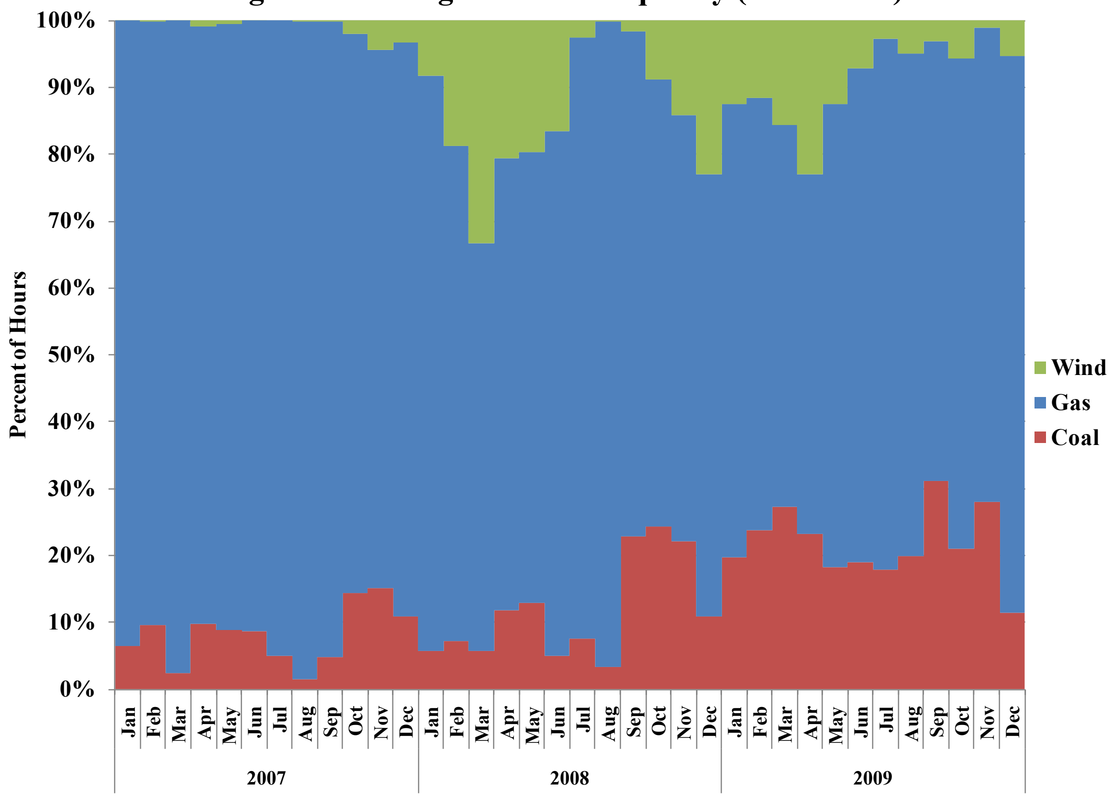


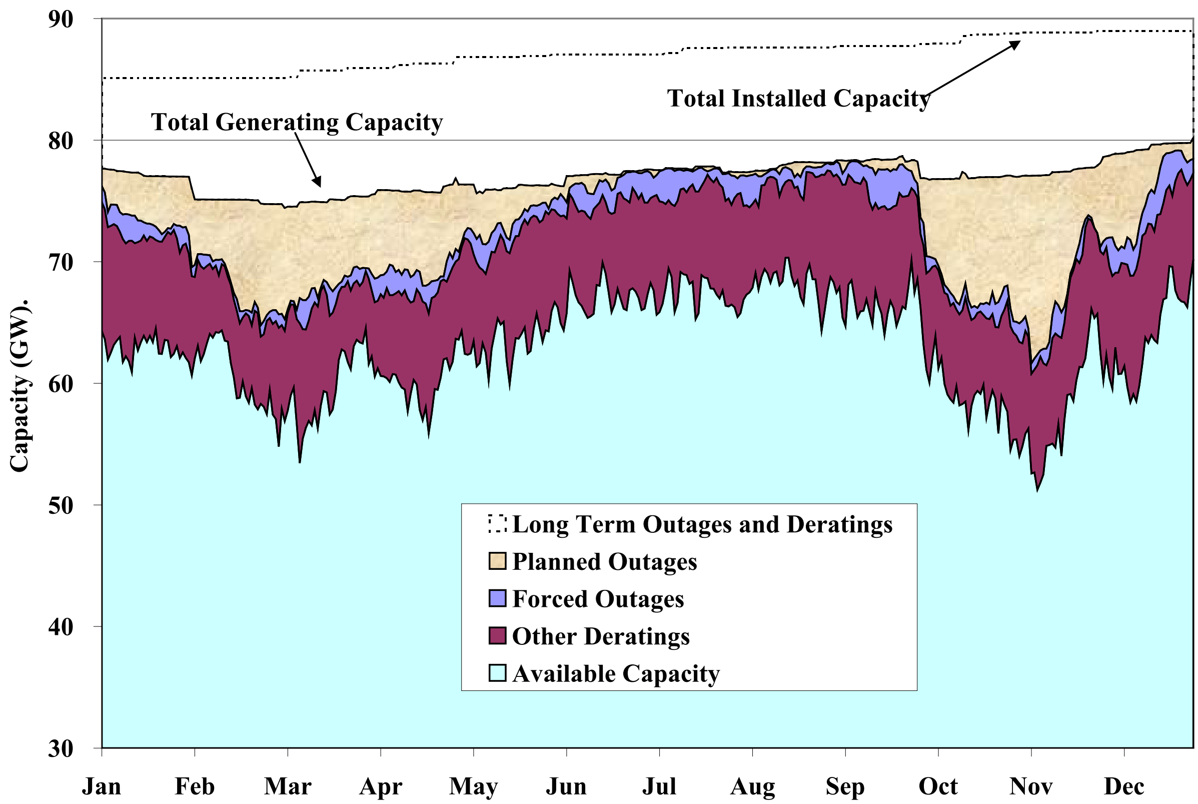
Figure 37 shows that the frequency at which coal was the price setting fuel for the West Zone also experienced a significant and sustained increase beginning in September 2008. This figure also shows that beginning in late 2007 the frequency at which wind was the price setting fuel for the West Zone increased dramatically. This increase is attributable to the growth in installed wind capacity that far exceed the load in the West Zone combined with existing transmission capability that limits the export capability from the West Zone, as discussed in more detail in Section III.

1. Generation Outages and Deratings

Figure 35 in the prior subsection shows that installed capacity is approximately 85 GW including mothballed units and all wind capacity, and approximately 71 GW excluding mothballed capacity and including only 8.7 percent of wind capacity. Hence, the installed capacity exceeds the capacity required to meet annual peak load plus ancillary services requirements of 67 GW. This might suggest that the adequacy of resources is not a concern for ERCOT in the near-term. However, resource adequacy must be evaluated in light of the resources that are actually available on a daily basis to satisfy the energy and operating reserve requirements in ERCOT. A substantial portion of the installed capability is frequently unavailable due to generator deratings. A derating is the difference between the maximum installed capability of a generating resource and its actual capability (or "rating") in a given hour. Generators may be fully derated (rating equals 0) due to a forced or planned outage. It is also very common for generating capacity to be partially derated (*e.g.*, by 5 to 10 percent) because the resource cannot achieve its installed capability level due to technical or environmental factors (*e.g.*, component equipment failures or ambient temperature conditions).

In this subsection, we evaluate long-term and short-term deratings to inform our evaluation of ERCOT capacity levels. Figure 38 shows a breakdown of total installed capability for ERCOT on a daily basis during 2009. This analysis includes all in-service and switchable capacity. The capacity in this analysis is separated into five categories: (a) long-term outages and deratings, (b) short-term planned outages, (c) short-term forced outages, (d) other short-term deratings, and (e) available and in-service capability.

Figure 38: Short and Long-Term Deratings of Installed Capability*



* Includes all outages and deratings lasting greater than 60 days and all mothballed units.

* Switchable capacity is included under installed capacity in this figure.

Figure 38 shows that long-term outages and other deratings fluctuated between 14 and 22 GW. These outages and deratings reduce the effective resource margins in ERCOT from the levels reported above. A large component of the "other deratings" is associated with limited wind resources resulting in generating resources that are not capable of producing up to the full installed capability. Other causes of these deratings reflect:

- Cogeneration resources unavailable to serve market load because they are being used to serve self-serve load;
- Resources out-of-service for economic reasons (*e.g.*, mothballed units); or
- Resources out-of-service for extended periods due to maintenance requirements.

With regard to short-term deratings and outages, the patterns of planned outages and forced outages were consistent with expectations:

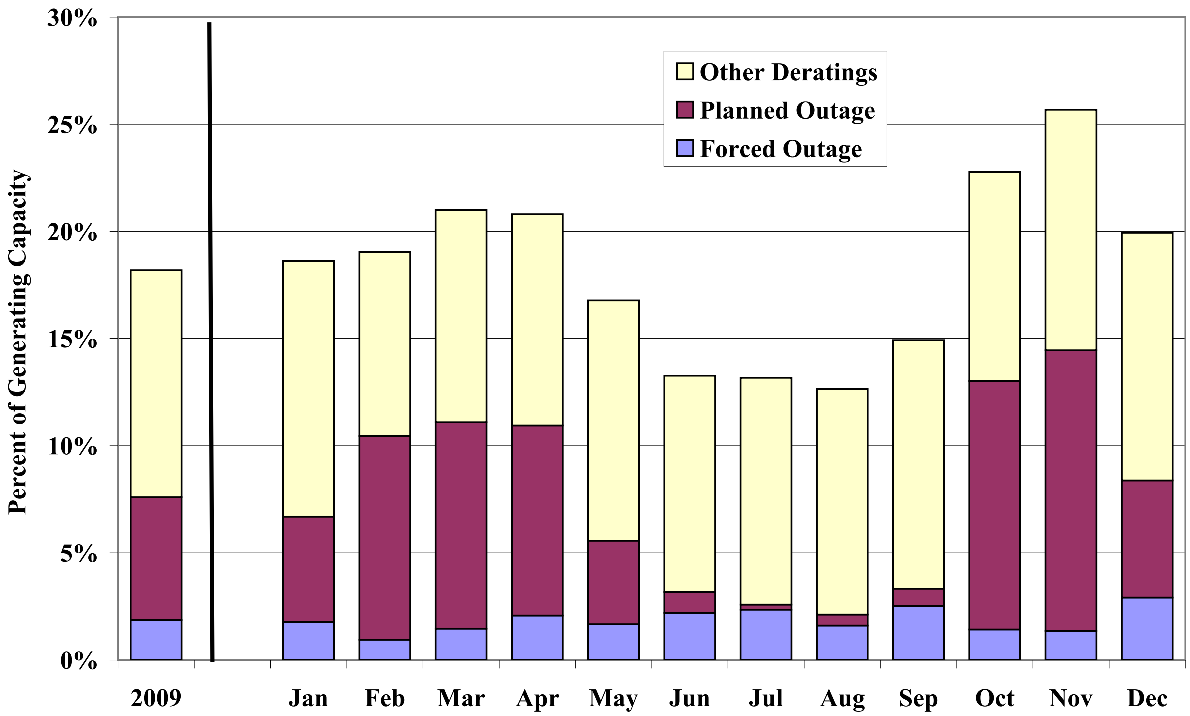
- Forced outages occurred randomly over the year and the forced outage rates were relatively low (although all forced outages may not be reported to ERCOT).

- Planned outages were relatively large in the spring and fall and extremely small during the summer.

Although the total installed capacity was higher in 2009 than in 2008, the annual average of daily available capacity was unchanged. Further, the average of daily available capacity during the summer months (May through September) decreased 1,180 MW from 2008, which can be primarily attributed to higher quantities of derating due to wind resource availability.

The next analysis focuses specifically on the short-term forced outages and other short-term deratings. Figure 39 shows the average magnitude of the outages and deratings lasting less than 60 days for the year and for each month during 2009.

Figure 39: Short-Term Outages and Deratings*



* Excludes all outages and deratings lasting greater than 60 days and all mothballed units.

Figure 39 shows that total short-term deratings and outages were as large as 25 percent of installed capacity in the spring and fall, and dropping to as low as 12 percent for the summer. Most of this fluctuation was due to anticipated planned outages, which ranged as high as 8 to 13 percent of installed capacity during February through April, and October through November. Short-term forced outages occurred more randomly, as would be expected, ranging between one and three percent of total capacity on a monthly average basis during 2009. These rates are relatively low in comparison to other operating markets for two reasons. First, these outages include only full outages (*i.e.*, where the resource's rating equals zero). In contrast, an equivalent forced outage rate is frequently reported for other markets, which includes both full and partial outages. Hence, the forced outage rate shown in Figure 39 can be expected to be lower than equivalent forced outage rates of other markets. Second, because forced outage information is self-reported by generators, we are not confident that the available data includes all forced outages that actually occurred.

The largest category of short-term deratings was the "other deratings" that occur for a variety of reasons. The other deratings would include any short-term forced or planned outage that was not reported or correctly logged by ERCOT. This category also includes deratings due to ambient temperature conditions, cogeneration uses, wind deratings due to variable wind conditions and other factors described above. Furthermore, suppliers may delay maintenance on components such as boiler tubes, resulting in reduced capability. Because these deratings can fluctuate day to day or seasonally, some of the deratings are included in the "long-term outages and deratings" category while the others are included in this category. The other deratings were approximately 10 percent on average during the summer in 2009 and as high as 11 percent in other months. In conclusion, the patterns of outages do not indicate patterns of physical withholding or raise other competitive concerns. However, this issue is analyzed in more detail in Section IV of this report.

2. Daily Generator Commitments

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

This subsection evaluates the commitment patterns in ERCOT by examining the levels of excess capacity. Excess capacity is defined as the total online capacity plus quick-start[19] units minus the demand for energy, responsive reserve, up regulation and non-spinning reserve provided from online capacity or quick-start units. To evaluate the commitment of resources in ERCOT,

19 For the purposes of this analysis, "quick-start" includes simple cycle gas turbines that are qualified to provide balancing energy.

Figure 40 plots the excess capacity in ERCOT during 2009. The figure shows the excess capacity in only the peak hour of each weekday because the largest generation commitment usually occurs at the peak hour. Hence, one would expect larger quantities of excess capacity in other hours.

Figure 40: Excess On-Line and Quick Start Capacity During Weekday Daily Peaks

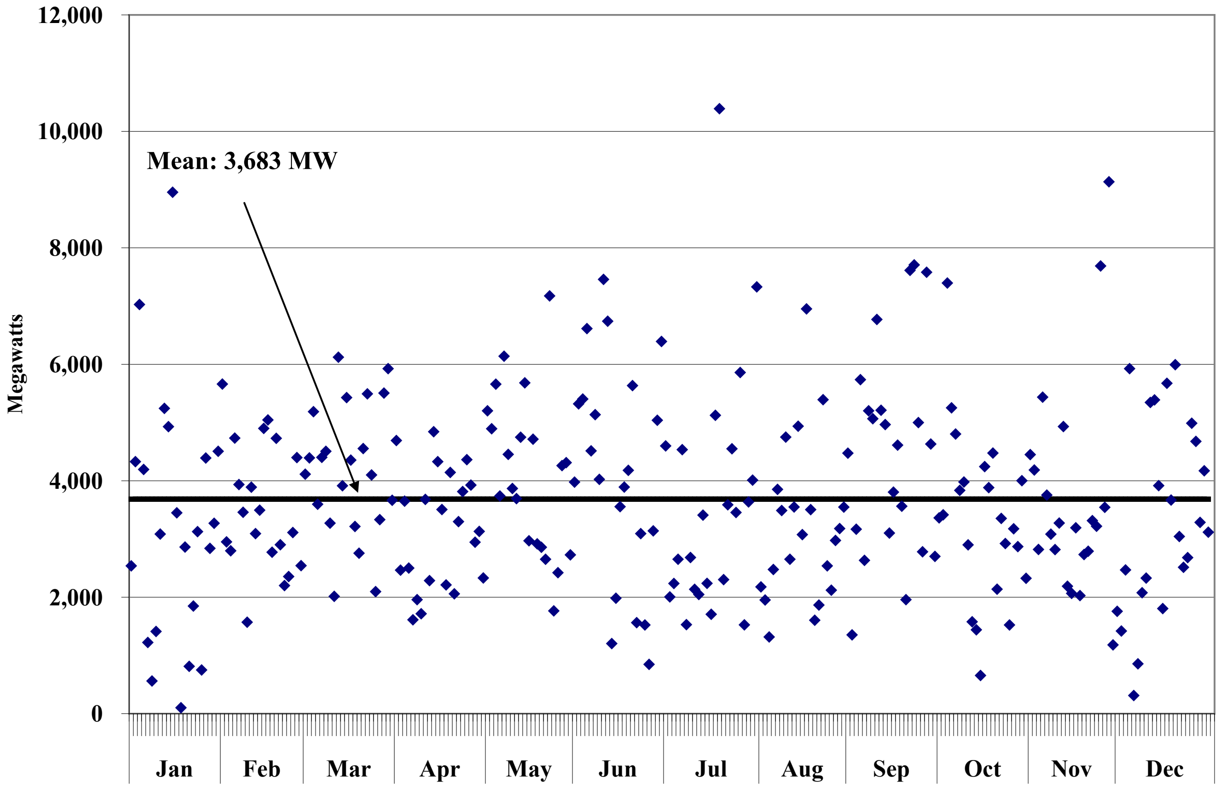


Figure 40 shows that the excess on-line capacity during daily peak hours on weekdays averaged 3,683 MW in 2009, which is approximately 11.9 percent of the average load in ERCOT. This is an increase of more than 600 MW from prior years. One explanation for the increase in excess on-line capacity in 2009 is the increase in the number of quick-start resources that are qualified to provide balancing energy service. Quick-start resources are actually off-line until dispatched; however, these resources are included in the on-line capacity calculation. The use of quick-start resources for balancing energy service results in a more efficient commitment of resources to managed uncertainties that materialize near real-time than does a process of making firm commitment decisions in the day ahead. For this reason, increases in the excess on-line capacity that are associated with the existence of additional quick -tart resources are not an efficiency concern.

The overall trend in excess on-line capacity in recent years indicates a movement toward more efficient unit commitment across the ERCOT market; however, the current market structure is still based primarily upon a decentralized unit commitment process whereby each participant makes independent generator commitment decisions that are not likely to be optimal. Further contributing to the suboptimal results of the current unit commitment process is that the decentralized unit commitment is comprised of non-binding resource plans that form the basis for ERCOT's day-ahead planning decisions. However, these non-binding plans can be modified by market participants after ERCOT's day-ahead planning process has concluded causing ERCOT to take additional actions that may be more costly and less efficient. Hence, the introduction of a day-ahead energy market with centralized Security Constrained Unit Commitment ("SCUC") that is financially binding under the nodal market design promises substantial efficiency improvements in the commitment of generating resources.

D. Demand Response Capability

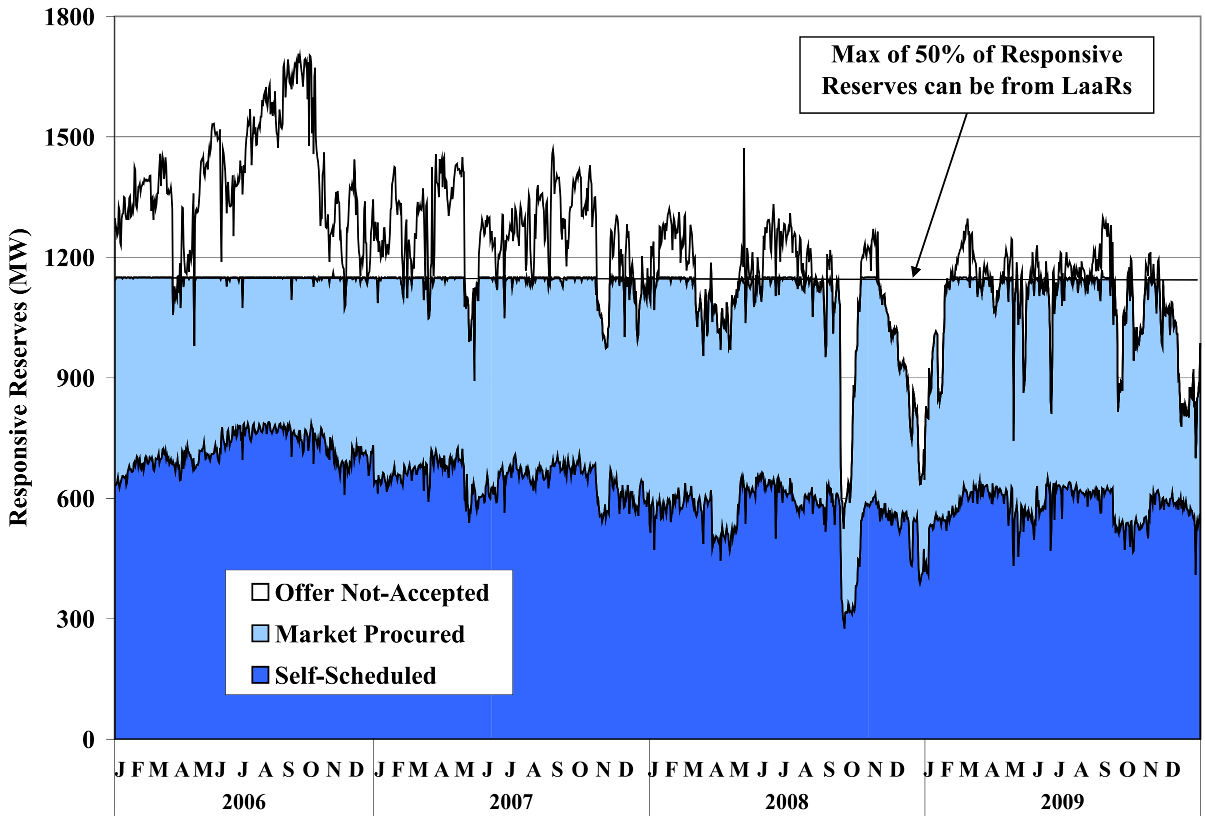
Demand response is a term that broadly refers to actions that can be taken by end users of electricity to reduce load in response to instructions from ERCOT or in response to certain market or system conditions. The ERCOT market allows participants with demand-response capability to provide energy and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

ERCOT allows qualified LaaRs to offer responsive reserves and non-spinning reserves into the day-ahead ancillary services markets. Qualified LaaRs can also offer blocks of energy in the balancing energy market. LaaRs providing up balancing energy must have telemetry and must be capable of responding to ERCOT energy dispatch instructions in a manner comparable to generation resources. Those providing responsive reserves must have high set under-frequency relay ("UFR") equipment. A load with UFR equipment is automatically tripped when the frequency falls below 59.7 Hz, which will typically occur only a few times in each year.

BULs are loads that are qualified to offer demand response capability in the balancing energy market. These loads must have an Interval Data Recorder to qualify and do not require telemetry. BULs may provide energy in the balancing energy market, but they are not qualified to provide reserves or regulation service.

As of December 2009, over 2,200 MW of capability were qualified as LaaRs. These resources regularly provided reserves in the responsive reserves market, but never participated in the balancing energy market and only a very small portion participated in the non-spinning reserves market. Figure 41 shows the amount of responsive reserves provided from LaaRs on a daily basis in 2009.

Figure 41: Provision of Responsive Reserves by LaaRs
Daily Average



The high level of participation by demand response in the ancillary service markets sets ERCOT apart from other operating electricity markets. Figure 41 shows that the amount of responsive reserves provided by LaaRs has held fairly constant at 1,150 MW since the beginning of 2006. (For reliability reasons, 1,150 MW is the limit of participation in the responsive reserve market by LaaRs.) Exceptions include a decrease in September of 2008 corresponding to the Texas landfall of Hurricane Ike and a more prolonged reduction from November 2008 through January 2009 that was likely a product of the economic downturn and its effect on industrial operations.

Although LaaRs are active participants in the responsive reserves market, they did not offer into the balancing energy, regulation or non-spinning reserve services markets in 2009. This is not surprising because the value of curtailed load tends to be very high, and providing responsive reserves offers substantial revenue with very little probability of being deployed. In contrast, providing non-spinning reserves introduces a much higher probability of being curtailed. Participation in the regulation services market requires technical abilities that most LaaRs cannot meet at this point.

E. Net Revenue Analysis

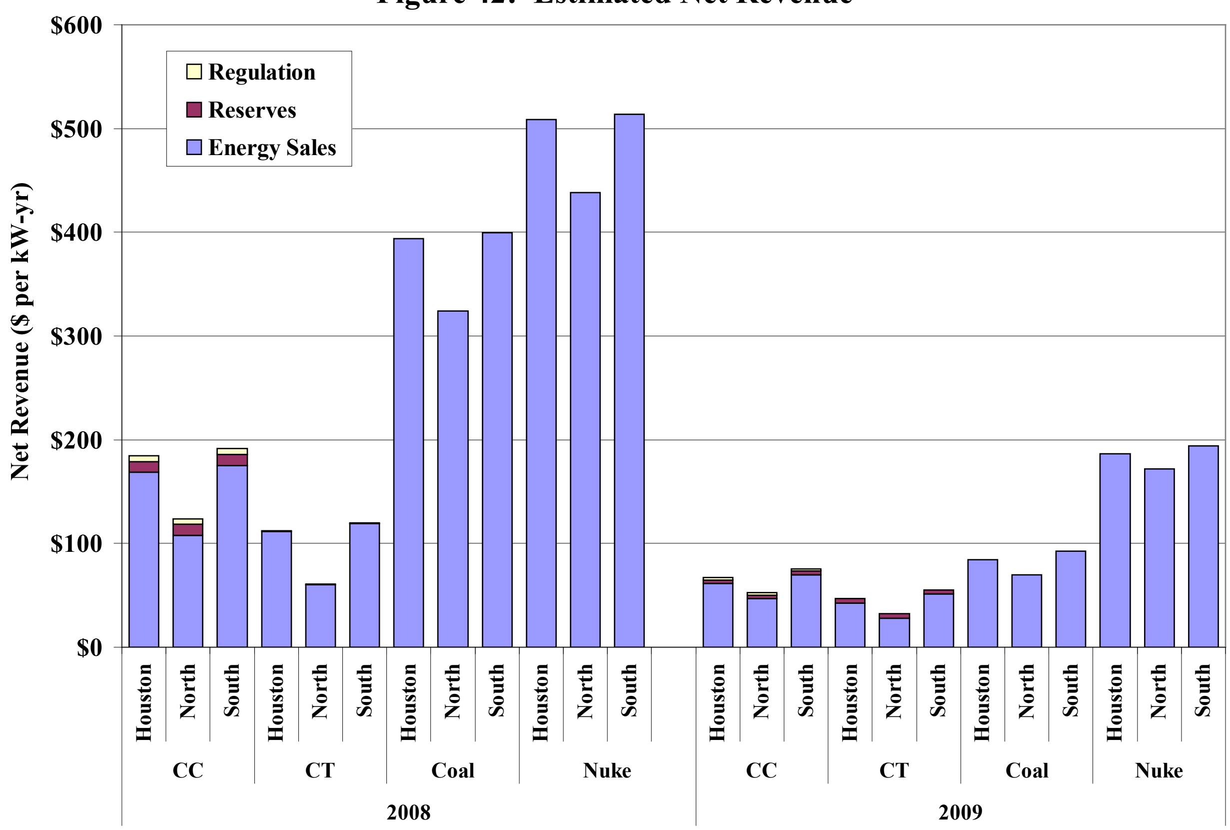
Net revenue is defined as the total revenue that can be earned by a generating unit less its variable production costs. Hence, it is the revenue in excess of short-run operating costs and is available to recover a unit's fixed and capital costs. Net revenues from the energy, operating reserves, and regulation markets together provide the economic signals that inform suppliers' decisions to invest in new generation or retire existing generation. In a long-run equilibrium, the markets should provide sufficient net revenue to allow an investor to break-even on an investment in a new generating unit. In the short-run, if the net short-run revenues produced by the market are not sufficient to justify entry, then one or more of three 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; or
- Market rules are causing revenues to be reduced inefficiently.

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

Figure 42 shows the results of the net revenue analysis for four types of units in 2008 and 2009. These are: (a) a gas combined-cycle, (b) a combustion turbine, (c) a coal unit, and (d) a nuclear unit. In recent years, most new capacity investment has been in natural gas-fired technologies, although high prices for oil and natural gas have caused renewed interest in new investment in coal and nuclear generation. For the gas-fired technologies, net revenue is calculated by assuming the unit will produce energy in any hour for which it is profitable and by assuming it will be available to sell reserves and regulation in other hours that it is available (*i.e.*, when it is not experiencing a planned or forced outage). For coal and nuclear technologies, net revenue is calculated by assuming that the unit will produce at full output. The energy net revenues are computed based on the balancing energy price in each hour. Although most suppliers would receive the bulk of their revenues through bilateral contracts, the spot prices produced in the balancing energy market should drive the bilateral energy prices over time.

Figure 42: Estimated Net Revenue



For purposes of this analysis, we assume heat rates of 7 MMbtu per MWh for a combined cycle unit, 10.5 MMbtu per MWh for a combustion turbine, and 9.5 MMbtu per MWh for a new coal unit. We assume variable operating and maintenance costs of $4 per MWh for the gas units and

$5 per MWh for the coal unit. We assume fuel and variable operating and maintenance costs of $8 per MWh for the nuclear unit. For each technology, we assumed a total outage rate (planned and forced) of 10 percent.

Some units, generally those in unique locations that are used to resolve local transmission constraints, also receive a substantial amount of revenue through uplift payments (*i.e.*, Out-of-Merit Energy, Out-of-Merit Capacity, and Reliability Must Run payments). This source of revenue is not considered in this analysis. The analysis also includes simplifying assumptions that can lead to over-estimates of the profitability of operating in the wholesale market. The following factors are not explicitly accounted for in the net revenue analysis: (i) start-up costs, which can be significant; and (ii) minimum running times and ramp restriction, which can prevent the natural gas generators from profiting during brief price spikes. Despite these limitations, the net revenue analysis provides a useful summary of signals for investment in the wholesale market.

Figure 42 shows that the net revenue decreased substantially in 2009 compared to each zone compared in 2008 and 2007. Based on our estimates of investment costs for new units, the net revenue required to satisfy the annual fixed costs (including capital carrying costs) of a new gas turbine unit ranges from $70 to $95 per kW-year. The estimated net revenue in 2009 for a new gas turbine was approximately $55, $47 and $32 per kW-year in the South, Houston and North Zones, respectively. For a new combined cycle unit, the estimated net revenue requirement is approximately $105 to $135 per kW-year. The estimated net revenue in 2009 for a new combined cycle unit was approximately $76, $67 and $52 per kW-year in the South, Houston and North Zones, respectively. These values indicate that the estimated net revenue in 2009 was well below the levels required to support new entry for a new gas turbine or a combined cycle unit in the ERCOT region.

Prior to 2005, net revenues were well below the levels necessary to justify new investment in coal and nuclear generation. However, high natural gas prices through 2008 allowed energy prices to remain at levels high enough to support new entry for these technologies. The production costs of coal and nuclear units did not change significantly over this period, leading to a dramatic rise in net revenues. With the significant decline in natural gas and energy prices in 2009, these results changed dramatically from recent years. For a new coal unit, the estimated net revenue requirement is approximately $190 to $245 per kW-year. The estimated net revenue in 2009 for a new coal unit was approximately $93, $84 and $70 per kW-year in the South, Houston and North Zones, respectively. For a new nuclear unit, the estimated net revenue requirement is approximately $280 to $390 per kW-year. The estimated net revenue in 2009 for a new nuclear unit was approximately $194, $187 and $172 per kW-year in the South, Houston and North Zones, respectively. These values indicate that the estimated net revenue for a new coal and nuclear unit in the South, Houston and North Zones was well below the levels required to support new entry in 2009.

Although estimated net revenue declined considerably in 2009 compared to the prior four years, there are other factors that determine incentives for new investment. First, market participants must anticipate how prices will be affected by the new capacity investment, future load growth, and increasing participation in demand response. Second, net revenues can be inflated when prices clear above competitive levels as a result of market power being exercised. Thus, a market participant may be deterred from investing in new capacity if it believes that prevailing net revenues are largely due to an exercise of market power that would not be sustainable after the entry of the new generation. Third, the nodal market design will have an effect on the profitability of new resources. In a particular location, nodal prices could be higher or lower than the prices in the current market depending on the pattern of congestion.

To provide additional context for the net revenue results presented in this section, we also compared the net revenue for natural gas-fired technologies in the ERCOT market with net revenue in other centralized wholesale markets. Figure 43 compares estimates of net revenue for each of the auction-based wholesale electricity markets in the U.S.: the ERCOT North Zone, the California ISO, the New York ISO, and PJM. The figure includes estimates of net revenue from energy, reserves and regulation, and capacity. ERCOT does not have a capacity market, and thus, does not have any net revenue from capacity sales.[20]

[20] The California ISO does not report capacity and ancillary services net revenue separately, so it is shown as a combined block in Figure 43. Generally, estimates were performed for a theoretical new combined-cycle unit with a 7,000 BTU/kWh heat rate and a theoretical new gas turbine with a 10,500 BTU/kWh heat rate. However, the California ISO reports net revenues for 7,650 and 9,500 BTU/kWh units.

Figure 43: Comparison of Net Revenue of Gas-Fired Generation between Markets

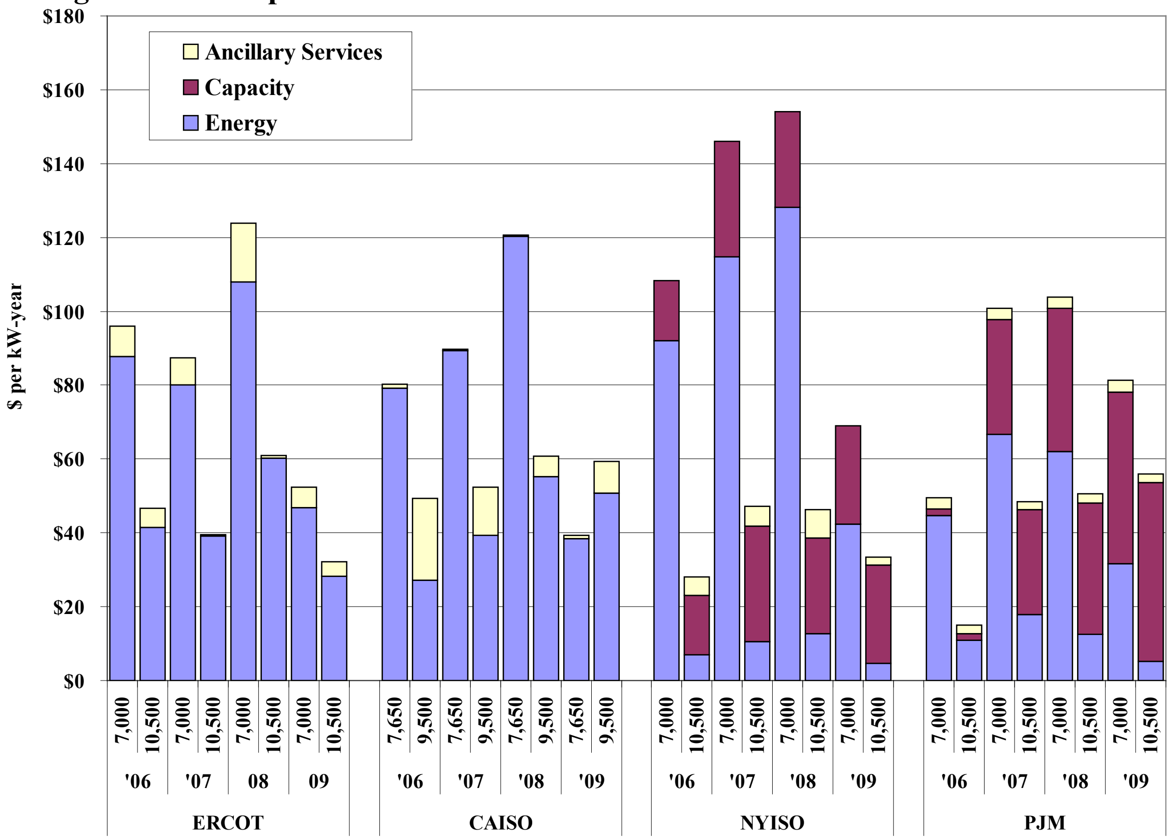


Figure 43 shows that net revenues decreased in all markets from 2008 to 2009, with the exception of gas peaking units in California ISO and PJM that remained flat. In the figure above, net revenues are calculated for central locations in each of the five markets. However, there are load pockets within each market where net revenue and the cost of new investment may be higher. Thus, even if new investment is not generally profitable in a market, it may be economic in certain areas. Finally, resource investments are driven primarily by forward price expectations, so historical net revenue analyses do not provide a complete picture of the future pricing expectations that will spur new investment.

F. Effectiveness of the Scarcity Pricing Mechanism

The PUCT adopted rules in 2006 that define the parameters of an energy-only market. These rules include a Scarcity Pricing Mechanism ("SPM") that relaxed the existing system-wide offer cap by gradually increasing it to $1,500 per MWh on March 1, 2007, $2,250 per MWh on March 1, 2008, and to $3,000 per MWh shortly after the implementation of the nodal market.

Additionally, market participants controlling less than five percent of the capacity in ERCOT by definition do not possess market power under the PUCT rules. Hence, these participants can submit very high-priced offers that, per the PUCT rule, will not be deemed to be an exercise of market power. However, because of the competition faced by the small market participants, the quantity offered at such high prices – if any – is very small.

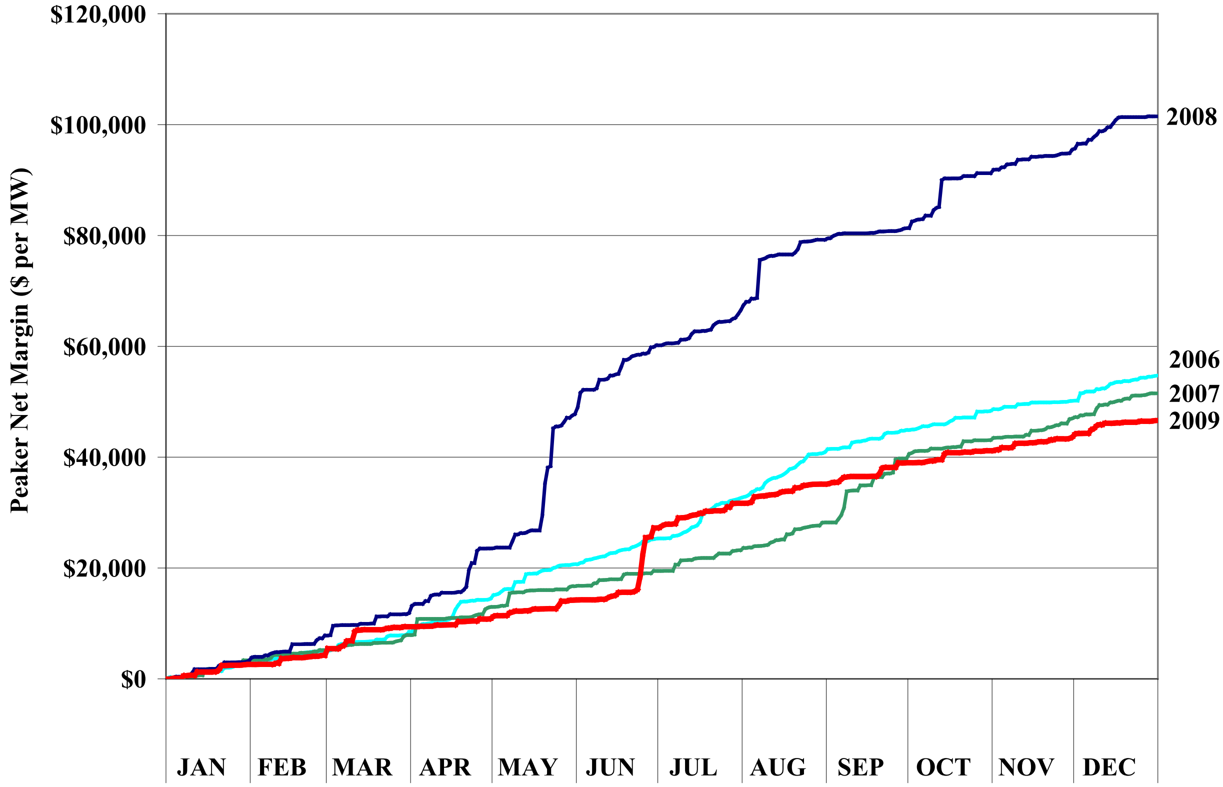
PUCT Subst. Rule 25.505 provides that the IMM may conduct an annual review of the effectiveness of the SPM. This subsection provides an assessment of the effectiveness of the SPM in 2009 under ERCOT's energy-only market structure.

Unlike markets with a long-term capacity market where fixed capacity payments are made to resources across the entire year regardless of the relationship of supply and demand, the objective of the energy-only market design is to allow energy prices to rise significantly higher during legitimate shortage conditions (*i.e.*, when the available supply is insufficient to simultaneously meet both energy and operating reserve requirements) such that the appropriate price signal is provided for demand response and new investment when required. During non-shortage conditions (*i.e.*, most of the time), the expectation of competitive energy market outcomes is no different in energy-only than in capacity markets.

Hence, in an energy-only market, it is the expectation of both the magnitude of the energy price during shortage conditions and the frequency of shortage conditions that will attract new investment when required. In other words, the higher the price during shortage conditions, the fewer shortage conditions that are required to provide the investment signal, and vice versa. While the magnitude of price expectations is determined by the PUCT energy-only market rules, it remains an empirical question whether the frequency of shortage conditions over time will be optimal such that the market equilibrium produces results that satisfy the reliability planning requirements (*i.e.*, the maintenance of a minimum 12.5 percent planning reserve margin).

The SPM includes a provision termed the Peaker Net Margin ("PNM") that is designed to measure the annual net revenue of a hypothetical peaking unit. Under the rule, if the PNM for a year reaches a cumulative total of $175,000 per MW, the system-wide offer cap is then reduced to the higher of $500 per MWh or 50 times the daily gas price index. Figure 44 shows the cumulative PNM results for each year from 2006 through 2009.[21]

Figure 44: Peaker Net Margin



As previously noted, the net revenue required to satisfy the annual fixed costs (including capital carrying costs) of a new gas turbine unit is approximately $70 to $95 per kW-year (i.e., $70,000 to $95,000 per MW-year). Thus, as shown in Figure 44 and consistent with the previous findings in this section relating to net revenue, the PNM reached the level sufficient for new entry in only one of the last four years (2008). In 2008, the peaker net margin and net revenue values rose substantially, surpassing the level required to support new peaker entry. However, a significant portion of the net revenue increase in 2008 was associated with extremely inefficient transmission congestion management and inefficient pricing mechanisms associated with the deployment of non-spinning reserves.[22] Both of these issues were corrected in the zonal market and will be further improved with the implementation of the nodal market in late 2010. With

[21] 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.

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

these issues addressed, the peaker net margin dropped substantially in 2009. Net revenues also dropped substantially for other technologies largely due to significant decreases in natural gas prices in 2009, but decreased natural gas price are not the driver for the reduction in net revenues for peaking resources. Beyond the correction of the market design inefficiencies that existed in 2008, there were three other factors that influenced the effectiveness of the SPM in 2009:

- A continued strong positive bias in ERCOT's day-ahead load forecast – particularly during summer on-peak hours – that creates the tendency to regularly commit online resources in excess of the quantity required to meet expected demand and operating reserve requirements;
- The implementation of PRR 776, which allows for quick-start gas turbines providing non-spinning reserves to offer the capacity into the balancing energy market; and
- The dependence on market participants to submit offers at or near the offer cap to produce scarcity level prices during legitimate operating reserve shortage conditions.

1. ERCOT Day-Ahead Load Forecast Error

ERCOT procedures include the operation of a day-ahead Replacement Reserve Service ("RPRS") market that is designed to ensure that adequate capacity is available on the system to meet reliability criteria for each hour of the following operating day. This includes an assessment of the capacity necessary to meet forecast demand and operating reserve requirements, as well as capacity required resolve transmission constraints.

An integral piece of the RPRS market is the day-ahead load forecast. If the day-ahead load forecast is significantly below actual load and no subsequent actions are taken, ERCOT may run the risk of there not being enough generating capacity online to meet reliability criteria in real-time. In contrast, if the day-ahead load forecast is significantly high, the outcome may be an inefficient commitment of excess online capacity in real-time.

Figure 45 shows the day-ahead load forecast error data for 2007 through 2009 with the average megawatt error displayed for each month in four hour blocks (hours ending). This figure shows a continuing bias toward over-forecasting summer peak loads by an average of 2,000MW.

Figure 45: Average Day Ahead Load Forecast Error by Month and Hour Blocks

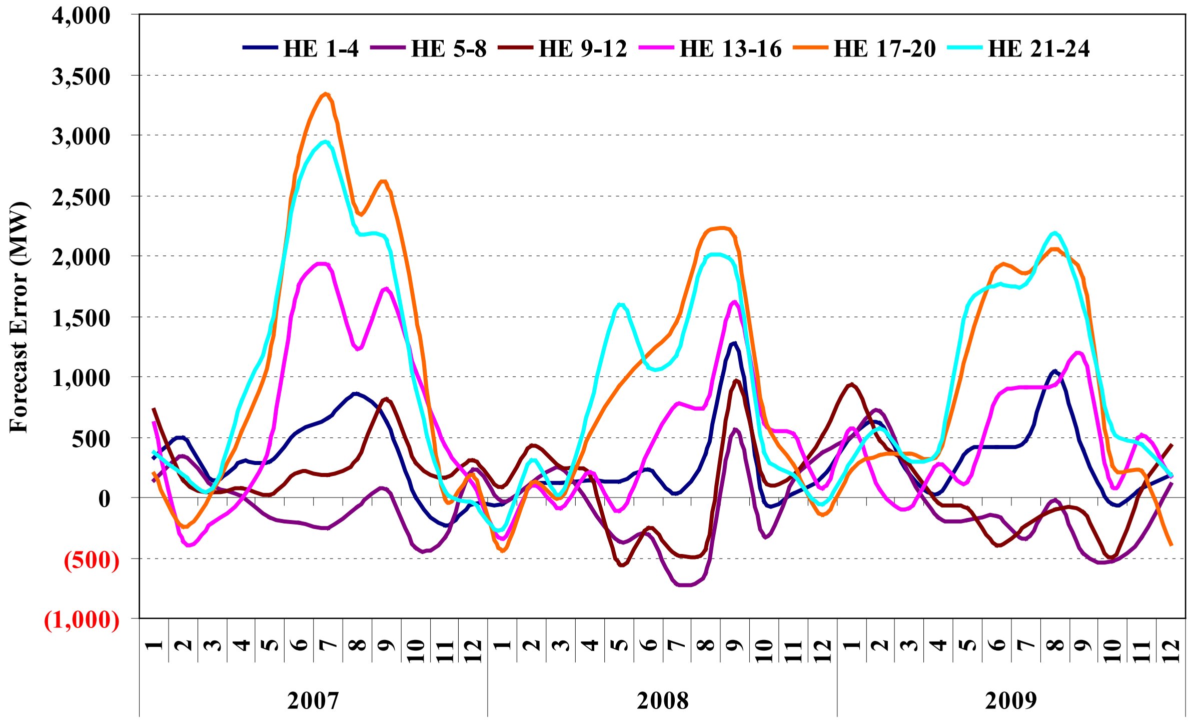
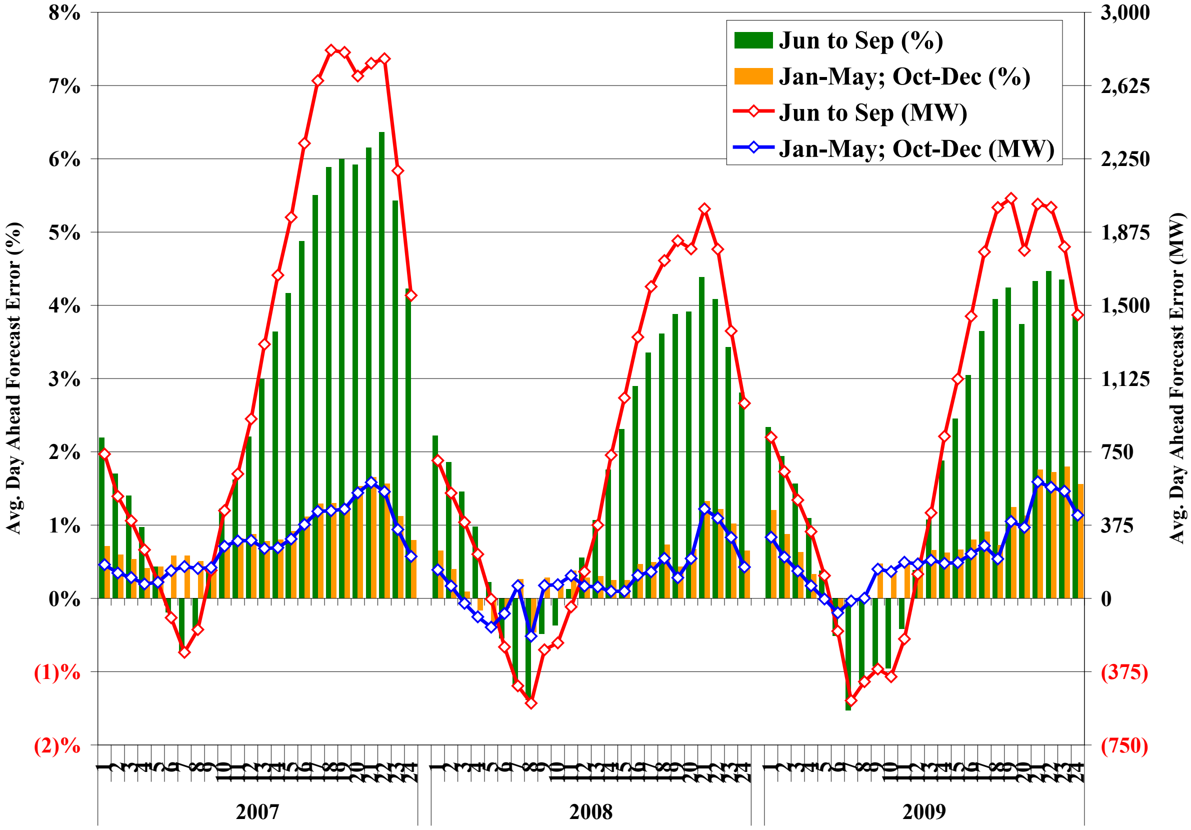


Figure 46 shows the average hourly day-ahead load forecast error for the summer months of June through September, and also for the months of January through May and October through December for 2007 through 2009. In this figure, positive values indicate a day-ahead load forecast that was greater than the actual real-time load. These data indicate a positive bias (*i.e.*, over-forecast) in the day-ahead load forecast over almost all hours in 2007 through 2009, with a particularly strong positive bias during the peak demand hours in the summer months. In terms of quantity, hour 17, for example, exhibited an average over-forecast of 300 MW for the non-summer months, and an average over-forecast of 2,000 MW for the four summer months in 2009. Figure 46 clearly shows that the positive day-ahead load forecast bias observed in 2007 and 2008 persists in 2009.

Figure 46: Average Day Ahead Hourly Load Forecast Error by Season



The existence of such a strong and persistent positive bias in the day-ahead load forecast will tend to lead to an inefficient over-commitment of resources and to the depression of real-time prices relative to a more optimal unit commitment. To the extent load uncertainty is driving the bias in the day-ahead load forecast, such uncertainty is more efficiently managed through the procurement of ancillary services such as non-spinning reserve, or through supplemental commitments of short-lead time resources at a time sufficiently prior to, but closer to real-time as uncertainty regarding real-time conditions diminishes.

In response to these observations in 2009 and prior years, the 2010 ERCOT ancillary service procurement methodologies was modified to adjust the ERCOT day-ahead load forecast to account for the historically measured net load forecast bias, and to compensate for this adjustment by increasing the quantity of non-spinning reserves procured. Although this solution is not ideal because it does not directly address the source of the forecast error bias, it is expected to have a positive effect toward reducing the average forecast error bias in 2010.

2. Implementation of PRR 776

Protocol Revision No. 776 related to the deployment and pricing of non-spinning reserve deployments was implemented in May 2009. Among other changes, the implementation of PRR 776 was expected to provide the following improvements related to non-spinning reserve deployments:

- Eliminate the previous *ex post* re-pricing provisions to provide for *ex ante* pricing during non-spinning reserve deployments, thereby providing more pricing certainty for resources and loads and significantly reducing the probability of *ex post* scarcity level prices during non-scarcity conditions;
- Allow quick start units providing non-spinning reserves to offer in the balancing energy market at a market-based price reflecting the cost and risks of starting and deploying these resources; and
- Reduce the probability of transitional shortages by providing more timely access to these reserves through the balancing energy market instead of manual operator deployments.

Generally, the implementation of PRR 776 performed as expected in 2009, providing increased efficiencies in market operations and pricing during the deployment of non-spinning reserves. As expected, the implementation of PRR 776 also significantly reduced the number of shortage intervals in 2009, as further discussed in the next subsection.

3. Dependence on High-Priced Offers by Market Participants

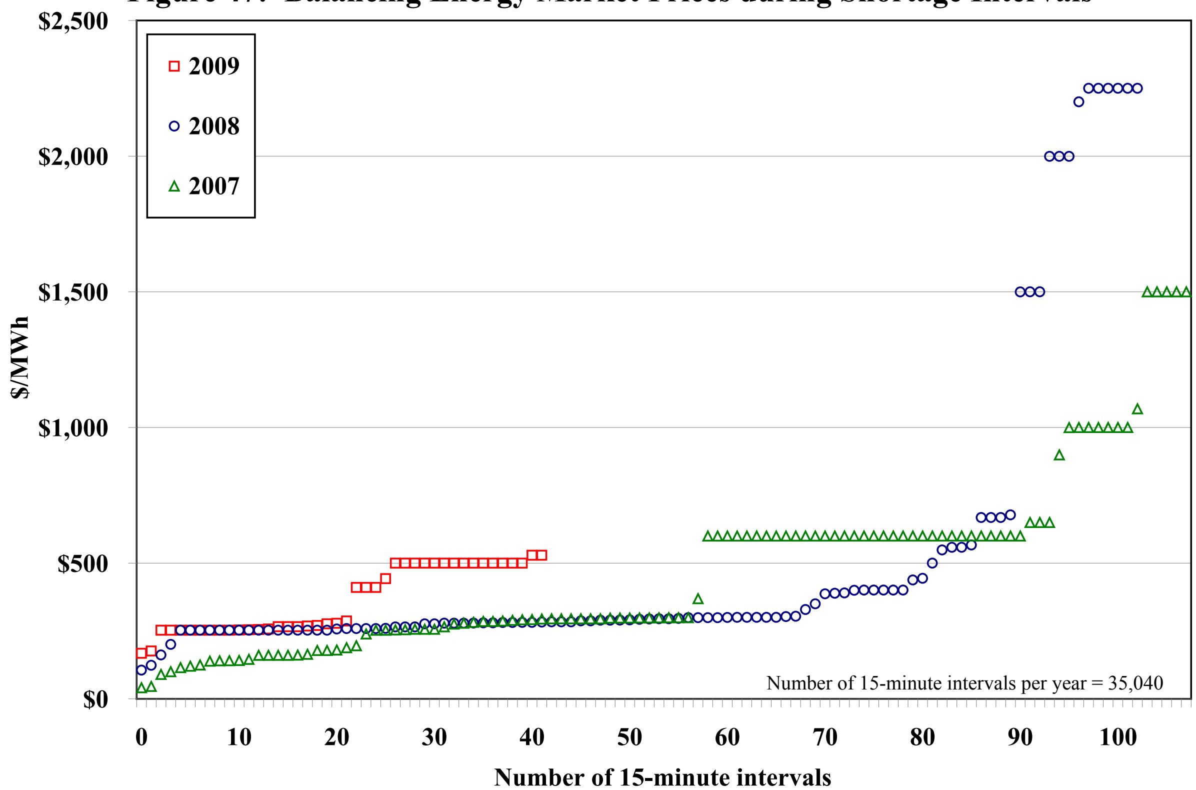
As a general principle, competitive and efficient market prices should be consistent with the marginal cost of the marginal action taken to satisfy the market's demand. In the vast majority of hours, the marginal cost of the marginal action is that associated with the dispatch of the last generator required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

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

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

Under the PUCT rules governing the energy-only market, the mechanism that allows for such pricing during shortage conditions relies upon the submission of high-priced offers by small market participants. Figure 47 shows the balancing market clearing prices during the 15-minute shortage intervals in 2007-2009.

Figure 47: Balancing Energy Market Prices during Shortage Intervals

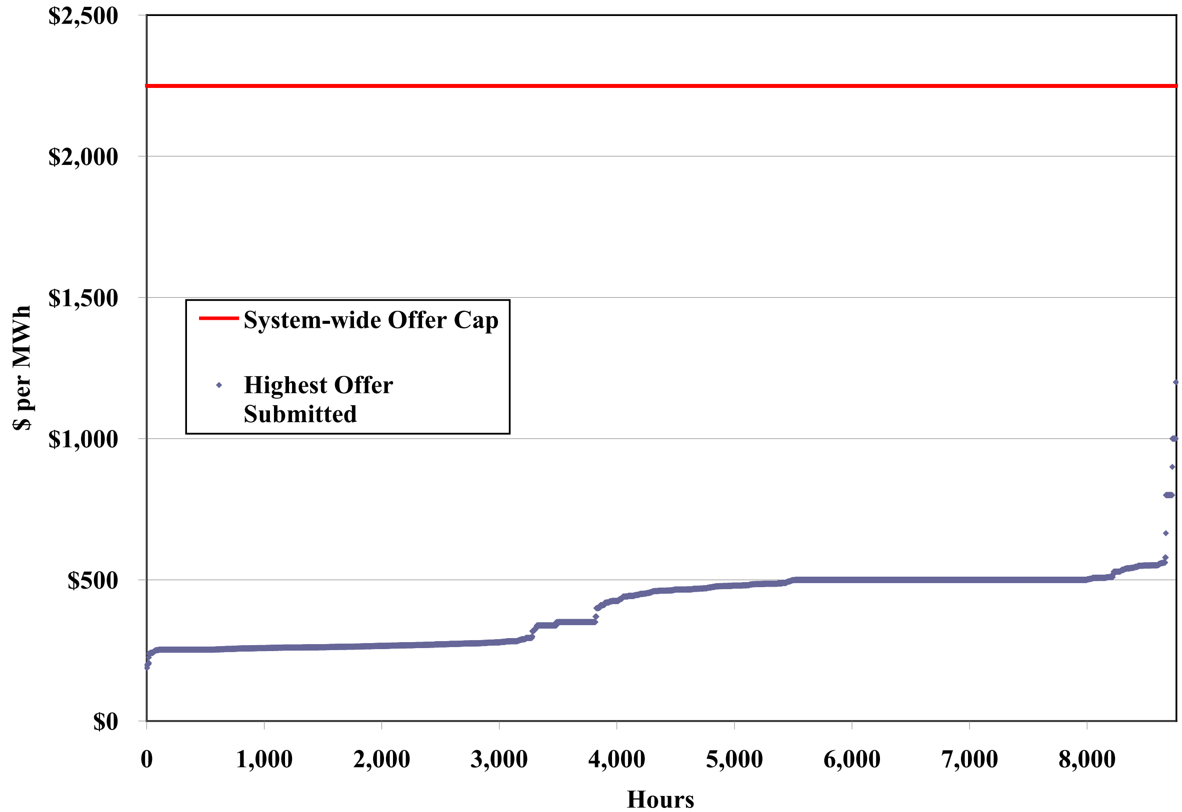


The 42 shortage intervals in 2009 are significantly fewer than the 108 and 103 shortage intervals that occurred in 2007 and 2008, respectively. This reduction can be primarily attributed to the implementation of PRR 776, which allowed more timely access to non-spinning reserves through the balancing energy market, thereby reducing the probability of transitional shortages of the core operating reserves. As shown in Figure 47, the prices during these 42 shortage intervals in 2009 ranged from $168 per MWh to $529 per MWh, with an average price of $364 per MWh and a median price of $283 per MWh.

Although each of the data points in Figure 47 represents system conditions in which the market was in shortage, the pricing outcomes are widely varied, with the majority of prices reflecting the marginal offer of the most expensive generation resource dispatched as opposed to the value of foregone operating reserves. Had an offer been submitted that established the MCPE at the system-wide offer cap in each of the 42 shortage intervals, the 2009 annual peaker net margin would have increased from $46,650 to $66,450 per MW-year, an increase of over 42 percent. The associated increase in the annual load-weighted average balancing energy price would have been less significant, increasing from $34.03 to $36.68 per MWh, an increase of 7.8 percent.

These results indicate that relying exclusively upon the submission of high-priced offers by market participants was generally not a reliable means of producing efficient scarcity prices during shortage conditions in 2007 through 2009. In fact, although the current system-wide offer cap is $2,250 per MWh (as represented by the maximum value of the y-axis in Figure 47), there were no hours in 2009 where an offer was submitted by a market participant that approached the offer cap. Figure 48 shows the highest balancing energy offer price submitted by all market participants in each hour of 2009, ranked from lowest to highest. This figure shows that there were only 33 hours (0.38 percent) with an offer that exceeded $1,000 per MWh, and the average of the highest offers submitted by any market participant in all hours in 2009 was approximately $400 per MWh.

Figure 48: Highest Hourly Balancing Energy Offer Prices



Despite the mixed and widely varied results of the SPM, private investment in generation capacity in ERCOT has continued, although such investment has been dominated by baseload (non-natural gas fueled) and wind generation. As indicated in the net revenue analyses, these investments have been largely driven by significant increases in natural gas prices in the four years prior to 2009. In contrast, private investment in peaking resources in ERCOT has been relatively thin. In an energy-only market, net revenue expectations for peaking resources are much more sensitive to the effectiveness of the shortage pricing mechanism than to the magnitude of natural gas prices.

More reliable and efficient shortage pricing could be achieved by establishing pricing rules that automatically produce scarcity level prices when operating reserve shortages exist. Such an approach would be more reliable because it would not be dependent upon the submission of high-priced offers by small market participants to be effective. It would also be more efficient during the greater than 99 percent of time in which shortage conditions do not exist because it would not be necessary for market participants to effectively withhold lower cost resources by offering relatively small quantities at prices dramatically higher than their marginal cost.

At least for the pendency of the zonal market, shortage pricing will remain dependent upon the existence of high-priced offers by market participants, and results such as those experienced in 2007 through 2009 will continue to frustrate the objectives of the energy-only market design. Further, although presenting some improvements, the nodal market design does not have a complete set of mechanisms to ensure the production of efficient prices during operating reserve shortage conditions. While important even in markets with a capacity market, efficient operating reserve shortage pricing is a particularly critical element in the ERCOT energy-only market to ensure that the long-term resource adequacy requirements are achieved.

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. In ERCOT, constraints on the transmission network are managed in two ways. First, ERCOT is made up of zones with the constraints between the zones managed through the balancing energy market. The balancing energy market model increases energy production in one zone and reduces it in another zone to manage the flows between the two zones when the interface constraint is binding, *i.e.*, when there is interzonal congestion. Second, all other constraints not defined as zonal constraints (*i.e.*, local congestion) are managed through the redispatch of individual generating resources. In this section of the report we evaluate the ERCOT transmission system usage and analyze the costs and frequency of transmission congestion.

A. Electricity Flows between Zones

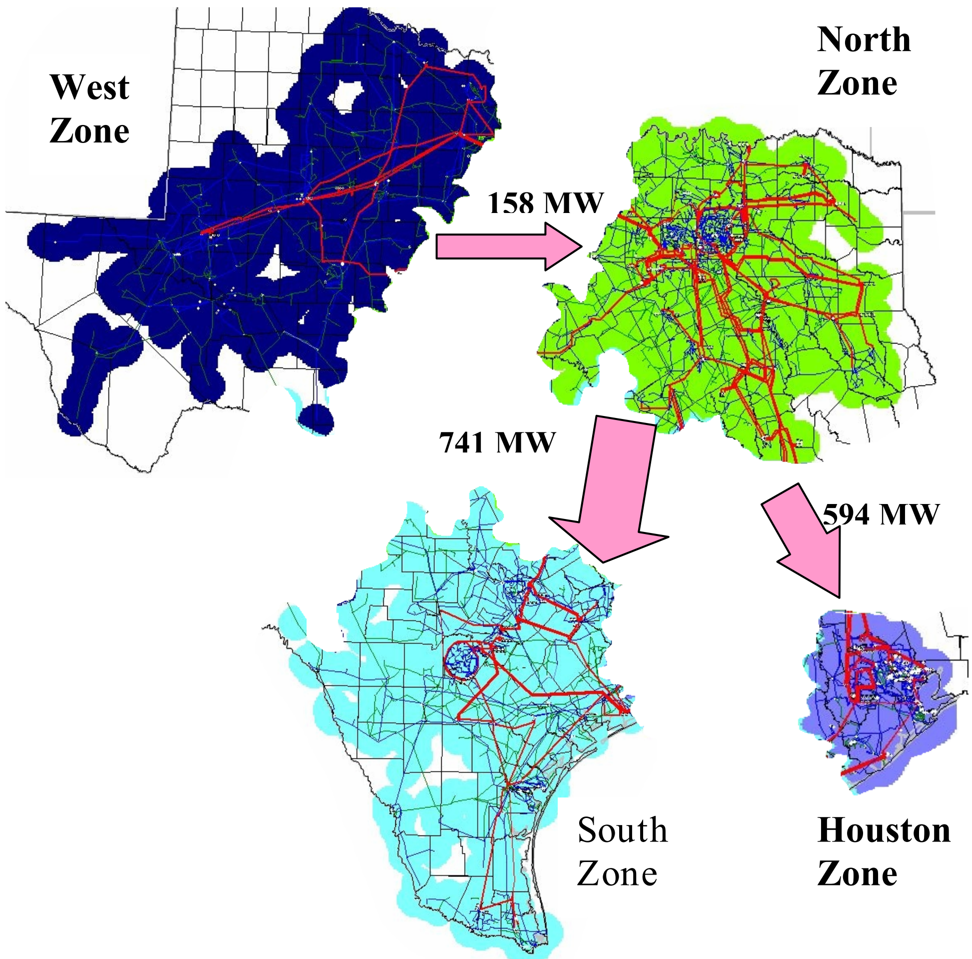
In 2009 there were four commercial pricing zones in ERCOT: (a) the North Zone, (b) the West Zone, (c) the South Zone, and (d) the Houston Zone. ERCOT operators use the Scheduling, Pricing and Dispatch ("SPD") software to economically dispatch balancing energy in each zone to serve load and manage congestion between zones. The SPD model embodies the market rules and requirements documented in the ERCOT protocols.

To manage interzonal congestion, SPD uses a simplified network model with four zone-based locations and five transmission interfaces. These five transmission interfaces, referred to as Commercially Significant Constraints ("CSCs"), are simplified representations of groups of transmission elements. ERCOT operators use planning studies and real-time information to set limits for each CSC that are intended to utilize the total transfer capability of the CSC. In this subsection of the report, we describe the SPD model's simplified representations of flows between zones and analyze actual flows in 2009.

The SPD model uses zonal approximations to represent complex interactions between generators, loads, and transmission elements. Because the model flows are based on zonal approximations, the estimated flows can depart significantly from real-time physical flows.

Estimated flows that diverge significantly from actual flows are an indication of inaccurate congestion modeling leading to inefficient energy prices and other market costs. This subsection analyzes the impact of SPD transmission flows and constraints on market outcomes.

Figure 49: Average SPD-Modeled Flows on Commercially Significant Constraints During All Intervals in 2009



Note: In the figure above, CSC flows are averaged taking the direction into account. So one arrow shows the average flow for the West-to-North CSC was 158 MW, which is equivalent to saying that the average for the North-to-West CSC was *negative* 158 MW.

Figure 49 shows the four ERCOT geographic zones as well as the five CSCs that interconnect the zones: (a) the West to North interface, (b) the South to North interface, (c) the North to South interface, (d) the North to Houston interface, and (e) the North to West interface. A single arrow is shown for the modeled flows of both the North to West and West to North CSCs and the South to North and North to South CSCs. Based on average SPD modeled flows, the North Zone exports a significant amount of power.

The most important simplifying assumption underlying the zonal model is that all generators and loads in a zone have the same effect on the flows over the CSC, or the same shift factor in relation to the CSC. [23] In reality, the generators and loads within each zone can have widely differing effects on the flows over a CSC. To illustrate this, we compared the flows calculated by using actual generation and zonal average shift factors to the average actual flows that occurred over each CSC. The flows over the North to West and South to North CSCs are not shown separately in the table below since they are equal and opposite the flows for the West to North CSC and North to South CSCs, respectively.

Table 2: Average Calculated Flows on Commercially Significant Constraints Zonal-Average vs. Nodal Shift Factors

CSC 2009	Flows Modeled by SPD (1)	Flows Calculated Using Actual Generation (2)	*Difference* *= (2) - (1)*	Actual Flows Using Nodal Shift Factors (3)	*Difference* *= (3) - (2)*
West-North	158	156	*-2*	207	*51*
North - South	741	688	*-53*	406	*-282*
North-Houston	594	541	*-53*	747	*206*

The first column in Table 2 shows the average flows over each CSC calculated by SPD. The second column shows the average flows over each CSC calculated by using zonal-average shift factors and actual real-time generation in each zone instead of the scheduled energy and balancing energy deployments used as an input in SPD. Although these flows are both calculated using the same zonal-average shift factors, they can differ when the actual generation varies from the SPD generation. This difference is shown in the third column (in italics). These differences indicate that the actual generation levels result in calculated flows on each CSC that vary only slightly from the flows modeled by SPD.

23 For a generator, a shift factor indicates the portion of the incremental output of a unit that will flow over a particular transmission facility. For example, a shift factor of 0.5 would indicate that half of any incremental increase in output from a generator would flow over the interface. A negative shift factor would indicate a decrease in flow on an interface resulting from an increase in generation.

The fourth column in Table 2 reports the actual average flows over each CSC by using nodal shift factors applied to actual real-time generation and load. The difference in flows between columns (3) and (2) is attributable to using zonal average shift factors versus nodal shift factors for generation and load in each zone. These differences in flows are shown in the fifth column (in italics).

These results show that the heterogeneous effects of generators and load in a zone on the CSC flows can cause the actual flows to differ substantially from the SPD-calculated flows. Table 2 shows that by using nodal (actual) shift factors reduced the calculated flows on the North to South interface by 282 MW and increased the calculated flows on the North to Houston CSC by 206 MW.

The use of simplified generation-weighted shift factors prevents the SPD model from efficiently resolving and assigning the costs of interzonal congestion. In the long run, the use of generation-weighted shift factors for loads systematically biases prices, so that buyers in some zones pay too much, and others pay too little. Further, the use of average zonal shift factors creates significant operational challenges for ERCOT in the real-time management of zonal congestion because the response to zonal dispatch instructions can often affect the actual flow on a CSC in a manner that is significantly different than that calculated by the simplified assumptions in the SPD model. In turn, ERCOT will tend to operate the system more conservatively to account for the operational uncertainties introduced by the simplified assumptions in the SPD model, the effect of which is discussed in more detail later in this section.

To provide additional understanding of the electricity flows between zones prior to discussing the details of interzonal congestion in the next subsection, Figure 50 shows the actual average imports of power for each zone in 2009. In this figure, positive values represent imports, and negative values indicate exports.[24]

[24] The Northeast Zone existed in 2005 and 2006, but was merged into a single North Zone in 2007 and 2008. The Northeast zone is included in the North zone for 2005 and 2006 in Figure 50.

Figure 50: Actual Zonal Net Imports

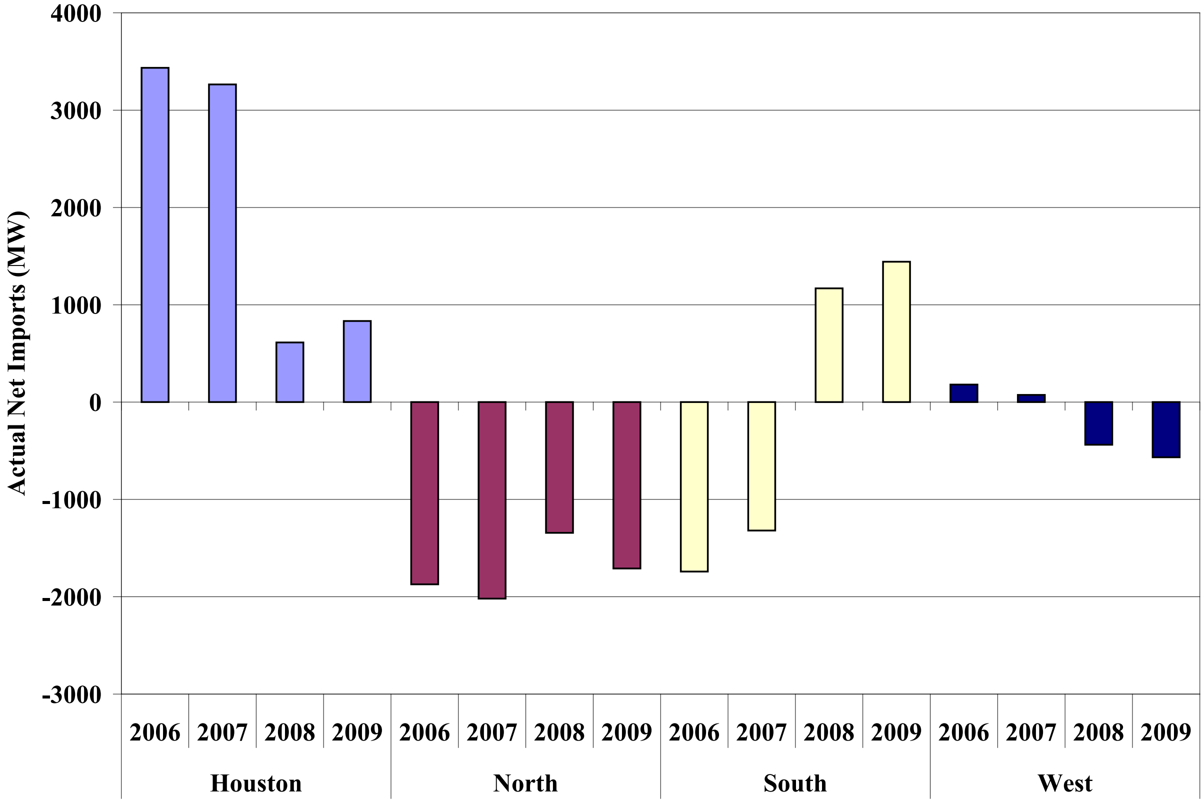


Figure 50 shows that the Houston Zone is a net importer of power, while the North Zone is a net exporter. The reduction in the Houston Zone imports in 2008 and 2009 and corresponding change in the South Zone from a net exporter to a net importer can be attributed to the movement of the 2,700 MW South Texas Nuclear Project from the South Zone to the Houston Zone in 2008. The West Zone transitioned from a net importer in 2006 and 2006 to a net exporter in 2008 and 2009. This reflects the significant increases in the installed capacity of wind resources in the West Zone that occurred over this time period.

B. Interzonal Congestion

The prior subsection showed the average interzonal flows calculated by SPD compared to actual flows in all hours. This subsection focuses on those intervals when the interzonal constraints were binding. Although this excludes most intervals, it is in these constrained intervals that the performance of the market is most critical.

Figure 51 shows the average SPD-calculated flows between the four ERCOT zones during constrained periods for the five CSCs. The arrows show the average magnitude and direction of the SPD-calculated flows during constrained intervals. The frequency with which these constraints arise is shown in parentheses.

Figure 51: Average SPD-Modeled Flows on Commercially Significant Constraints During Transmission Constrained Intervals in 2009

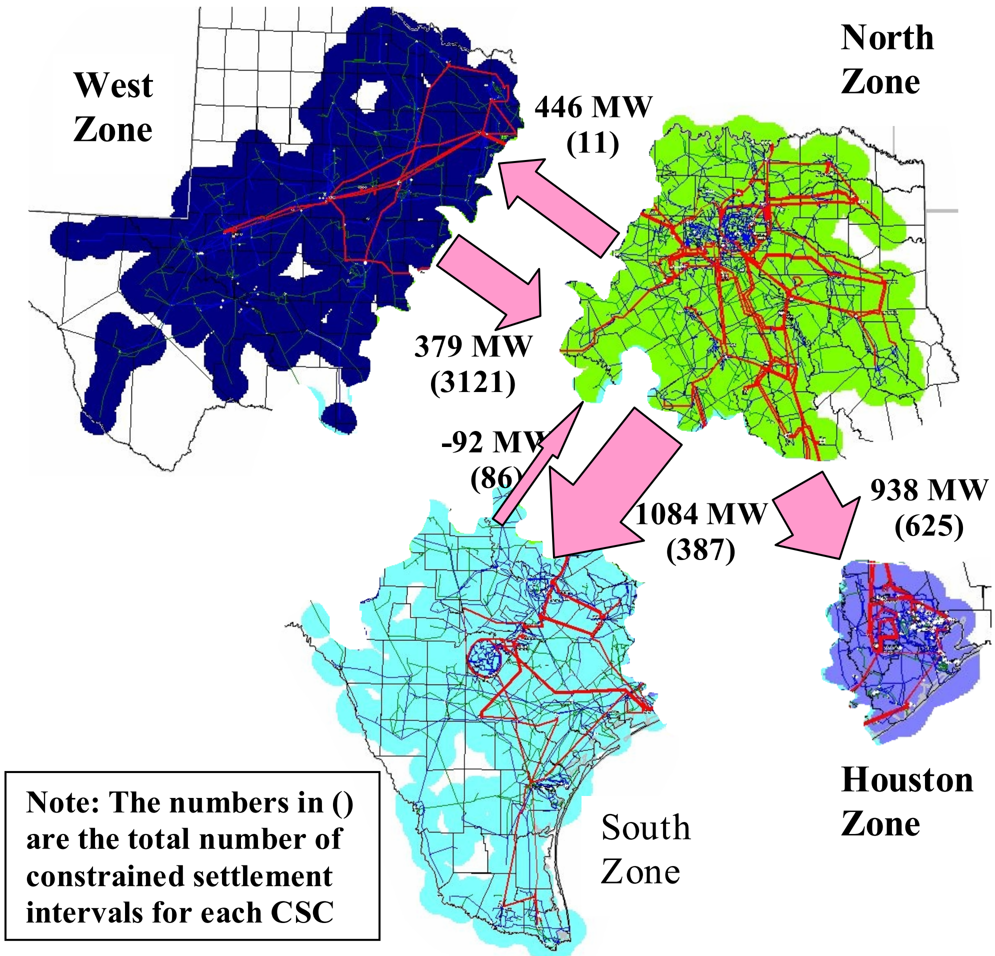


Figure 51 shows that inter-zonal congestion was most frequent in 2009 on the West to North and the North to Houston CSCs, followed by the North to South CSC. The West to North CSC exhibited SPD-calculated flows averaging 379 MW during 3,121 constrained 15-minute intervals (9 percent of the totals intervals in the year). The North to South CSC exhibited SPD-calculated flows averaging 1,084 MW during 387 constrained intervals (1 percent of the total intervals), and the SPD-calculated average flow for the North to Houston CSC was 938 MW during 625 constrained intervals (2 percent of the total intervals).

Table 3: Average Calculated Flows on Commercially Significant Constraints during Transmission Constrained Intervals
Zonal-Average vs. Nodal Shift Factors

CSC 2009	Flows Modeled by SPD (1)	Flows Calculated Using Actual Generation (2)	*Difference = (2) - (1)*	Actual Flows Using Nodal Shift Factors (3)	*Difference = (3) - (2)*
North - South	1084	1058	*-26*	906	*-152*
North - Houston	938	889	*-49*	1171	*282*
South - North	-92	-198	*-106*	209	*406*
West - North	379	435	*56*	383	*-52*
North - West	446	632	*186*	623	*-9*

Table 3 shows data similar to that presented in Table 2, except that the data in Table 3 is limited for each CSC to only those intervals in which the transmission constraint was binding. Table 3 shows that the average SPD-modeled flows for the West to North and North to West CSCs were relatively close to actual flows, whereas the average actual flows for the North to South, South to North and North to Houston CSCs varied significantly from the average flows modeled by SPD.

The following subsections provide a more detailed assessment of the actual occurrences of congestion for each CSC in 2009, with the exception of the North to West CSC that was binding in only eleven 15-minute intervals in 2009.

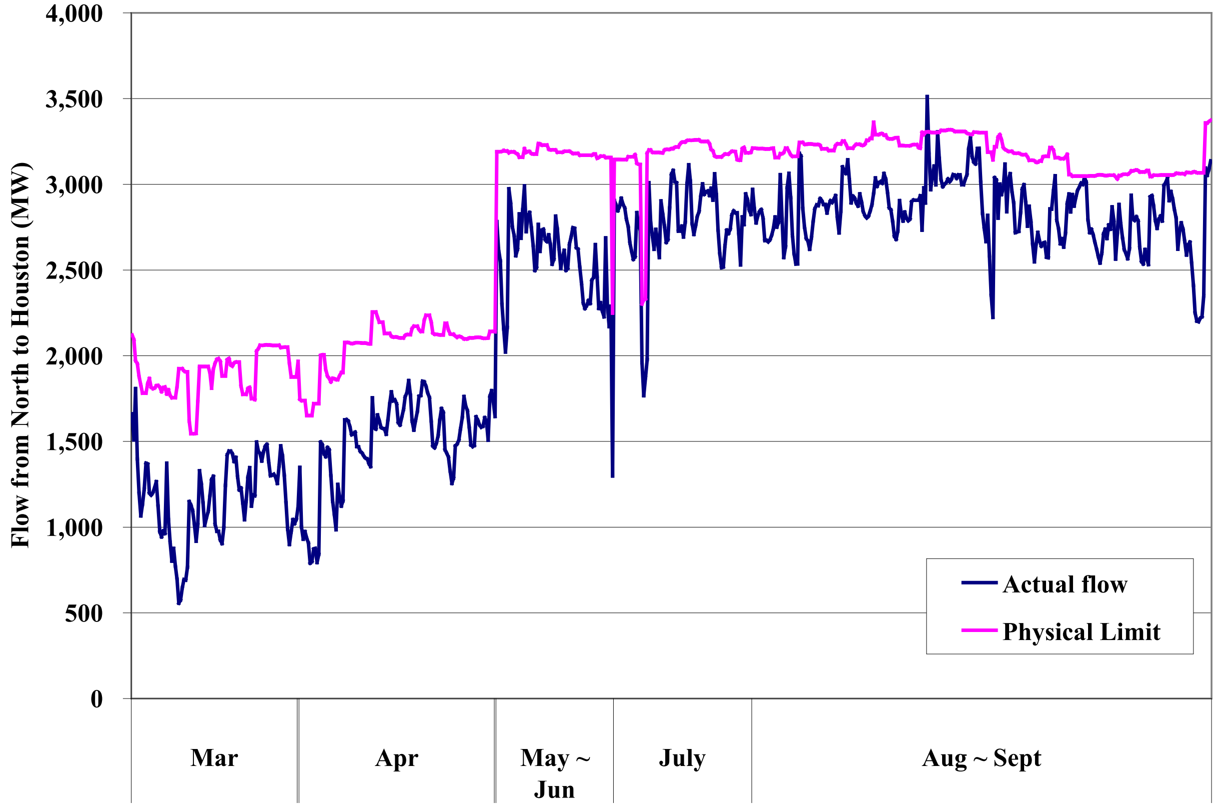
1. Congestion on the North to Houston CSC

The North to Houston CSC was binding in 625 15-minute intervals with an annual average shadow price of $2.01 per MW. These values represent a significant reduction in both the frequency and magnitude of congestion compared to 2008 when the North to Houston CSC was binding in 1,447 intervals with an annual average shadow price of $20.

The decreased congestion on the North to Houston CSC in 2009 is primarily attributable to the implementation of PRR 764 in June 2008 that revised the definition of valid zonal transmission constraints and improved the efficiency of transmission congestion management within the context of the zonal market model.[25] Figure 52 shows the actual flows versus the physical limit for the North to Houston CSC in 2009 during intervals when the CSC was binding.

25 *See* 2008 ERCOT SOM Report at 81-87.

Figure 52: Actual Flows versus Physical Limits during Congestion Intervals North to Houston



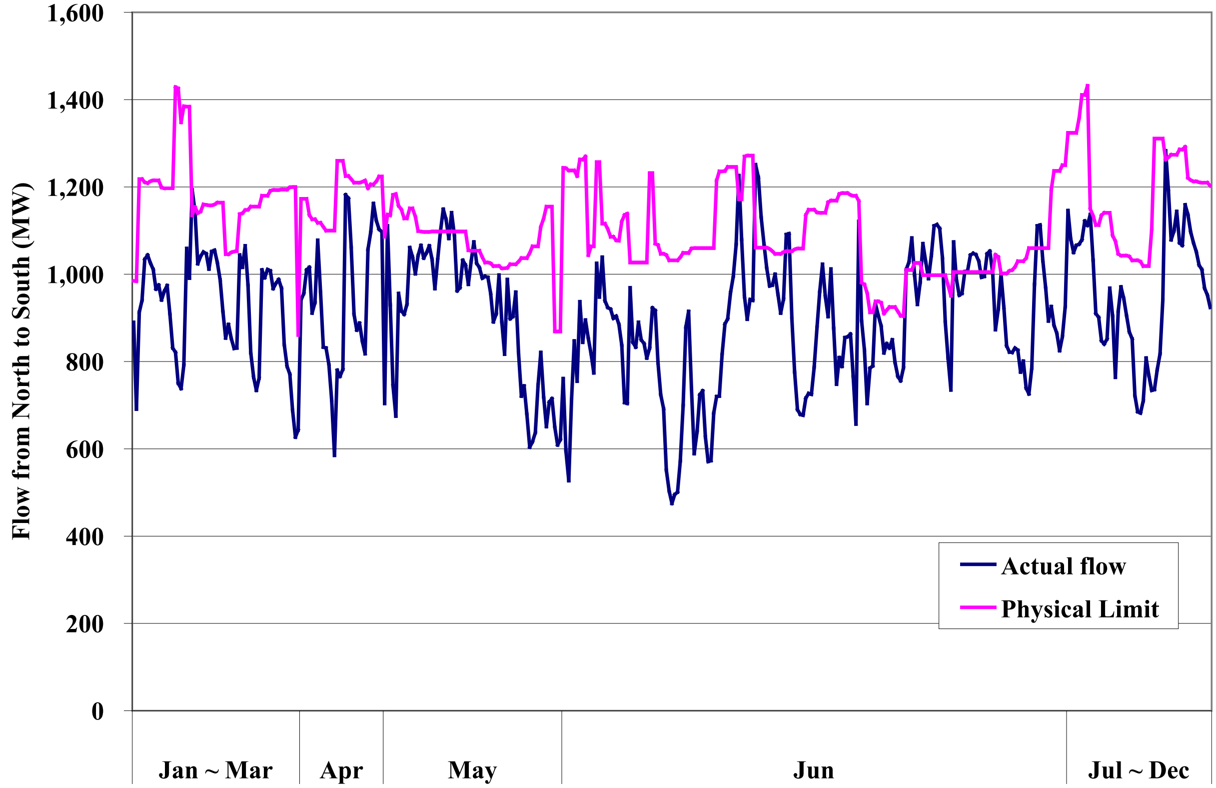
The average physical limit and actual flow for the North to Houston CSC during constrained intervals were 2,772 and 2,290 MW, respectively. Of the 625 intervals that the North to Houston CSC was binding, the actual flow was less than the physical limit in 623 intervals and greater than the physical limit in two intervals. In the 623 intervals where the actual flow was less than the physical limit and the constraint was binding, the average actual flow was 484 MW below the physical limit. In contrast, in the two intervals where the actual flow was greater than the physical limit, the average actual flow was 104 MW above the physical limit.

2. Congestion on the North to South CSC

In 2009 the North to South CSC was binding in 387 15-minute intervals with an annual average shadow price of $8.39 per MW. Like the North to Houston CSC, these values represent a significant reduction in both the frequency and magnitude of congestion compared to 2008 when the North to South CSC was binding in 2,531 intervals with an annual average shadow price of $22.

As was the case for the North to Houston CSC, the reduction in congestion on the North to South CSC in 2009 can be attributed to the implementation of PRR 764. Figure 53 shows the actual flows versus the physical limit for the North to South CSC in 2009 during intervals when the CSC was binding.

Figure 53: Actual Flows versus Physical Limits during Congestion Intervals North to South



The average physical limit and actual flow for the North to South CSC during constrained intervals were 1,117 and 906 MW, respectively, in 2009. Of the 387 intervals that the North to South CSC was binding, the actual flow was less than the physical limit in 353 intervals and greater than the physical limit in 34 intervals. In the 353 intervals where the actual flow was less than the physical limit and the constraint was binding, the average actual flow was 237 MW below the physical limit. In contrast, in the 34 intervals where the actual flow was greater than the physical limit, the average actual flow was 53 MW above the physical limit.

Figure 53 also shows that a significant percentage of the congestion on the North to South CSC occurred during June 2009. During this timeframe, the ERCOT market experienced very high temperatures and associated increases in load levels, as well as a number of outages at baseload generating facilities, particularly in the South Zone. This combination of events led to an increase in the frequency of congestion on the North to South CSC as well as local congestion related to import limitations into the San Antonio area from the north. In the zonal model, the most effective resolution to North to South congestion is to increase generation in the South Zone. However, effective zonal congestion management on the North to South CSC was affected by the local congestion in the San Antonio area, which is most effectively resolved by increasing generation in and South of San Antonio, and decreasing generation north of San Antonio. Because most of the generation resources located north of San Antonio required to decrease output to manage the local congestion in the San Antonio area are also in the South Zone that was broadly required to increase output to manage the zonal North to South congestion, competing reliability objectives were present that complicated the simultaneous resolution of both the North to South zonal congestion and the intrazonal San Antonio import-related congestion. Faced with these competing reliability objectives and the inability to resolve both reliability issues within the context of the zonal model and its bifurcated process of zonal and local transmission congestion management, ERCOT implemented a temporary transmission switching solution in late June that effectively increased the transfer capability on the North to South CSC, thereby resolving these competing reliability objectives under the atypical load and generator outage conditions experienced at that time.

3. Congestion on the West to North CSC

In 2009 the West to North CSC was binding in 3,121 15-minute intervals. This was more frequent than any other CSC in 2009 and, with the exception of the same CSC in 2008 that was binding for 5,320 intervals, more frequent than any other CSC since the inception of single control area operations in 2001. The primary reason for the high frequency of congestion on the West to North CSC in 2008 and 2009 is the significant increase in installed wind generation relative to the load in the West Zone and limited transmission export capability to the broader market.

Average load in the West Zone was 2,023 MW in 2009, with a minimum of 1,588 MW and a maximum of 2,744 MW. The average profile of West Zone wind production is negatively correlated with the load profile, with the highest wind production occurring primarily during the spring, fall and winter months, and predominately during off-peak hours. Figure 54 shows the average West Zone wind production for each month in 2009, with the average production in each month shown separately in four hour blocks.[26]

Figure 54: Average West Zone Wind Production

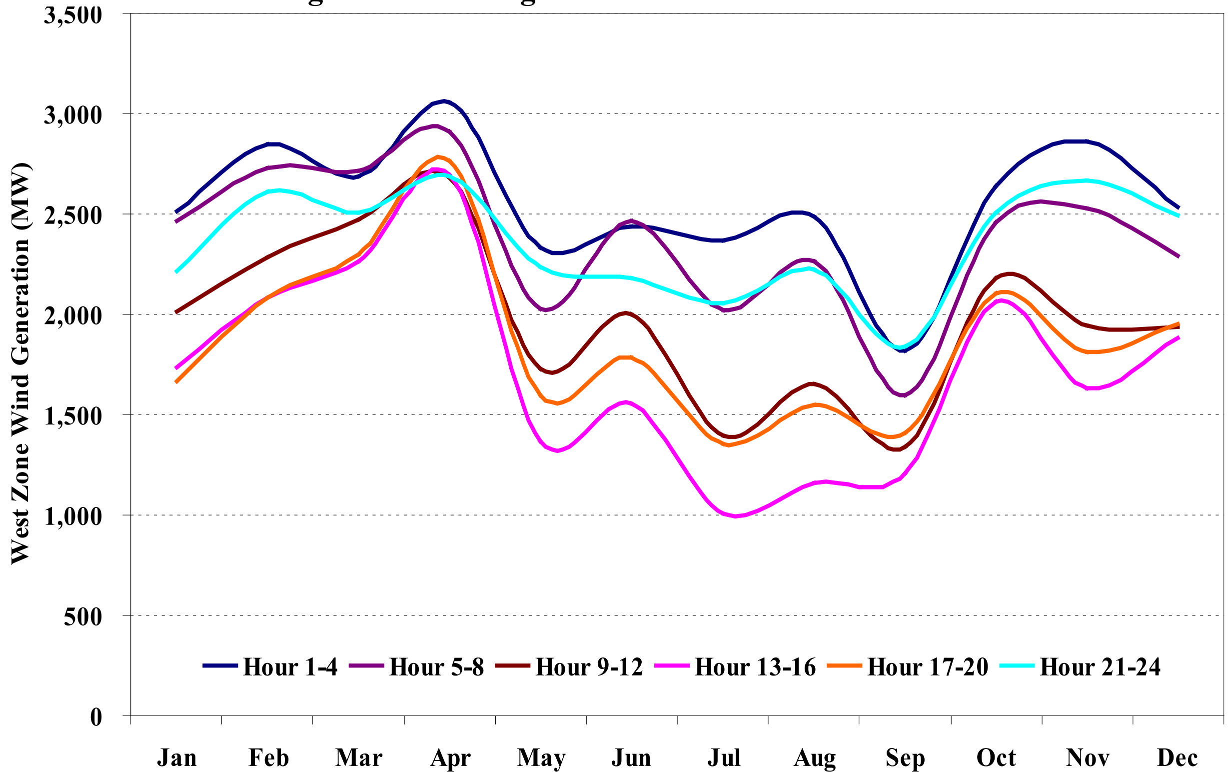
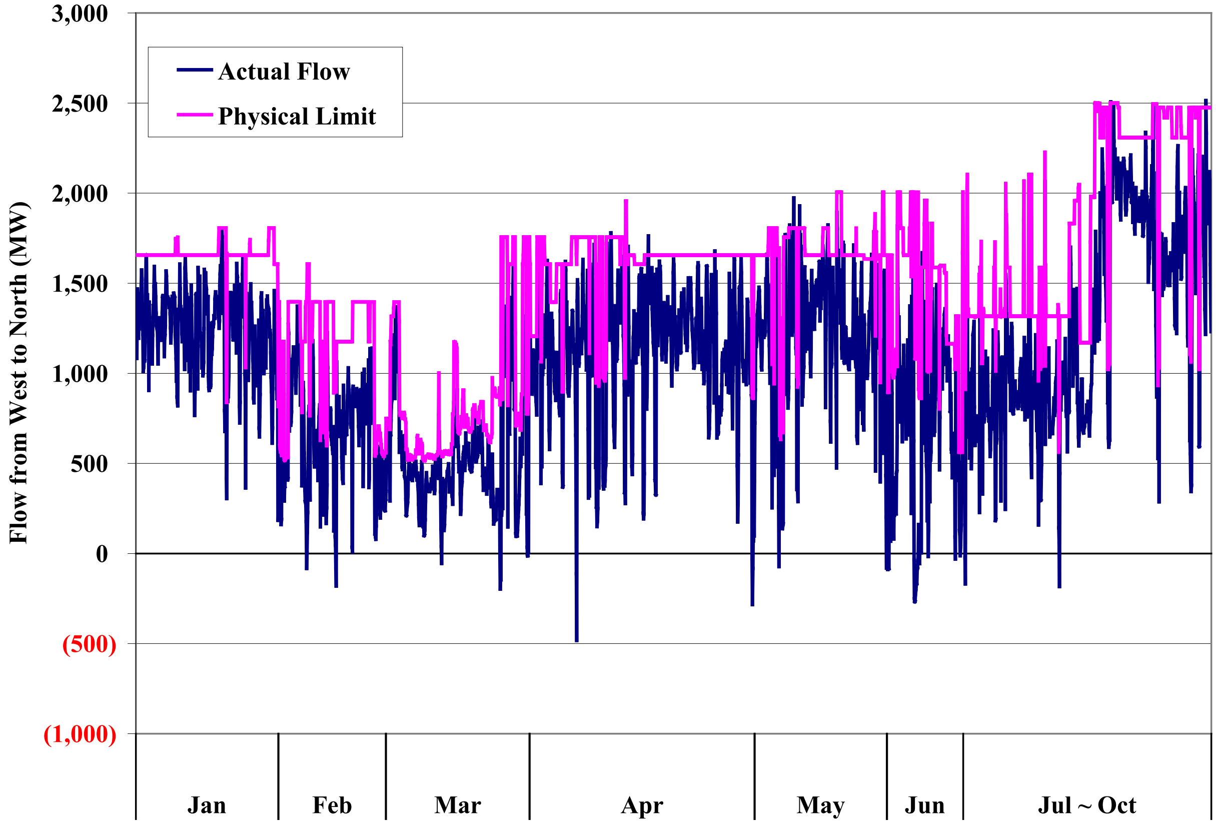


Figure 55 shows the actual flows and the physical limit for the West to North CSC in 2009 for intervals in which the CSC was binding. The average physical limit and actual flow for the West to North CSC during constrained intervals were 1,528 and 1,046 MW, respectively, in 2009. Of the 3,121 intervals that the West to North CSC was binding, the actual flow was less than the physical limit in 3,096 intervals and greater than the physical limit in 25 intervals. In the 3,096 intervals where the actual flow was less than the physical limit and the constraint was binding, the average actual flow was 487 MW below the physical limit. In contrast, in the 25 intervals where the actual flow was greater than the physical limit, the average actual flow was 42 MW above the physical limit.

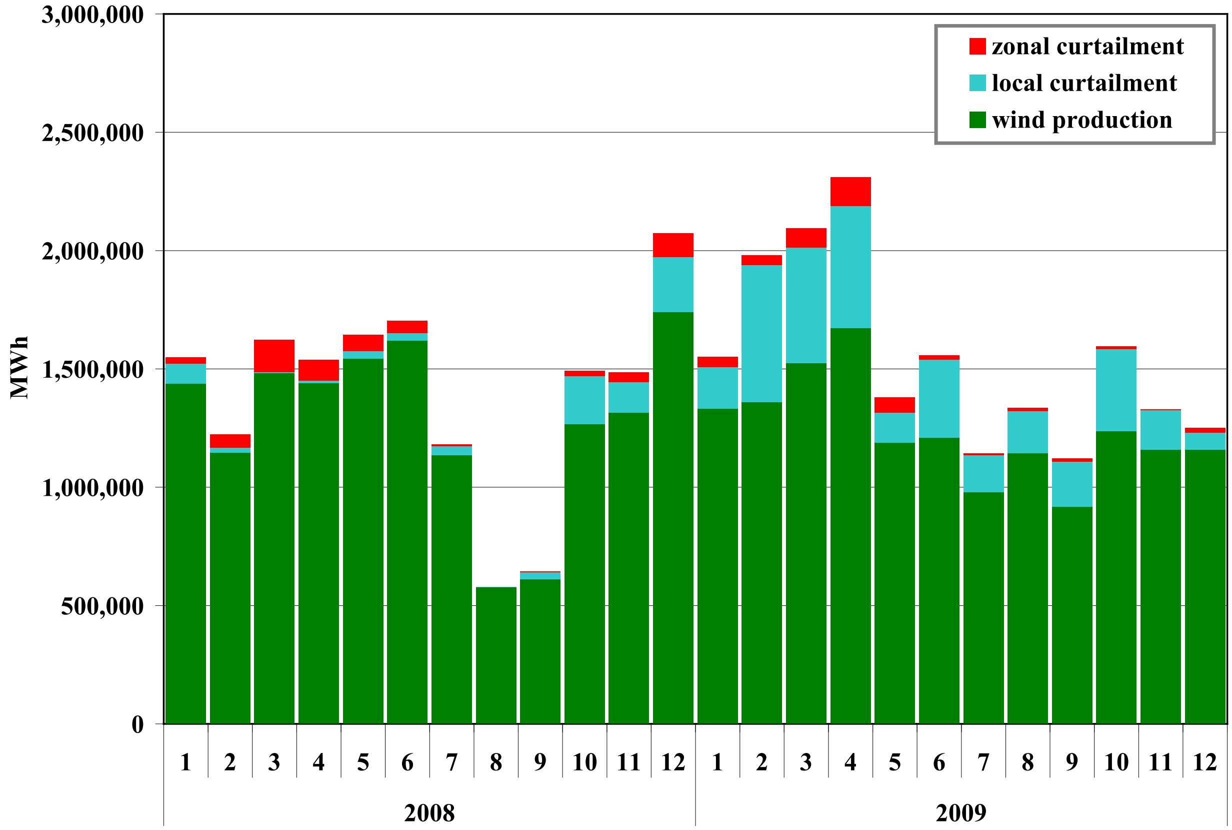
26 Figure 54 shows actual wind production, which was affected by curtailments at the higher production levels in 2008. Thus, the higher levels of actual wind production in Figure 54 are lower than the production levels that would have materialized absent transmission constraints.

Figure 55: Actual Flows versus Physical Limits during Congestion Intervals West to North



Although the frequency of zonal transmission congestion on the West to North CSC was very high in 2009, it was lower than in 2008. However, zonal congestion data do not provide a complete view of the congestion situation in the West Zone. Figure 56 shows the wind production and local and zonal curtailment quantities for the West Zone for each month of 2008 and 2009. This figure reveals that, while the quantity of zonal curtailments for wind resources in the West Zone was reduced from 604,000 MWh in 2008 to 442,000 MWh in 2009, the quantity of local curtailments increased significantly, rising from 812,000 MWh in 2008 to over 3,400,000 MWh in 2009.

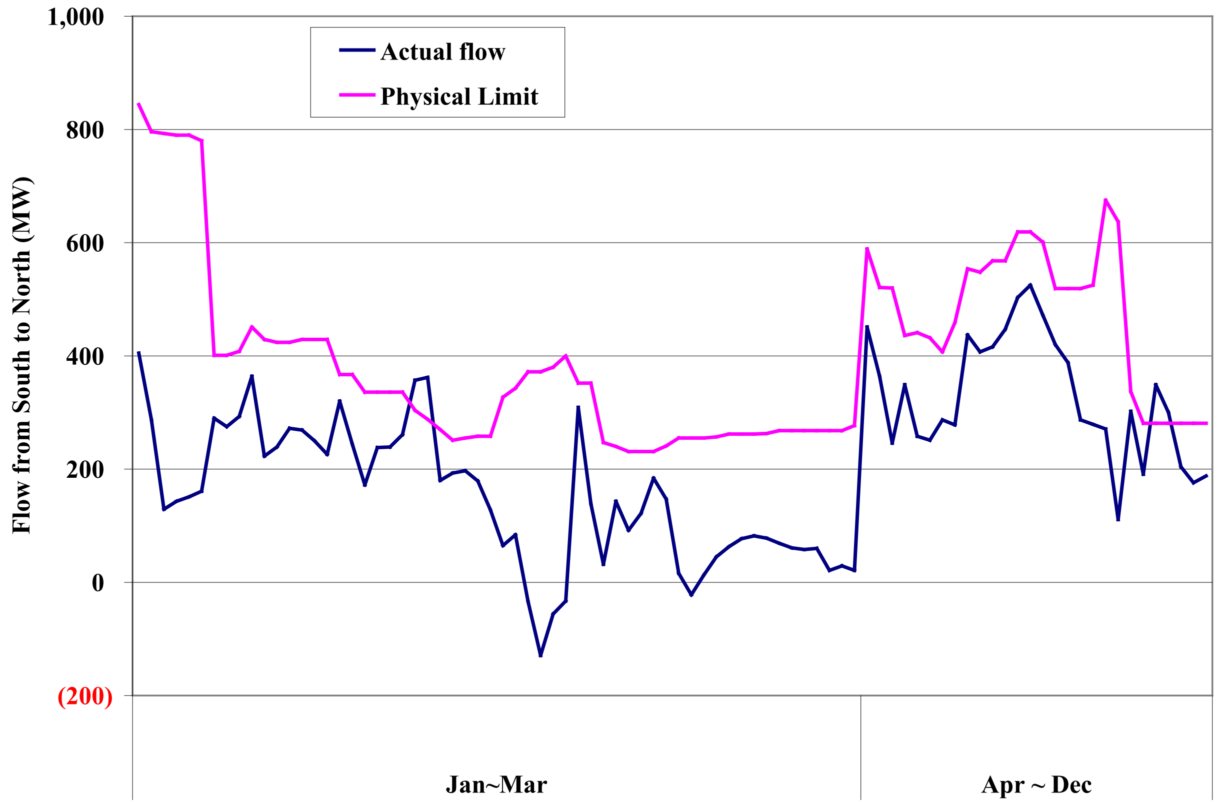
Figure 56: West Zone Wind Production and Curtailment



4. Congestion on the South to North CSC

The South to North CSC was binding in 86 15-minute intervals in 2009. Figure 57 shows the actual flows and the physical limit for the South to North CSC in 2009 for intervals in which the CSC was binding. The average physical limit and actual flow for the South to North CSC during constrained intervals were 402 and 209 MW, respectively, in 2009. Of the 86 intervals that the South to North CSC was binding, the actual flow was less than the physical limit in 82 intervals and greater than the physical limit in four intervals. In the 82 intervals where the actual flow was less than the physical limit and the constraint was binding, the average actual flow was 205 MW below the physical limit. In contrast, in the four intervals where the actual flow was greater than the physical limit, the average actual flow was 53 MW above the physical limit.

Figure 57: Actual Flows versus Physical Limits during Congestion Intervals South to North



5. Zonal Congestion Management Challenges

As discussed in the first part of this section, differences that exist between the commercial SPD model representation and the physical reality create operational challenges for ERCOT to efficiently manage zonal transmission congestion while also maintaining reliable operations. Table 4 shows the average physical limit, actual flow and the difference between the average physical limit and the actual flow for each CSC during binding intervals in 2009.

Table 4: CSC Average Physical Limits vs. Actual Flows during Constrained Intervals

CSC 2009	Average Physical Limit (MW)	Average Actual Flow (MW)	Avg. Physical Limit - Avg. Actual Flow (MW)
North to South	1117	906	211
North to Houston	2772	2290	483
South to North	401	208	193
West to North	1528	1046	483
North to West	780	623	157

Table 4 shows that, for all CSCs in 2009, the average actual flow was considerably less than the average physical limit. For all CSCs combined, the average actual flow was 23 percent less than the average physical limit. To maximize the economic use of the scarce transmission capacity, the ideal outcome would be for the actual flows to reach the physical limits, but not to exceed such limits to maintain reliable operations. However, primarily for the reasons discussed in the first part of this section, achieving such ideal outcomes is practically impossible in the context of the zonal market model. Further, as discussed in relation to the North to South CSC, the bifurcated process of resolving zonal and local congestion can at times lead to reliability conflicts that are difficult to resolve within the relatively inflexible framework of the zonal market design.

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

C. Congestion Rights Market

Interzonal congestion can be significant from an economic perspective, compelling the dispatch of higher-cost resources because power produced by lower-cost resources cannot be delivered over the constrained interfaces. When this constraint occurs market participants must compete to use the available transfer capability between zones. To allocate this capability efficiently, ERCOT establishes clearing prices for energy in each zone that will vary in the presence of congestion and charges the transactions between the zones the interzonal congestion price.

One means by which ERCOT market participants can hedge congestion charges in the balancing energy market is by acquiring Transmission Congestion Rights ("TCRs") or Pre-assigned Congestion Rights ("PCRs"). Both TCRs and PCRs entitle the holder to payments corresponding to the interzonal congestion price. Hence, a participant holding TCRs or PCRs for a transaction between two zones would pay the interzonal congestion price associated with the transaction and receive TCR or PCR payments that offset the congestion charges. TCRs are acquired by annual and monthly auctions (as explained in more detail below) while PCRs are allocated to certain participants based on historical patterns of transmission usage.

To analyze congestion rights in ERCOT, we first review the TCRs and PCRs that were auctioned or allocated for each CSC in 2009. Figure 58 shows the average number of TCRs and PCRs awarded for each of the CSCs in 2009 compared to the average SPD-modeled flows during the constrained intervals.

Figure 58: Transmission Rights vs. Real-Time SPD-Calculated Flows
Constrained Intervals - 2009

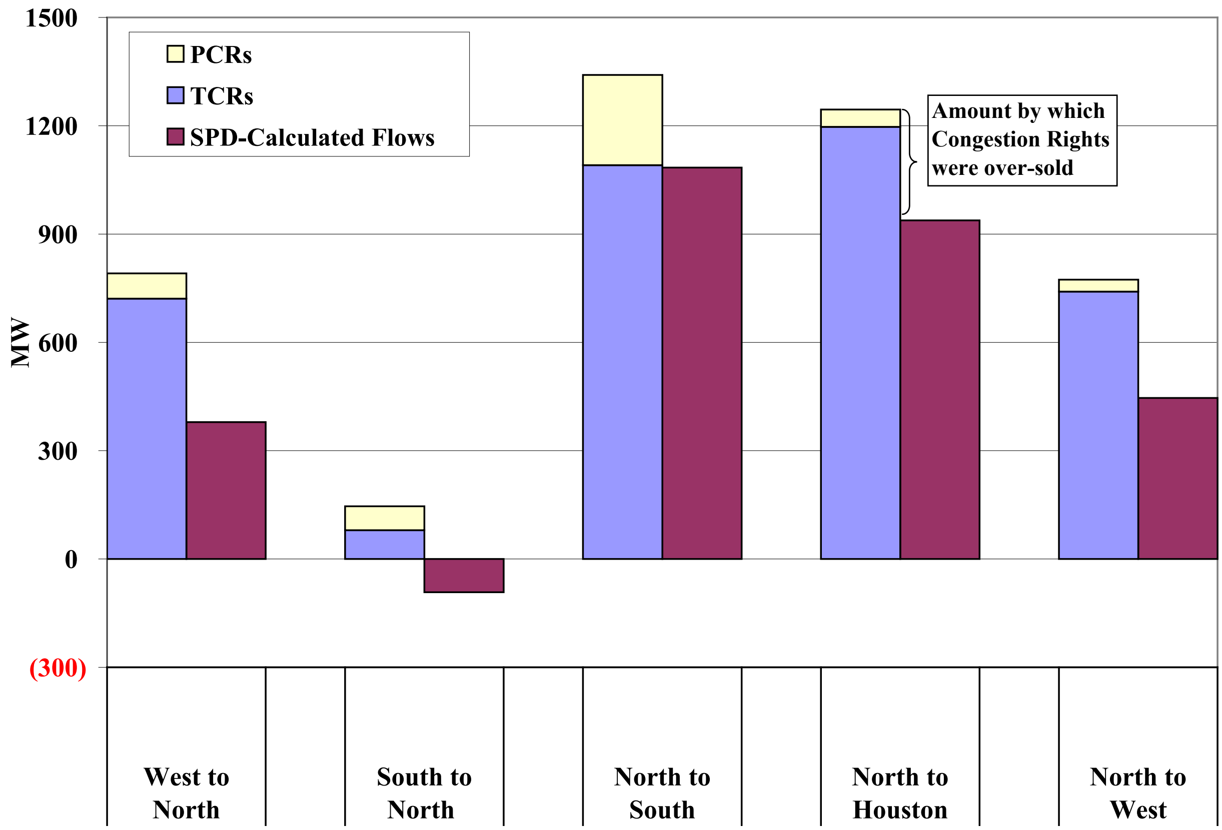


Figure 58 shows that total congestion rights (the sum of PCRs and TCRs) on all the interfaces exceeded the average real-time SPD-calculated flows during constrained intervals. These results indicate that the congestion rights were oversold in relation to the SPD-calculated limits. For example, congestion rights for the North to Houston CSC were oversold by an average of 328 MW. The average amount of TCRs awarded each month in 2009 is higher than in 2008.

Ideally the financial obligations to holders of congestion rights would be satisfied with congestion revenues collected from participants scheduling over the interface and through the sale of balancing energy flowing over the interface. When the SPD-calculated flows are consistent with the quantity of congestion rights sold over the interface, the congestion revenues will be sufficient to satisfy payments to the holders of the congestion rights. Alternatively, when the quantity of congestion rights exceeds the SPD-calculated flow over an interface, congestion revenues from the balancing energy market will not be sufficient to meet the financial obligations to congestion rights holders.

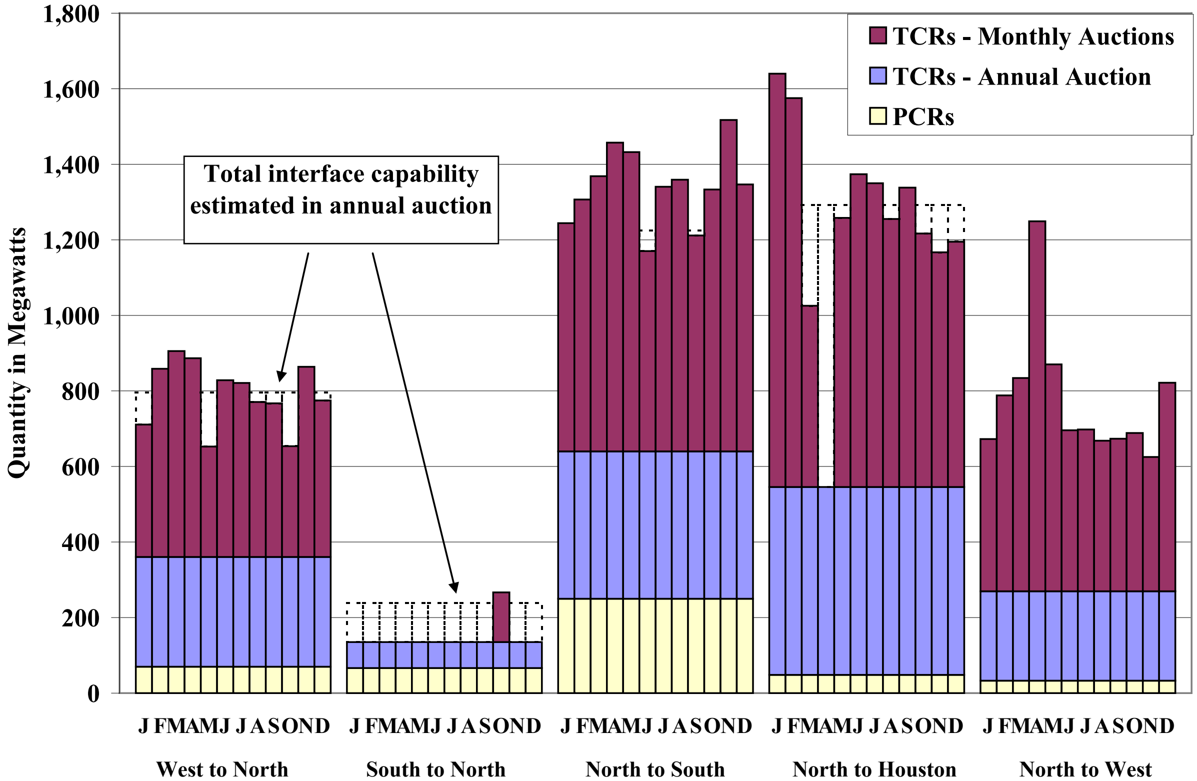
As an example, suppose the SPD-calculated flow limit is 300 MW for a particular CSC during a constrained interval and that holders of congestion rights own a total of 800 MW over the CSC. ERCOT will receive congestion rents from the balancing energy market to cover precisely 300 MW of the 800 MW worth of obligations. Thus, a revenue shortfall will result that is proportional to the shadow price of the constraint on the CSC in that interval (*i.e.*, proportional to the congestion price between the zones). In this case, the financial obligations to the congestion rights holders cannot be satisfied with the congestion revenue, so the shortfall is charged proportionately to all loads in ERCOT as part of the Balancing Energy Neutrality Adjustment ("BENA") charges.

To provide a better understanding of these relationships, we next review ERCOT's process to establish the quantity of congestion rights allocated or sold to participants. ERCOT performs studies to determine the capability of each interface under peak summer conditions. This summer planning study is the basis for offering 40 percent of the available TCRs for sale in the annual auction. These rights are auctioned during December for the coming year. Additional TCRs are offered for sale based on monthly updates of the summer study. Because the monthly studies tend to more accurately reflect conditions that will prevail in the coming month, the monthly designations tend to more closely reflect actual transmission limits.

However, the monthly studies used to designate the TCRs do not always accurately reflect real-time transmission conditions for two main reasons. First, transmission and generation outages can occur unexpectedly and can significantly reduce the transfer capability of a CSC. Even planned transmission outages may not be known to ERCOT when the summer studies are conducted. Second, conditions may arise causing the actual physical flow to be significantly different from the SPD modeled flow. As discussed above, ERCOT operators may need to respond by lowering the SPD-modeled flow limits to manage the actual physical flow. Accordingly, it is likely that the quantity of congestion rights awarded will be larger than available transmission capability in SPD.

To examine how these processes have together determined the total quantity of rights sold over each interface, Figure 59 shows the quantity of each category of congestion rights for each month during 2009. The quantities of PCRs and annual TCRs are constant across all months and were determined before the beginning of 2009, while monthly TCR quantities can be adjusted monthly.

Figure 59: Quantity of Congestion Rights Sold by Type

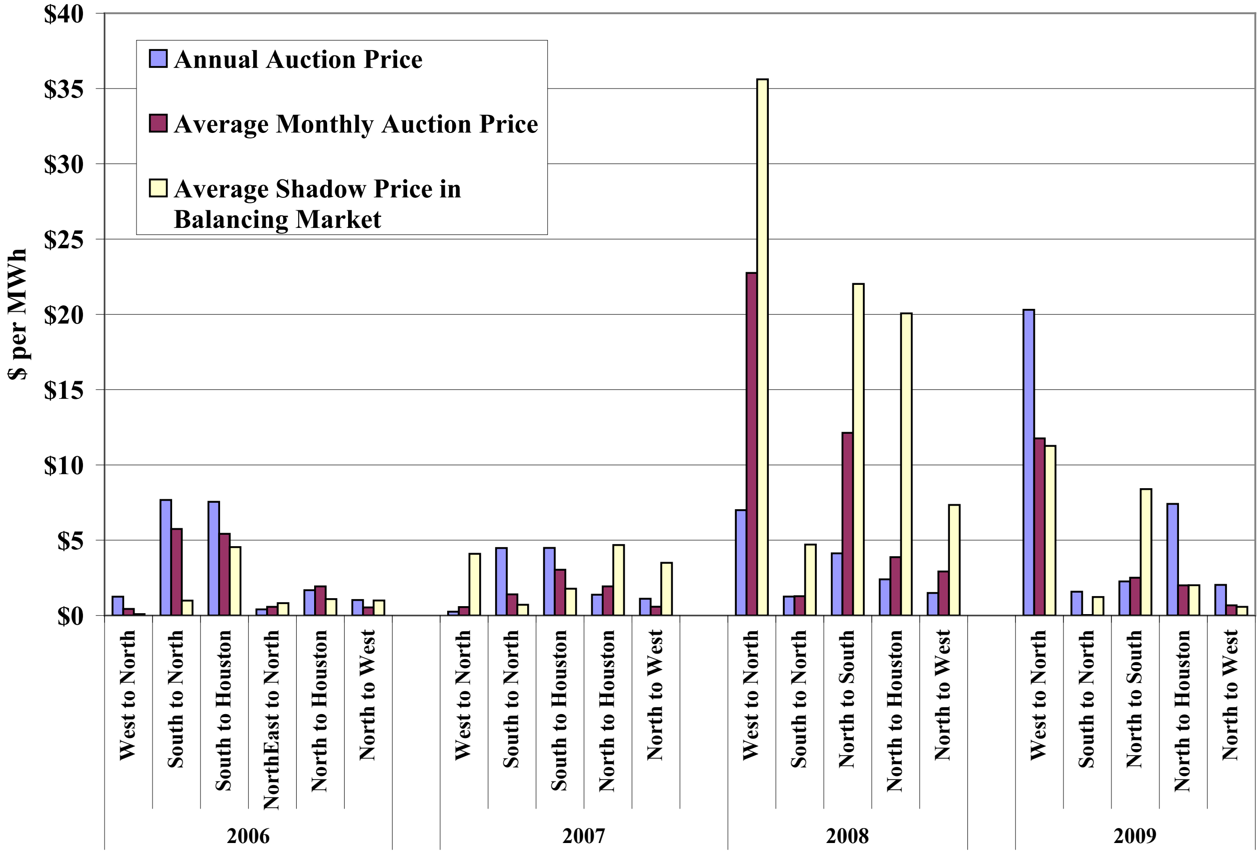


When the monthly planning studies indicate changes from the summer study, revisions are often made to the estimated transmission capability. Therefore, the auctioned congestion rights may increase or decrease relative to the amount estimated in the summer study. The shadow boxes in the figure represent the capability estimated in the summer study that is not ultimately sold in the monthly auction. When there is no shadow box in Figure 59, the total quantity of PCRs and TCRs sold in the annual and monthly auctions equaled or exceeded the summer estimate and therefore no excess capability is shown.

The South to North and North to Houston interfaces experienced the largest fluctuations in the estimates of transmission capacity between the annual auction and the monthly auctions. In fact, for several months South to North TCRs were not even offered for sale by ERCOT. The divergence between annual and monthly estimates of transmission capacity on the other interfaces was smaller.

Market participants who are active in congestion rights auctions are subject to substantial uncertainty. Outages and other contingencies occur randomly and can substantially change the market value of a congestion right. Real-time congestion prices reflect the cost of interzonal congestion and are the basis for congestion payments to congestion rights holders. In a perfectly efficient system with perfect forecasting by participants, the average congestion price should equal the auction price. However, we would not expect full convergence in the real-world, given uncertainties and imperfect information. To evaluate the results of the ERCOT congestion rights market, in Figure 60 we compare the annual auction price for congestion rights, the average monthly auction price for congestion rights, and the average congestion price for each CSC.

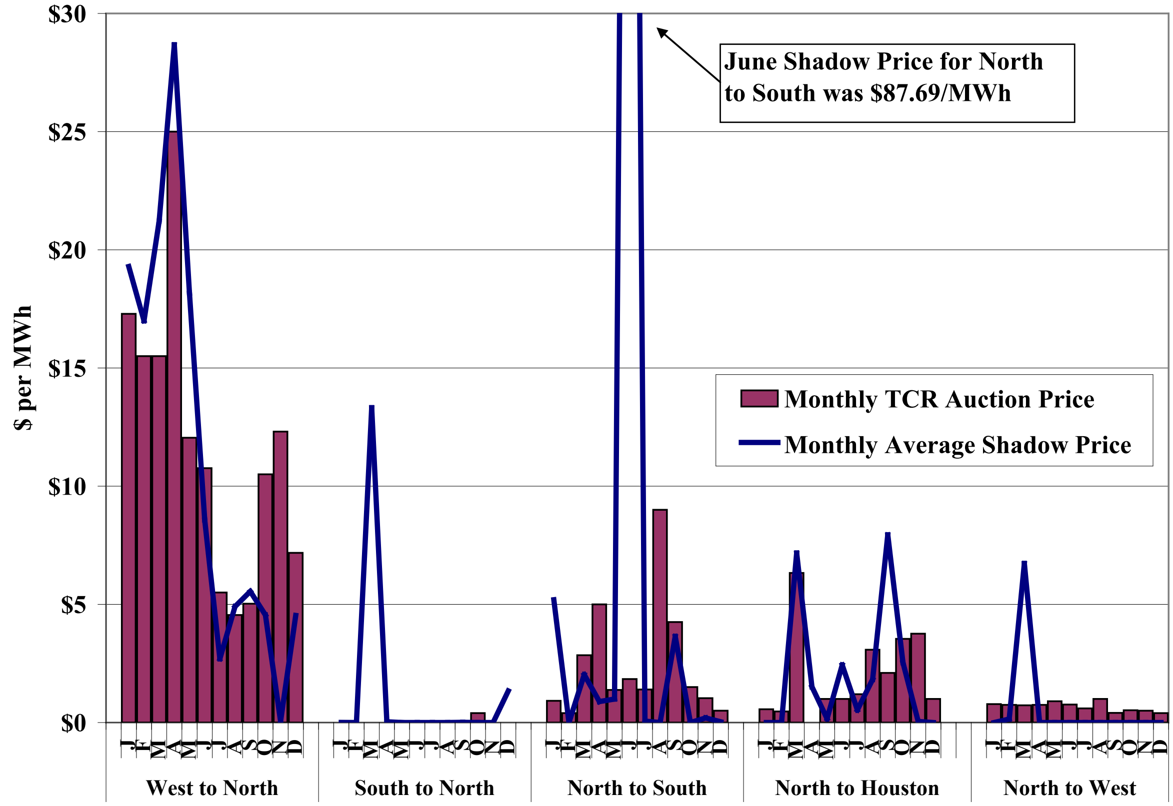
Figure 60: TCR Auction Prices versus Balancing Market Congestion Prices



This figure shows that the TCR annual auction prices were higher than the value of congestion in real-time for the West to North, North to Houston and North to West CSCs in 2009. In contrast, the annual auction price was significantly lower than the value of congestion in real-time for the North to South CSC in 2009. For the West to North, North to Houston and North West CSCs, the average monthly auction prices were more consistent with the value of congestion in real-time in 2009, indicating a more accurate forecast by the participants at the monthly auction than previous years for these CSCs. The North to South monthly auction price was significantly lower than the actual value of congestion in the real-time in 2009. This outcome is primarily due to the significant North to South congestion experienced in June 2009 that was influenced to a large degree by a number of baseload unit outages that were not foreseeable, as discussed previously in this Section.

Figure 61 compares monthly TCR auction prices with monthly average real-time CSC shadow prices from SPD for 2009. The TCR auction prices are expressed in dollars per MWh.

Figure 61: Monthly TCR Auction Price and Average Congestion Value



With the exception of the North to South CSC in June 2009 that diverged for the reasons previously discussed, the monthly TCR auction prices and the real-time shadow prices indicates that market participants improved their ability to predict and value the real-time cost of zonal congestion in 2009 compared to prior years.

To evaluate the total revenue implications of the issues described above, our next analysis compares the TCR auction revenues and obligations. Auction revenues are paid to loads on a load-ratio share basis. Market participants acquire TCRs in the ERCOT-run TCR auction market in exchange for the right to receive TCR credit payments (equal to the congestion price for a CSC times the amount of the TCR). If TCR holders could perfectly forecast shadow prices in the balancing energy market, auction revenues would equal credit payments to TCR holders. The credit payments to the TCR holders should be funded primarily from congestion rent collected in the real-time market from participants scheduling transfers between zones or power flows resulting from the balancing energy market.

The congestion rent from the balancing energy market is associated with the schedules and balancing deployments that result in interzonal transfers during constrained intervals (when there are price differences between the zones). For instance, suppose the balancing energy market deployments result in exports of 600 MWh from the West Zone to the North Zone when the price in the West Zone is $40 per MWh and the price in the North Zone is $55 per MWh. The customers in the North Zone will pay $33,000 (600 MWh * $55 per MWh) while suppliers in the West Zone will receive $24,000 (600 MWh * $40 per MWh). The net result is that ERCOT collects $9,000 in congestion rent ($33,000 – $24,000) and uses it to fund payments to holders of TCRs.[27] If the quantity of TCRs perfectly matches the capability of the CSC in the balancing energy market, the congestion rent will perfectly equal the amount paid to the holders of TCRs.

Figure 62 reviews the results of these processes by showing (a) monthly and annual revenues from the TCR auctions, (b) credit payments earned by the holders of TCRs based on real-time outcomes, and (c) congestion rent from schedules and deployments in the balancing energy market.

[27] This explanation is simplified for the purposes of illustration. Congestion rents are also affected by differences between calculated flows on CSCs from interzonal schedules using zonal average shift factors and actual flows on CSCs in real-time. As discussed in this Section, these differences can be significant.

Figure 62: TCR Auction Revenues, Credit Payments, and Congestion Rent

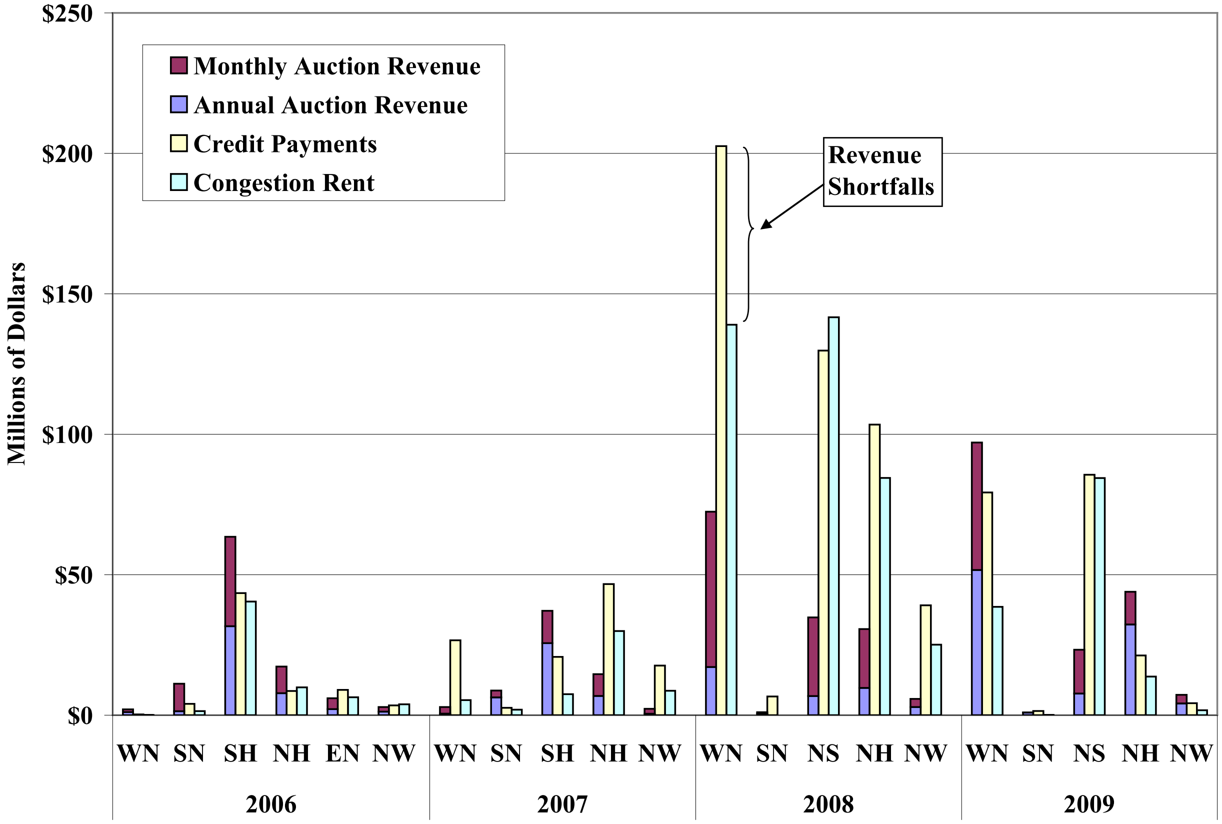


Figure 62 shows that the West to North and North to Houston had the most significant revenue shortfalls in 2009. When congestion rents fall significantly below payments to TCR holders, it implies that the SPD-calculated flows across constrained interfaces have been systematically lower than the amount of TCRs sold for the interfaces.

Figure 62 also shows that payments to TCR holders have consistently exceeded the congestion rents that have been collected from the balancing market in 2006 through 2009. Congestion rents covered 90, 47 and 79 percent of payments to TCR holders in 2006, 2007 and 2008, respectively. In 2009, Congestion rents covered 72 percent of the payments to TCR holders, with an annual net revenue shortfall of $53 million.

As described above, a revenue shortfall exists when the credit payments to congestion rights holders exceed the congestion rent. This shortfall is caused when the quantity of congestion rights exceeds the SPD-calculated flow limits in real-time. These shortfalls are included in the Balancing Energy Neutrality Adjustment charge and assessed to load ERCOT-wide. Collecting substantial portions of the congestion costs for the market through such uplift charges reduces the transparency and efficiency of the market. It also increases the risks of transacting and serving load in ERCOT because uplift costs cannot be hedged.

D. Local Congestion and Local/System Capacity Requirements

In this subsection, we address local congestion and local and system reliability requirements by evaluating how ERCOT manages the dispatch and commitment of generators when constraints and reliability requirements arise that are not recognized or satisfied by the current zonal markets. Local (or intrazonal) congestion occurs in ERCOT when a transmission constraint is binding that is not defined as part of a CSC or CRE. Hence, these constraints are not managed by the zonal market model. ERCOT manages local congestion by requesting that generating units adjust their output quantities (either up or down). When insufficient capacity is committed to meet local or system reliability requirements, ERCOT commits additional resources to provide the necessary capacity in either the day-ahead market or in the adjustment period, which includes the hours after the close of the day-ahead market up to one hour prior to real-time. Capacity required for local reliability constraints is procured through either the Replacement Reserve Service market ("Local RPRS") or as out-of-merit capacity ("OOMC"). Some of this capacity is also instructed to be online through Reliability Must Run ("RMR") contracts. Capacity required for system reliability requirements (*i.e.,* the requirement that the total system-wide online capacity be greater than or equal to the sum of the ERCOT load forecast plus operating reserves in each hour) is procured through either the RPRS market ("Zonal RPRS") or as OOMC.

As discussed above, when a unit's dispatch level is adjusted to resolve local congestion, the unit has provided out-of-merit energy or OOME. For the purposes of this report, we define OOME to include both Local Balancing Energy ("LBE") deployed by SPD and manual OOME deployments, both of which are used to manage local congestion and generally subject to the same settlement rules. Since the output of a unit may be increased or decreased to manage a constraint, the unit may receive an OOME up or an OOME down instruction from ERCOT. For the management of local congestion, a unit that ERCOT commits to meet its reliability requirements is an out-of-merit commitment or OOMC. The payments made to generators by ERCOT when it takes OOME, OOMC, Local RPRS, Zonal RPRS or RMR actions are recovered through uplift charges to the loads. The payments for each class of action are described below.

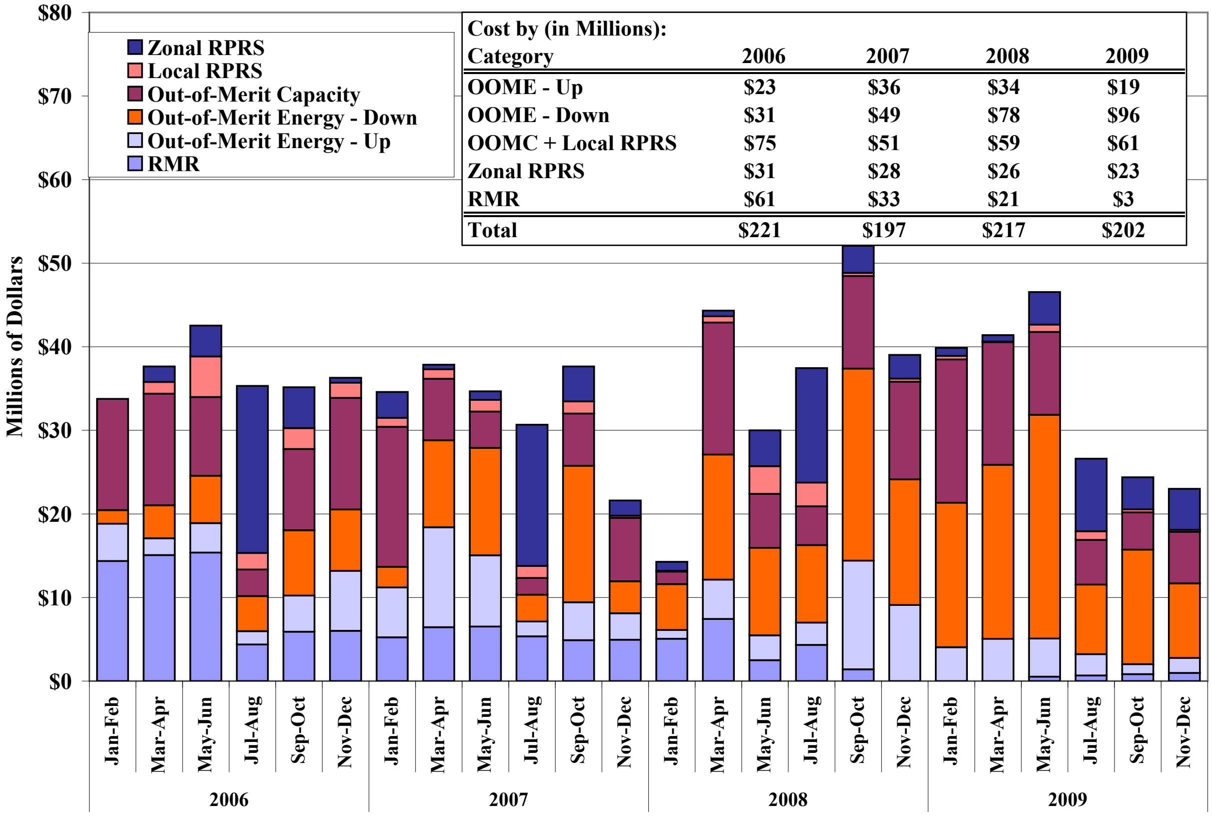
When a unit is dispatched out of merit (OOME up or OOME down), the unit is paid for a quantity equal to the difference between the scheduled output based on the unit's resource plan and the actual output resulting from the OOME instruction from ERCOT. The payment per MWh for OOME is a pre-determined amount specified in the ERCOT Protocols based on the type and size of the unit, the natural gas price, and the balancing energy price. The net payment to a resource receiving an OOME up instruction is equal to the difference between the formula-based OOME up amount and the balancing energy price. For example, for a resource with an OOME up payment amount of $60 per MWh that receives an OOME up instruction when the balancing energy price is $35 per MWh will receive an OOME up payment of $25 per MWh ($60-$35).

For OOME down, the Protocols establish an avoided-cost level based on generation type that determines the OOME down payment obligation to the participant. If a unit with an avoided cost under the Protocols of $15 per MWh receives an OOME down instruction when the balancing energy price is $35 per MWh, then ERCOT will make an OOME down payment of $20 per MWh.

A unit providing capacity under an OOMC or Local RPRS instruction is paid a pre-determined amount, defined in the ERCOT Protocols, based on the type and size of the unit, natural gas prices, the duration of commitment, and whether the unit incurred start-up costs. Owners of a resource receiving an OOMC or Local RPRS instruction from ERCOT are obligated to offer any available energy from the resource into the balancing energy market. Zonal RPRS is selected based upon offer prices for startup and minimum energy and resources procured for Zonal RPRS are paid the market clearing price for this service.

Finally, RMR units committed or dispatched pursuant to their RMR agreements receive cost-based compensation. Since October 2002, ERCOT has entered into several RMR agreements with older, inefficient units that were planned to be retired. As a part of the RMR exit strategy process, all units were removed from RMR status by October 2008; however, two additional units entered into RMR agreements in May 2009. Units contracted to provide RMR service to ERCOT are compensated for start-up costs, energy costs, and are also paid a standby fee. Figure 63 shows each of the four categories of uplift costs from 2006 to 2009.

Figure 63: Expenses for Out-of-Merit Capacity and Energy



Cost by (in Millions):

Category	2006	2007	2008	2009
OOME - Up	$23	$36	$34	$19
OOME - Down	$31	$49	$78	$96
OOMC + Local RPRS	$75	$51	$59	$61
Zonal RPRS	$31	$28	$26	$23
RMR	$61	$33	$21	$3
Total	$221	$197	$217	$202

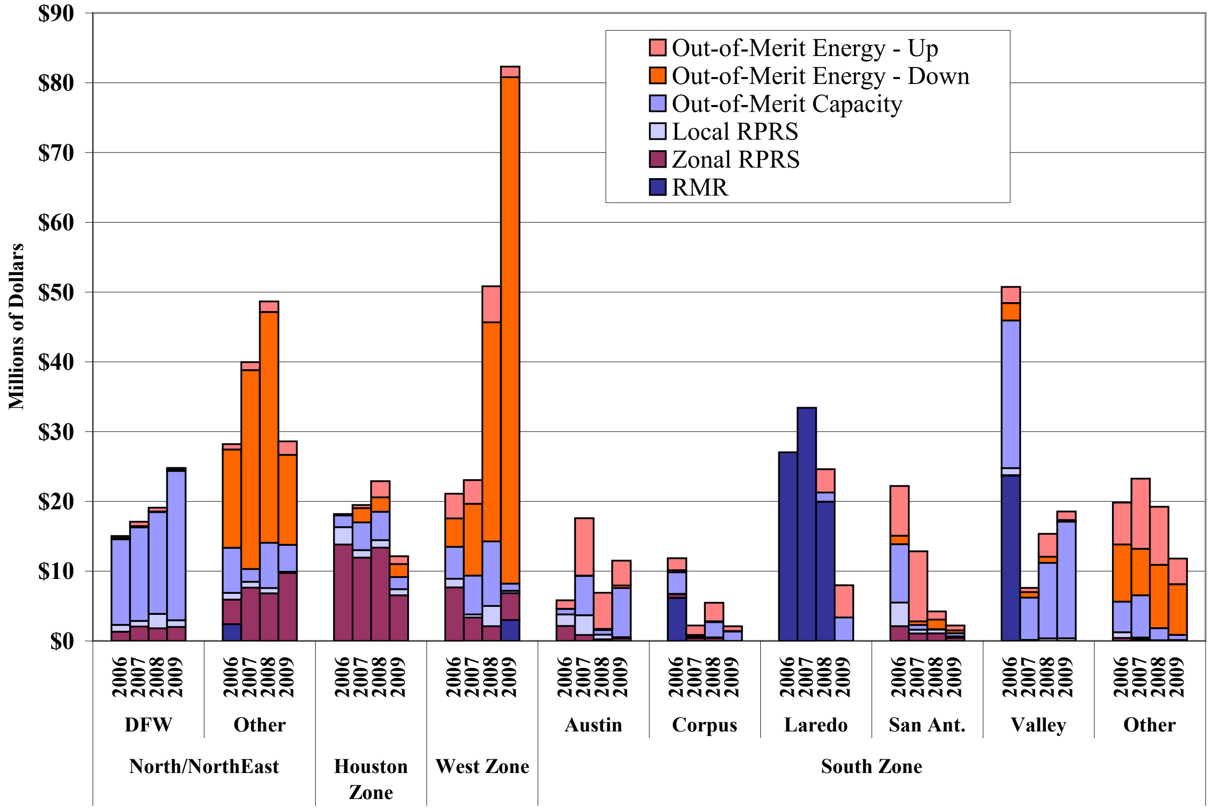
The results in Figure 63 show that overall uplift costs for RMR units, OOME units, OOMC/Local RPRS and Zonal RPRS units were $202 million in 2009, which is a $15 million decrease over the $217 million in 2008. [28] OOME Down and RMR costs accounted for the most significant portion of the change in 2009. OOME down increased from $78 million in 2008 to $96 million in 2009. These values represent significant increases in OOME Down costs from 2006 and 2007, and are primarily attributable to increases in OOME Down instructions for wind resources in the West Zone. RMR cost decreased from $21 million in 2008 to $3 million in 2009. Figure 63 also shows that the highest Zonal RPRS costs occur in July and August when electricity demand in the ERCOT region is at its highest levels.

Although the costs are borne by load throughout ERCOT, the costs are caused in specific locations because these actions, with the exception of zonal RPRS, are taken to maintain local

[28] Zonal RPRS for system adequacy is deployed at the second stage of the RPRS run, which is affected by the deployment at the first stage of the RPRS run, or the local RPRS deployment. Because ERCOT Protocols allocate the costs of local and zonal RPRS in the same manner, we have included both as local congestion costs.

reliability. The rest of the analyses in this section evaluate in more detail where these costs were caused and how they have changed between 2006 and 2009. Figure 64 shows these payments by location.

Figure 64: Expenses for OOME, OOMC and RMR by Region



The most significant changes in local congestion costs in 2009 compared to 2008 shown in Figure 64 are as follows:

- OOME Down costs in the West Zone increased by $42 million in 2009. This increase was associated with the significant addition of wind capacity in the West Zone. OOMC cost in the West Zone decreased by $8 million in 2009.
- OOME Down costs in the North Zone decreased by $20 million in 2009. This decrease can be attributed to fewer transmission outages requiring the reduced output of coal/lignite units.
- RMR costs in the Laredo area of the South Zone decreased by $20 million to zero in 2009. This decrease was associated with the termination of the Laredo RMR contract in October 2008.
- OOMC costs in the Valley area of the South Zone increased by $6 million in 2009. This increase was associated with the more frequent need for local capacity to be online to maintain Rio Grande Valley import limits.

IV. ANALYSIS OF COMPETITIVE PERFORMANCE

In this section, we evaluate competition in the ERCOT market by analyzing the market structure and the conduct of the participants during 2009. We examine market structure by using a pivotal supplier analysis that indicates suppliers were pivotal in the balancing energy market in 2009 much less frequently than in 2007 and 2008 and significantly less frequently than in 2005 and 2006. These results highlight the trend of continued improvement in the structural competitiveness of the balancing energy market over the last five years. This analysis also shows that the frequency with which a supplier was pivotal increased at higher levels of demand, which is consistent with observations in prior years. To evaluate participant conduct we estimate measures of physical and economic withholding. We examine withholding patterns relative to the level of demand and the size of each supplier's portfolio. Based on these analyses, we find the overall performance of the ERCOT wholesale market to be competitive in 2009.

A. Structural Market Power Indicators

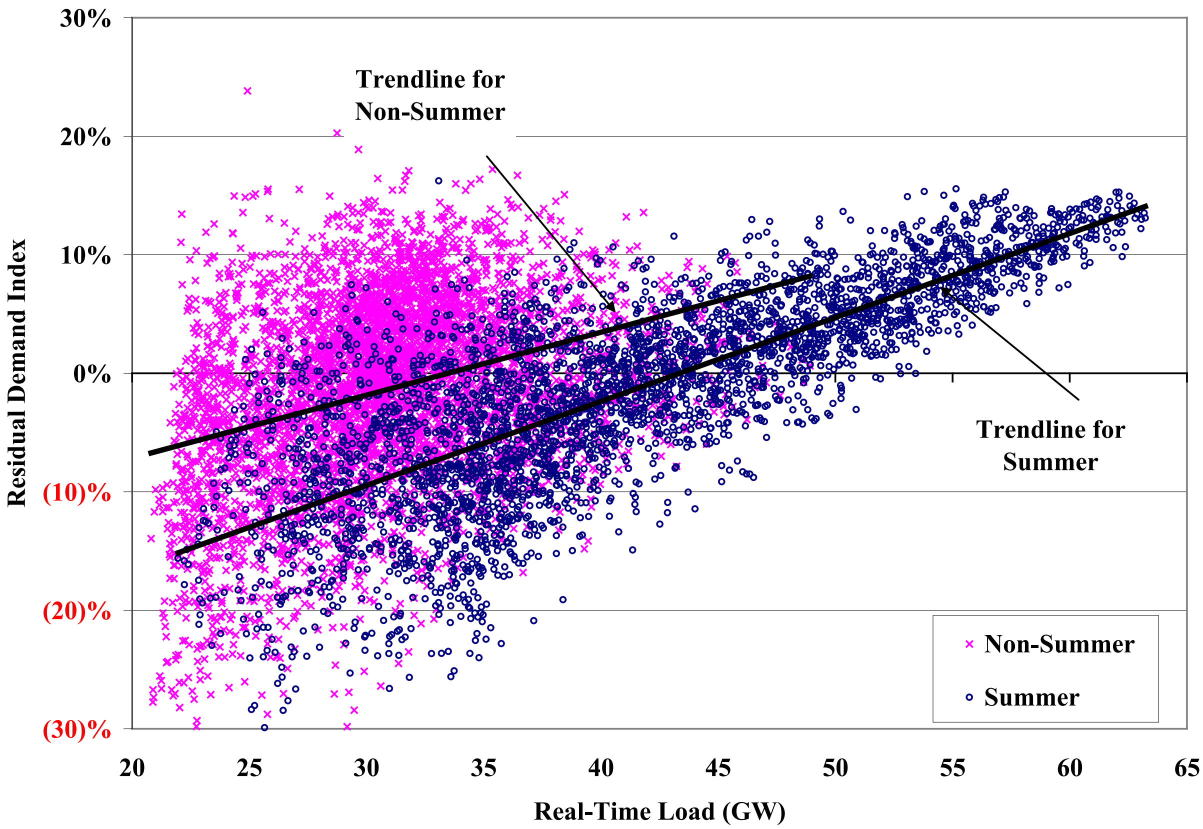
We analyze market structure by using the Residual Demand Index ("RDI"), a statistic that measures the percentage of load that could not be satisfied without the resources of the largest supplier. When the RDI is greater than zero, the largest supplier is pivotal (*i.e.*, its resources are needed to satisfy 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, it does identify conditions under which a supplier would have the *ability* to raise prices significantly by withholding resources.

Figure 65 shows the RDI relative to load for all hours in 2009. The data are divided into two groups: (i) hours during the summer months (from May to September) are shown by darker points, while (ii) hours during other months are shown by lighter points. The trend lines for each data series are also shown and indicate a strong positive relationship between load and the RDI.

This analysis shown below is done at the QSE level because the largest suppliers that determine the RDI values own a large majority of the resources they are scheduling or offering. It is possible that they also control the remaining capacity through bilateral arrangements, although we do not know whether this is the case. To the extent that the resources scheduled by the largest QSEs are not controlled or providing revenue to the QSE, the RDIs will tend to be slightly overstated.

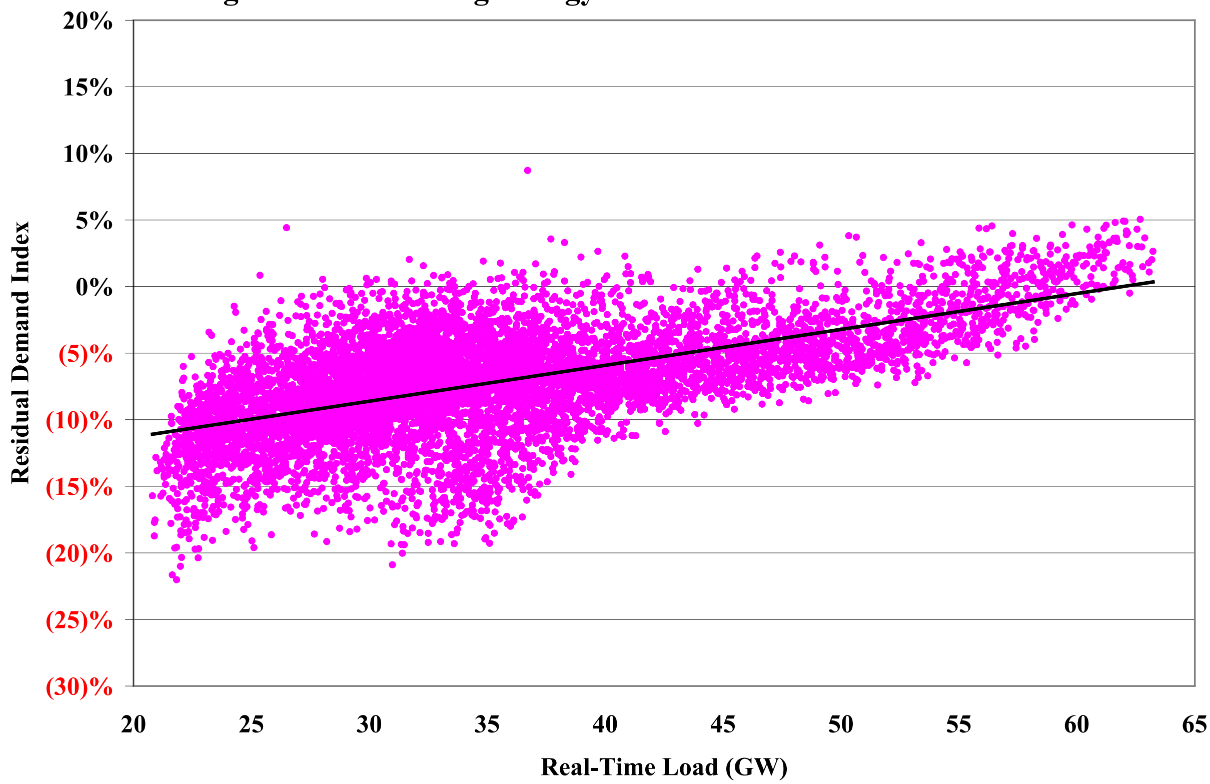
Figure 65: Residual Demand Index



The figure shows that the RDI for the summer (i.e. May to September) was usually positive in hours when load exceeded 40 GW. During the summer, the RDI was greater than zero in approximately 46 percent of all hours, reduced from 60 percent in 2008. The RDI was comparable at lower load levels during the spring and fall due to the large number of generation planned outages and less commitment. Hence, although the load was lower outside the summer, our analysis shows that a QSE was pivotal in approximately 46 percent of all hours during the non-summer period, reduced from 70 percent in 2008. It is important to recognize that inferences regarding market power cannot be made solely from this data. Retail load obligations can affect the extent of market power for large suppliers, since such obligations cause them to be much smaller net sellers into the wholesale market than the analysis above would indicate. Bilateral contract obligations can also affect a supplier's potential market power. For example, a smaller supplier selling energy in the balancing energy market and through short-term bilateral contracts may have a much greater incentive to exercise market power than a larger supplier with substantial long-term sales contracts. The RDI measure shown in the previous figure does not consider the contractual position of the supplier, which can increase a supplier's incentive to exercise market power compared to the load-adjusted capacity assumption made in this analysis.

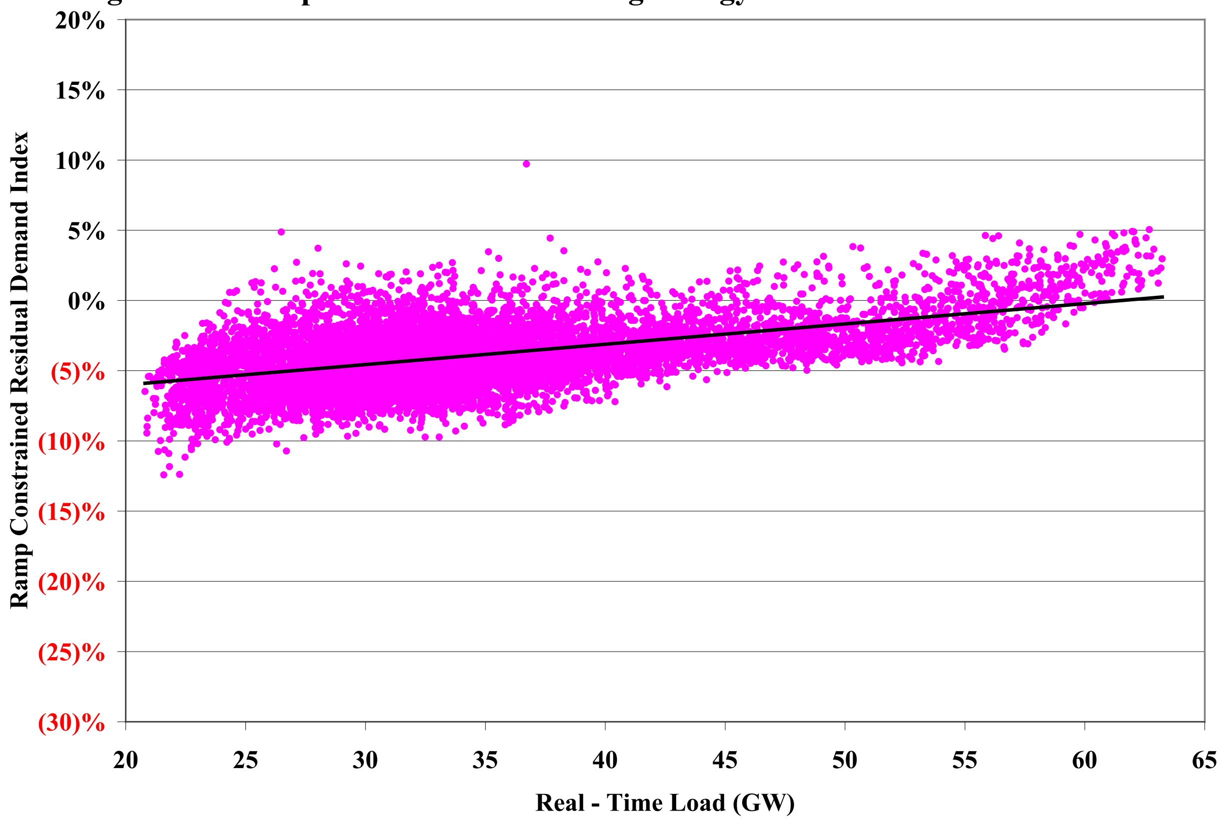
In addition, a supplier's ability to exercise market power in the current ERCOT balancing energy market may be higher than indicated by the standard RDI. Hence, a supplier may be pivotal in the balancing energy market when it would not have been pivotal according to the standard RDI shown above. To account for this, we developed RDI statistics for the balancing energy market. Figure 66 shows the RDI in the balancing energy market relative to the actual load level.

Figure 66: Balancing Energy Market RDI vs. Actual Load



Ordinarily, 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[29] owned by other suppliers. Figure 66 limits the other supplier's capacity to the capacity offered in the balancing energy market. When the RDI is greater than zero, the largest supplier's balancing energy offers are necessary to prevent a shortage of offers in the balancing energy market. Figure 67 shows the same data as in Figure 66 except that the balancing energy offers are further limited by portfolio ramp constraints in each interval.

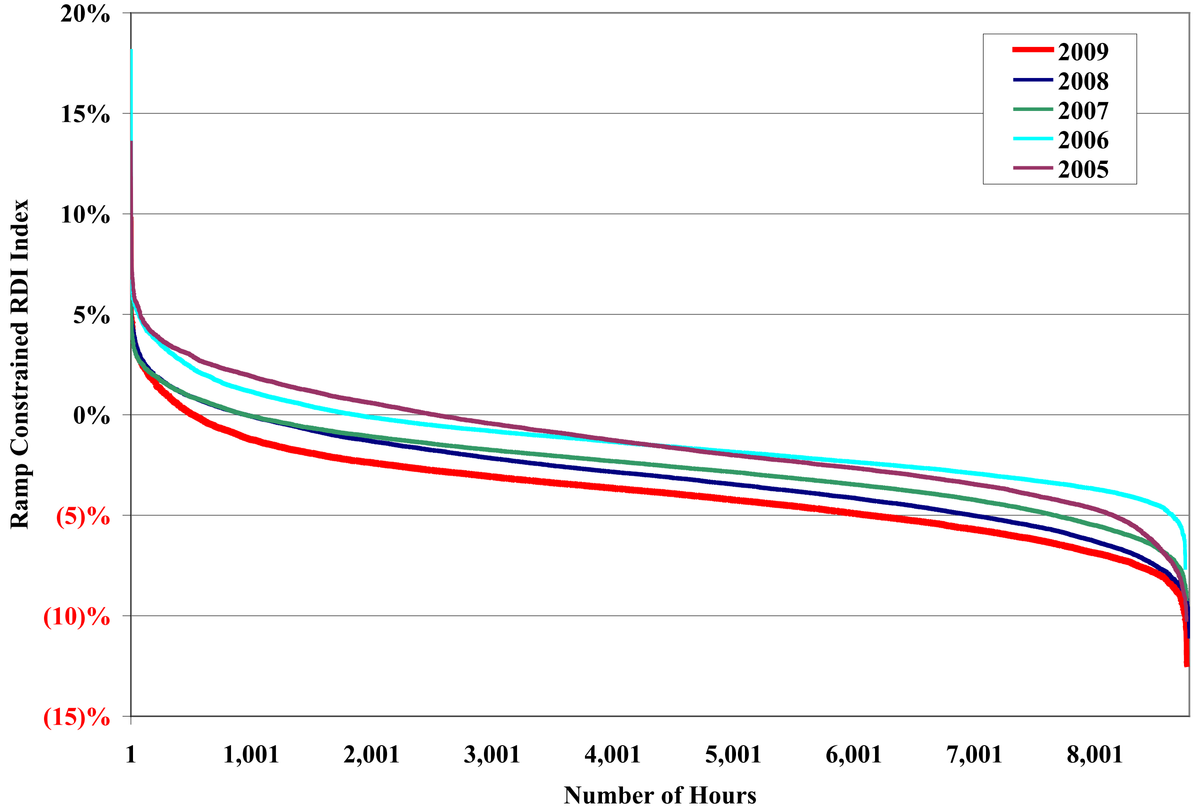
Figure 67: Ramp-Constrained Balancing Energy Market RDI vs. Actual Load



In 2009, the instances when the RDI was positive occurred over a wide range of load levels, from 25 GW to 63 GW. The balancing energy market RDI data and trend line for 2009 are similar in shape to prior years, with the frequency with which a supplier was pivotal generally increasing at higher levels of demand. However, the frequency of data points that are positive in 2009 is smaller than the frequency in prior years. This difference is highlighted in Figure 68, which compares the balancing energy market RDI duration curves for 2005 through 2009.

29 For the purpose of this analysis, "quick-start" includes off-line simple cycle gas turbines that are flagged as on-line in the resource 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 deployment instruction from the balancing energy market.

Figure 68: Ramp-Constrained Balancing Energy Market RDI Duration Curve



The frequency with which at least one supplier was pivotal in the balancing energy market *(i.e.*, an RDI greater than zero) has fallen consistently over the last five years from 29 and 21 percent of the hours in 2005 and 2006, respectively, to less than 11 percent of the hours in 2007 and 2008, to less than 6 percent of the hours in 2009. These results highlight the trend of continued improvement in the structural competitiveness of the balancing energy market over the last five years.

B. Evaluation of Supplier Conduct

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

In a single-price auction like the balancing energy market auction, 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 balancing energy market. Because forward prices will generally be highly correlated with spot prices, price increases in the balancing 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. Balancing Energy Market Offer Patterns

In this section, we evaluate balancing energy offer patterns by analyzing the rate at which capacity is offered.[30] Figure 69 shows the average amount of capacity offered to supply up balancing service relative to all available capacity.

Figure 69 shows a seasonal variation in 2009 over time in quantities of energy available and offered to the balancing energy market. Up balancing offers are divided into the portion that is capable of being deployed in one interval and the portion which would take longer due to portfolio ramp rate offered by the QSE (*i.e.*, "Ramp-Constrained Offers"). Capacity that is available but un-offered is represented by the white dashed portion of each column in the chart.

30 The methodology for determining the quantities of un-offered capacity is detailed in the 2006 SOM Report (2006 SOM Report at 63-65).

Figure 69: Balancing Energy Offers Compared to Total Available Capacity Daily Peak Load Hours

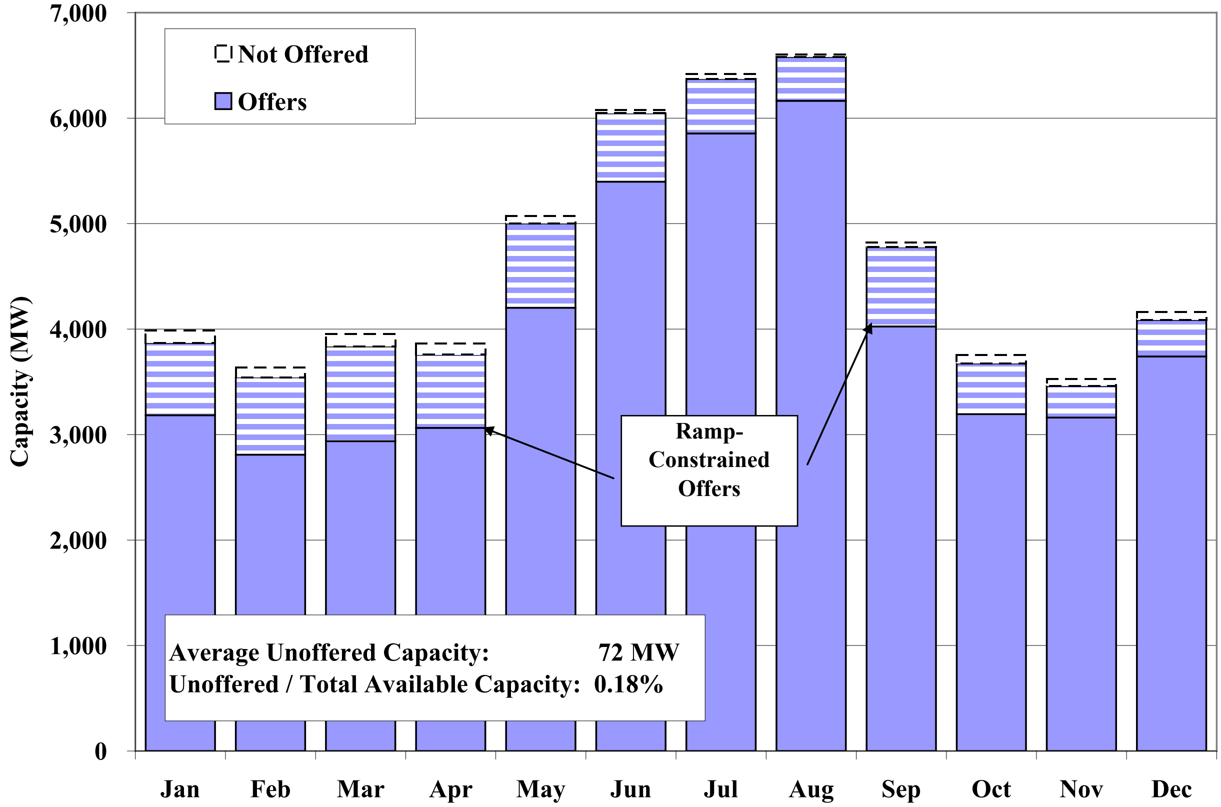


Figure 69 shows a seasonal variation in the quantity of energy available and offered in the balancing energy market, with higher quantities in the summer months than in the non-summer months. This figure also shows that the quantities of un-offered capacity were relatively small in all months in 2009.

Un-offered capacity can raise competitive concerns to the extent that it reflects withholding by a dominant supplier that is attempting to exercise market power. To investigate whether this has occurred, Figure 70 shows the same data as the previous figure, but arranged by load level for daily peak hours in 2009. Because prices are most sensitive to withholding under the tight conditions that occur when load is relatively high, increases in the un-offered capacity at high load levels would raise competitive concerns.

Figure 70: Balancing Energy Offers Compared to Total Available Capacity Daily Peak Load Hours

The figure indicates that in 2009 the average amount of capacity available to the balancing market increased as demand increased. Conversely, the quantity of un-offered capacity decreased as demand increased.

The pattern of un-offered capacity shown in Figure 70 does not raise significant competitive concerns. If the capacity were being strategically withheld from the market, we would expect it to occur under market conditions most susceptible to the exercise of market power. Thus, we would expect significantly more un-offered capacity under higher load conditions. However, the figure shows that portions of the available capacity that are un-offered decreases as load levels increase. Based on this analysis and the additional analyses in this section at the supplier level, we do not find that the un-offered capacity raises potential competitive concerns.

2. Evaluation of Potential Physical Withholding

Physical withholding occurs when a participant makes resources unavailable for dispatch that are otherwise physically capable of providing energy and that are economic at prevailing market prices. This can be done either by derating a unit or declaring it as forced out of service. Because generator deratings and forced outages are unavoidable, the goal of the analysis in this section is to differentiate justifiable deratings and outages from physical withholding. We test for physical withholding by examining deratings and forced 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 65 through Figure 67 indicate that the potential for market power abuse rises at higher load levels as the frequency of positive RDI values increases. Hence, if physical withholding is a problem in ERCOT, we would expect to see increased deratings and forced outages at the highest load levels. Conversely, because competitive prices increase as load increases, deratings and forced outages in a market performing competitively will tend to decrease as load approaches peak levels. Suppliers that lack market power will take actions to maximize the availability of their resources since their output is generally most profitable in these peak periods.

Figure 71 shows the average relationship of short-term deratings and forced outages as a percentage of total installed capacity to real-time load level during the summer months 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.

We focus on the summer months to eliminate the effects of planned outages and other discretionary deratings that customarily occur in off-peak periods. Long-term deratings are not included in this analysis because they are unlikely to constitute physical withholding given the cost of such withholding. Renewable and cogeneration resources are also excluded from this analysis because of the high variation in the availability of these classes of resources. The large supplier category includes the four largest suppliers in ERCOT. The small supplier category includes the remaining suppliers (as long as the supplier controls at least 300 MW of capacity).

Figure 71: Short-Term Deratings by Load Level and Participant Size June to August, 2009

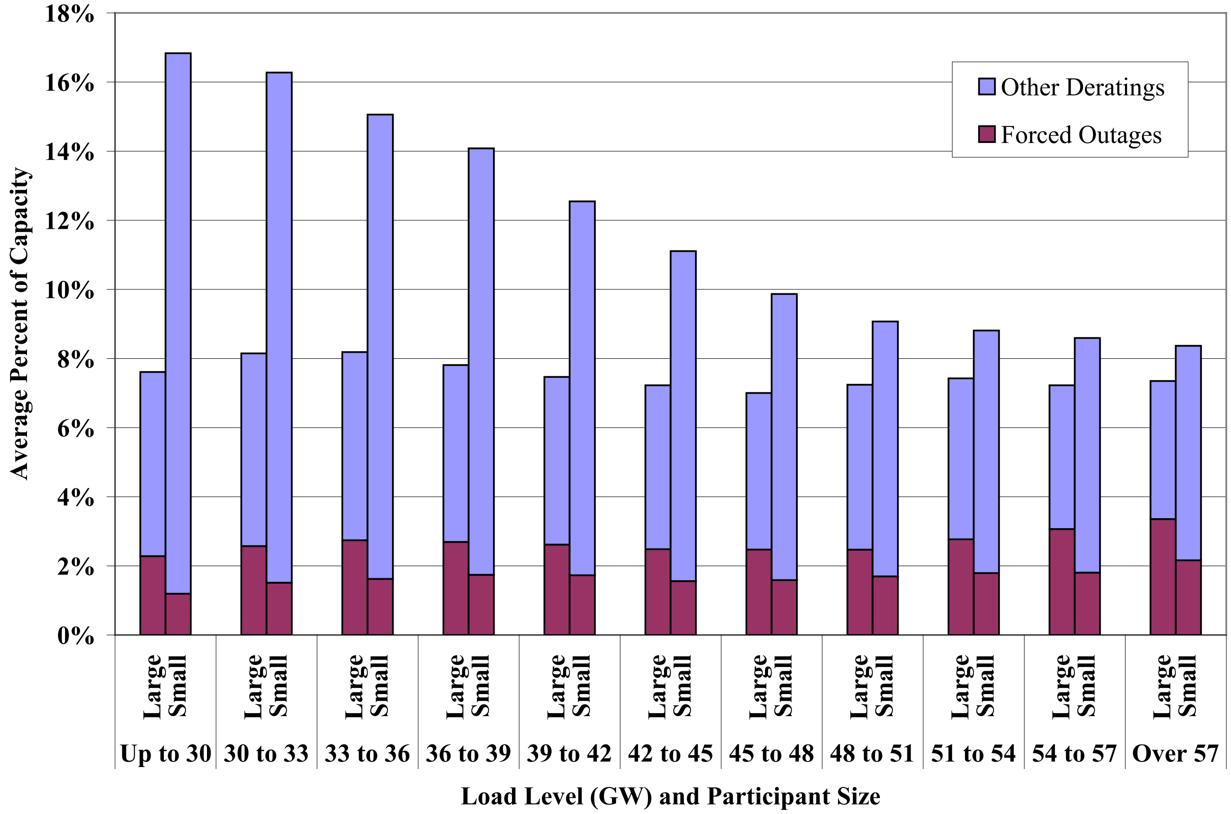


Figure 71 suggests that as electricity demand increases, small market participants tend to make more capacity available to the market, whereas the capacity available from large suppliers is relatively constant across all levels of system demand. For small suppliers, the combined short-term derating and forced outage rates decreased from approximately 17 percent at low demand levels to about 8 percent at load levels above 57 GW. Large suppliers have derating and outage rates that are lower than those of small suppliers across the entire range of load levels. For large suppliers, the combined short-term derating and forced outage rates remained constant, between 7 and 8 percent, across all load levels.

Given that the market is more vulnerable to market power at the highest load levels, these derating patterns do not indicate physical withholding by the large suppliers.

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 balancing energy price. A participant can economically withhold resources, as measured by the output gap, by raising its balancing energy offers so as not to be dispatched or by not offering unscheduled energy in the balancing energy market.

Resources can be included in the output gap when they are committed and producing at less than full output or when they are uncommitted and producing no energy. Unscheduled energy from committed resources is included in the output gap if the balancing energy price exceeds the estimated marginal production cost of energy from that resource by at least $50 per MWh. The output gap excludes capacity that is necessary for the QSE to fulfill its ancillary services obligations. Uncommitted capacity is considered to be in the output gap if the unit would have been profitable given day-ahead bilateral zonal market prices as published in *Megawatt Daily*. The resource is counted in the output gap for commitment if its net revenue (market revenues less total cost, which includes startup and operating costs) exceeds the total cost of committing and operating the resource by a margin of at least 25 percent for the standard 16-hour delivery time associated with on-peak bilateral contracts.[31]

As was the case for outages and deratings, the output gap will frequently detect conduct that can be competitively justified. Hence, it is important to evaluate the correlation of the output gap patterns to those factors that increase the potential for market power, including load levels and portfolio size. Figure 72 compares the real-time load to the average incremental output gap for all market participants as a percentage of the real-time system demand from 2005 through 2009.

[31] The operating costs and startup costs used for this analysis are the generic costs for each resource category type as specified in the ERCOT Protocols.

Figure 72: Incremental Output Gap by Load Level

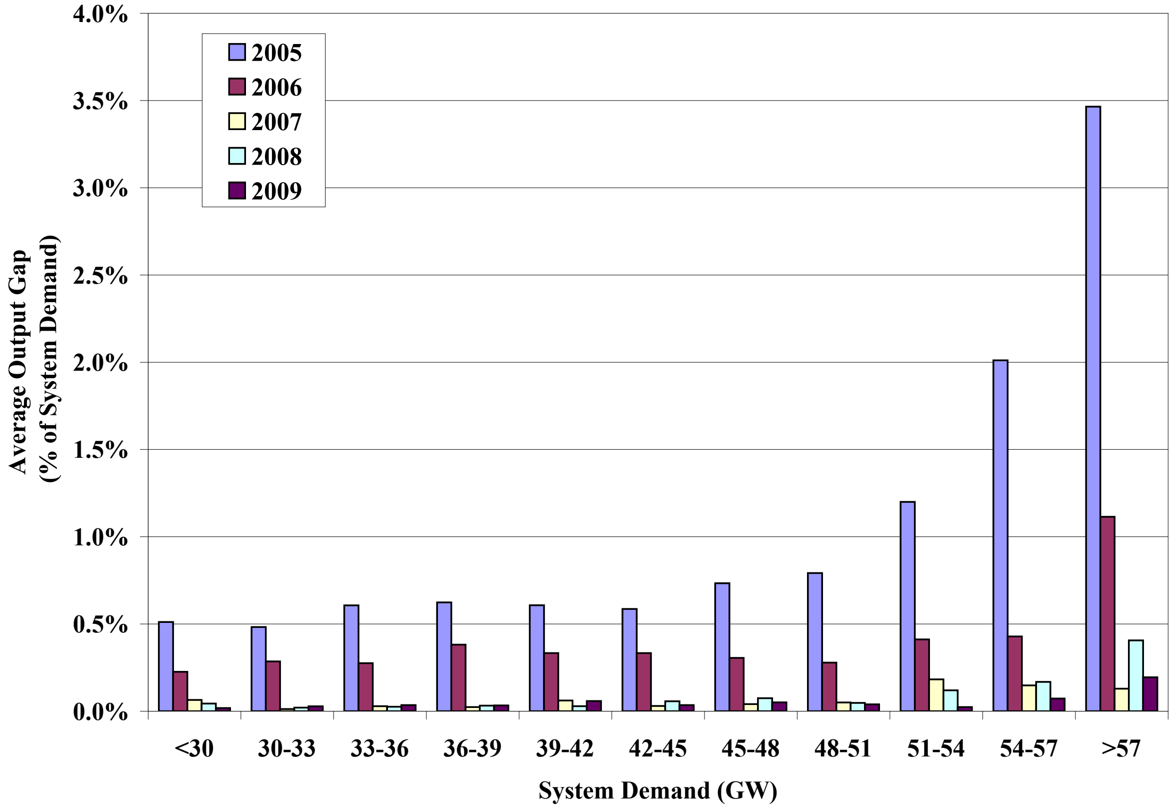


Figure 72 shows that the competitiveness of supplier offers improved considerably in 2006 compared to 2005, followed by even more substantial improvement in 2007 through 2009. In 2009, the overall magnitude of the incremental output gap remains very small and does not raise significant economic withholding concerns.

Figure 73 compares real-time load to the average output gap as a percentage of total installed capacity by participant size. The large supplier category includes the four largest suppliers in ERCOT, whereas the small supplier category includes the remaining suppliers that each controls more than 300 MW of capacity. The output gap is separated into (a) quantities associated with uncommitted resources and (b) quantities associated with incremental output ranges of committed resources.

Figure 73: Output Gap by Load Level and Participant Size

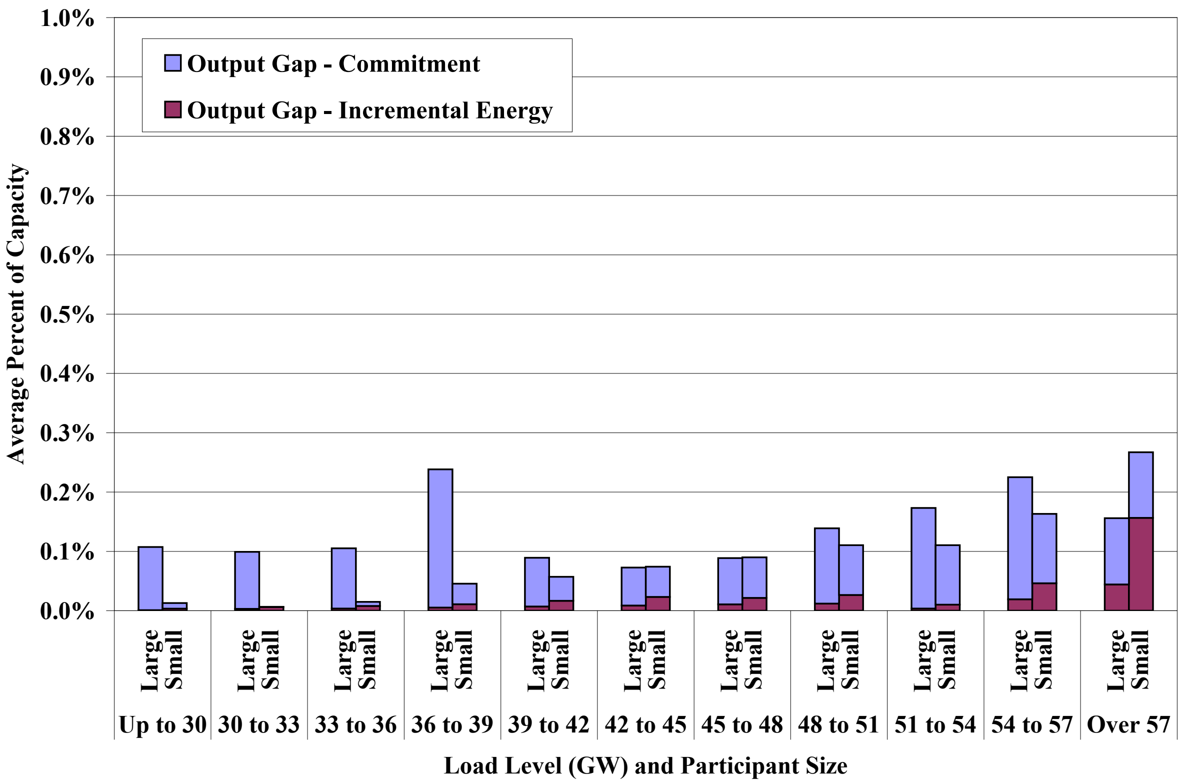


Figure 73 shows that the output gap quantities for incremental energy of large and small suppliers were very low across all load levels. Overall, the output gap measures in 2009 were comparable with the levels in 2008 and 2007, with all the years showing significant improvement over 2005 and 2006.[32] Figure 73 also shows that the increase in the incremental output gap for all market participants in 2009 at the highest load levels is not only small in overall magnitude, but is higher for small participants than for large participants, and therefore does not raise competitive concerns.

Overall, based upon the analyses in this section, we find that the ERCOT wholesale market performed competitively in 2009.

32 *See* 2005, 2006, 2007 and 2008 SOM Reports.