POTOMAC ECONOMICS

2008 STATE OF THE MARKET REPORT
FOR THE
ERCOT WHOLESALE ELECTRICITY MARKETS

POTOMAC ECONOMICS, LTD.

Independent Market Monitor for the
ERCOT Wholesale Market

August 2009

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

This report reviews and evaluates the outcomes of the ERCOT wholesale electricity markets in 2008. 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 Public Utility Commission of Texas ("PUCT") Substantive Rule 25.505(g).

Our analysis indicates that the market performed competitively in 2008. However, the report generally confirms prior findings that the current market rules and procedures are resulting in systemic inefficiencies. Many of these findings can be found in six previous reports we have issued regarding the ERCOT electricity markets.[1] These reports included a number of recommendations designed to improve the performance of the current ERCOT markets. Many of these recommendations were considered by ERCOT working groups and some were embodied in protocol revision requests ("PRRs"). Most of the remaining recommendations will be addressed by the introduction of the nodal market design in late 2010.

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. As discussed in previous reports, this is also one of the most significant shortcomings of the current ERCOT zonal market design. The zonal market structure is an inherently inefficient model for managing transmission congestion. The zonal market model also suffers from the need to predict and define ahead of time those constraints that can be reasonably managed by using zonal congestion management techniques. Given the dynamic nature of supply, demand and the topology of the transmission system, such predictions can often be incorrect. This was the case in 2008, resulting in significant price excursions in the South and

[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 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"); and "ERCOT State of the Market Report 2007", Potomac Economics, August 2008 ("2007 SOM Report").

Houston Zones during the months of April, May and early June until an expedited PRR that modified ERCOT congestion management procedures was implemented.

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 also allow ERCOT to increase the economic and reliable utilization of scarce transmission resources well beyond that attainable in the zonal market. Finally, the nodal market will produce price signals that better indicate where new generation is most needed for managing congestion and maintaining reliability. In the long-term, these enhancements to overall market efficiency should translate into substantial savings for consumers.

A. Review of Market Outcomes

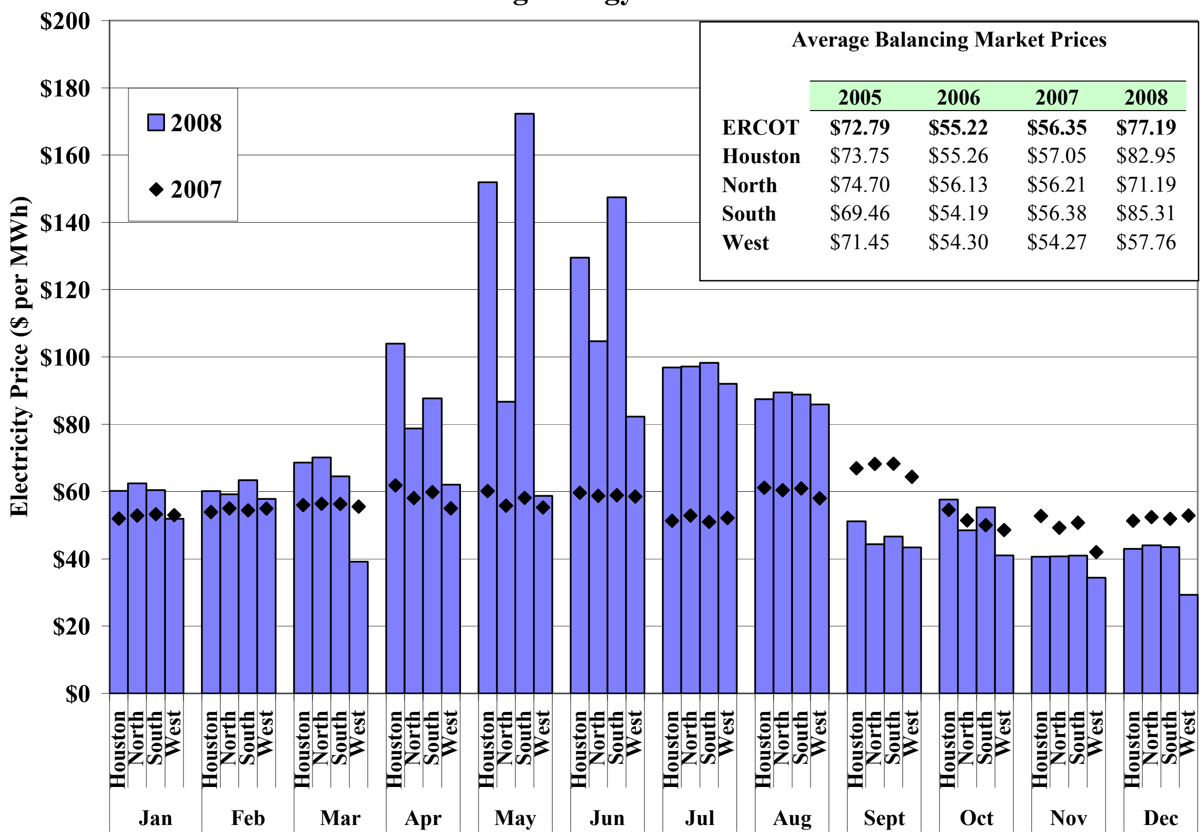
1. Balancing Energy Prices

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, balancing energy market prices were 37 percent higher in 2008 than in 2007, with May and June 2008 showing the largest increases from the same months in 2007. The average natural gas price in 2008 increased 28 percent over 2007 levels, with monthly changes ranging from a 87 percent increase in July ($5.91 per MMBtu in July 2007 and $11.05 per MMBtu in July 2008) to an 20 percent decrease in December ($6.63 per MMBtu in December 2007 and $5.29 per MMBtu in December 2008). Natural gas is typically the marginal fuel in the ERCOT market. Hence, the movements in wholesale energy prices from 2007 to 2008 were largely a function of natural gas price levels.

Balancing Energy Market Prices



	2005	2006	2007	2008
ERCOT	**$72.79**	**$55.22**	**$56.35**	**$77.19**
Houston	$73.75	$55.26	$57.05	$82.95
North	$74.70	$56.13	$56.21	$71.19
South	$69.46	$54.19	$56.38	$85.31
West	$71.45	$54.30	$54.27	$57.76

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. At least five other factors provided a meaningful contribution to price outcomes in 2008.

First, as discussed in Section II, ERCOT peak demand and installed capacity were relatively flat in 2008, and energy production increased only slightly in 2008 compared to 2007. These results were similar to 2007 compared to 2006. In contrast to years prior to 2007 that experienced increasing demand and decreasing supply, the static supply and demand characteristics from 2007 to 2008 contributed to comparable wholesale pricing outcomes over the course of these two years, with the exception of the second factor, which is transmission congestion.

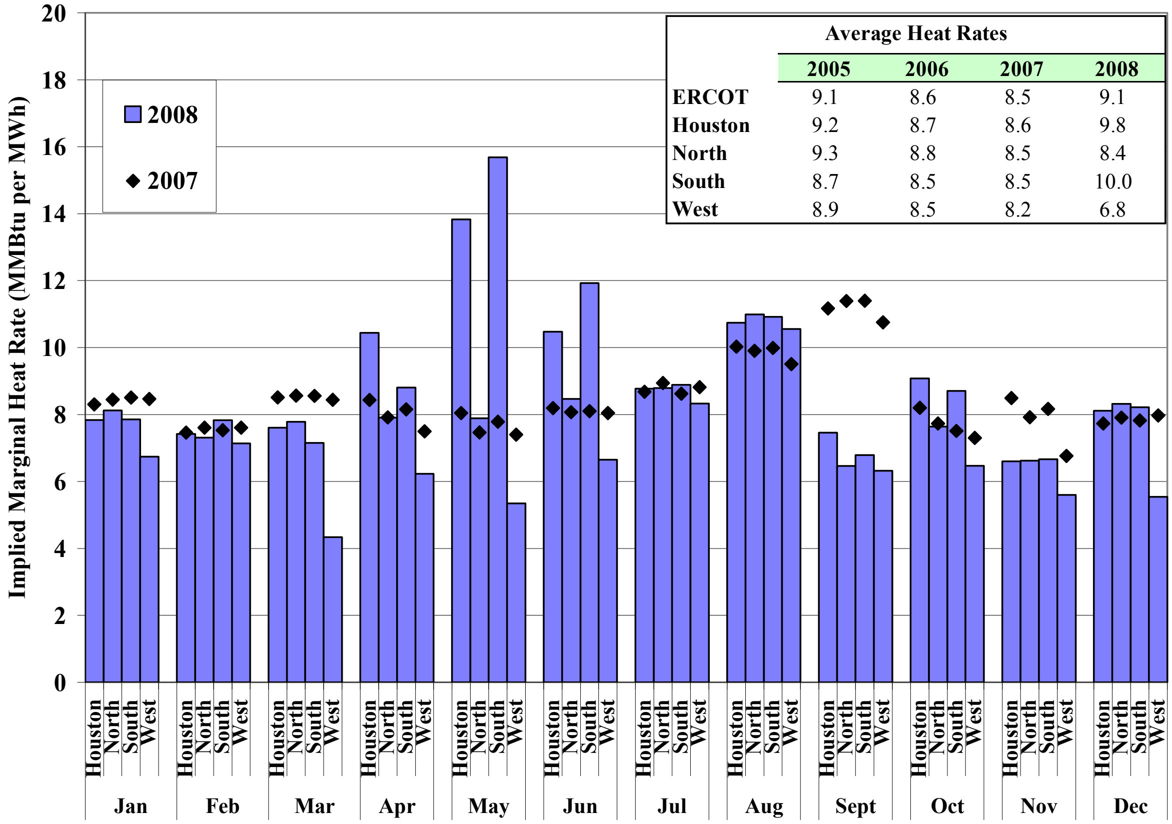
As discussed in Section III, chronic and severe transmission congestion from North to South and North to Houston materialized in April, May and June 2008 that had a significant effect on balancing energy pricing outcomes, particularly in the Houston and South zones. In addition, significant increases in installed wind generation in the West Zone led to an increase in West to North congestion, in turn producing a load-weighted average price in the West Zone that was approximately 26 percent below the ERCOT average price in 2008, with wind resources frequently being the marginal generation source in the West Zone.

Third, aside from the effect of wind generation on the West Zone prices, the continued increase in wind production in 2008 served to displace more costly generation resources when the wind was producing. This will tend to lower average prices across the market, but the intermittent nature of wind can also lead to transitory price spikes as other generation resources may be required on short notice to fill the gap left by significantly lower than expected or rapidly declining wind output.

Fourth, the balancing energy offer cap increased to $2,250 per MWh on March 1, 2008, consistent with Commission rule. Prior to March 1, the rule had set the offer cap at $1,500 per MWh. The increased offer cap is intended to produce higher prices during system shortage conditions as a part of the PUCT's rules that rely upon energy prices exclusively to ensure generation resource adequacy as opposed to the reliance on both capacity and energy prices used in most other domestic organized electricity markets. As discussed in Section II, this mechanism was not always effective in achieving this intended outcome, and some of ERCOT's reliability-based actions can often disrupt the market-based balance of supply and demand, thereby frustrating the long-term success of the energy-only market.

Finally, as discussed in Section IV, the overall competitive performance of the market exhibited continued improvement in 2008, which will tend to lower prices. 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



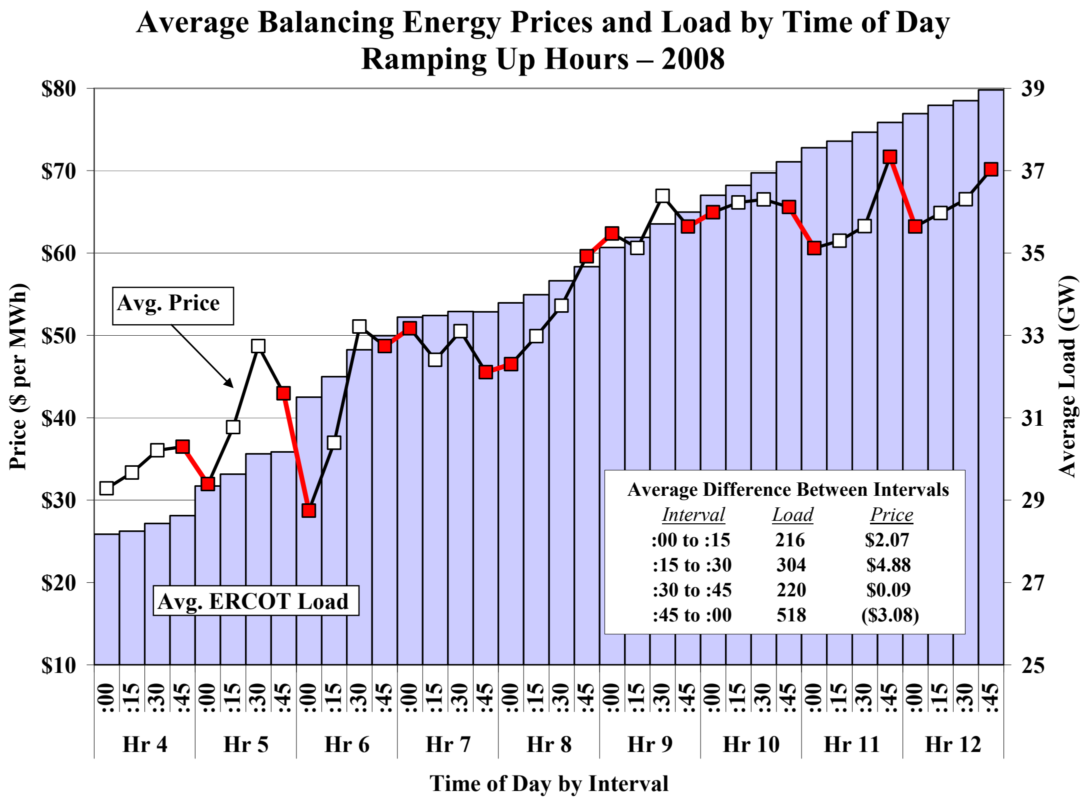
Adjusted for gas price influence, the above figure shows that average implied heat rate for all hours of the year increased by 6.8 percent from 8.50 in 2007 to 9.08 in 2008.[2] The average implied heat rate was higher in 2008 than in 2007 for the months of April, May, June, August and October. The increases in implied heat rates during April through June compared to 2007 are explained primarily by significant transmission congestion that affected the Houston and South Zones most significantly, and is discussed in more detail in Section III. The increase in the implied heat rate in August and October was due to a greater number of shortage intervals in these two months in 2008 compared to 2007, as well as the effect of *ex post* pricing adjustments

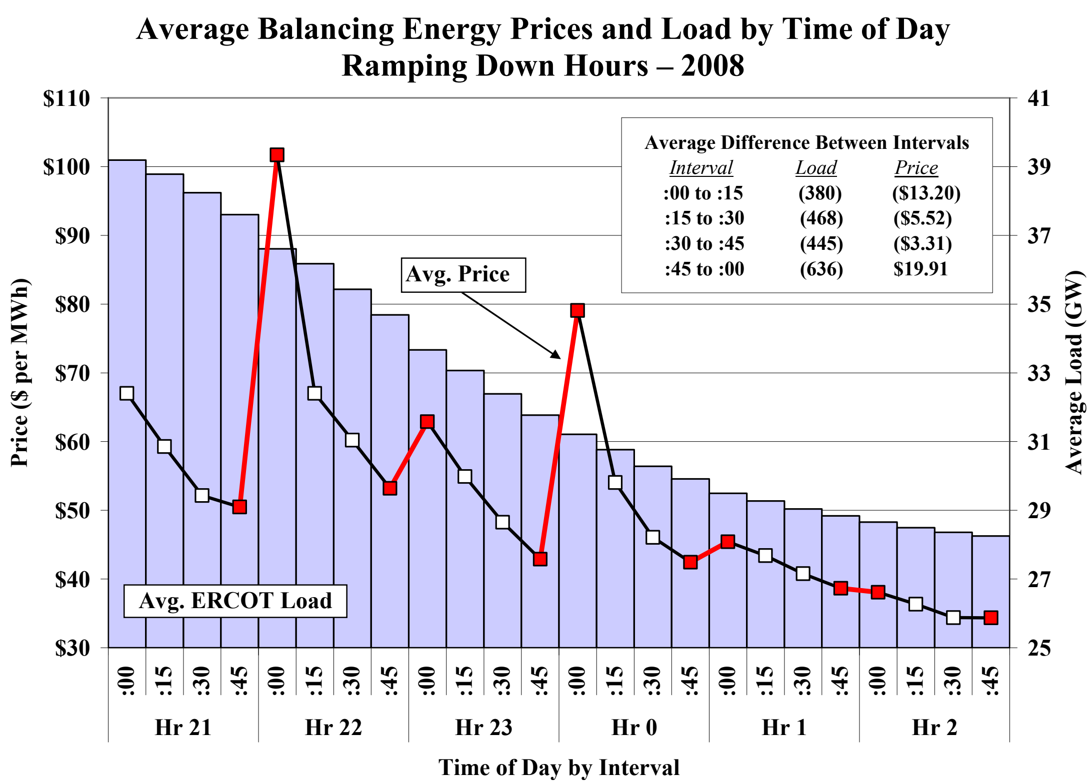
2 The *Implied Marginal Heat Rate* equals the *Balancing Energy Market Price* divided by the *Natural Gas Price.*

during the deployment of non-spinning reserves applied under then-existing ERCOT Protocols, which is discussed in more detail in Section II. In contrast, the implied heat rate in September 2008 was significantly lower than in September 2007. This is explained by two factors. First, September 2007 experienced more shortage intervals than September 2008, which led to an increase in the implied heat rate in September 2007. Second, demand in the ERCOT region was significantly reduced in September 2008 because of the landfall of Hurricane Ike causing widespread and prolonged outages in the Houston area. This suppressed demand and in turn resulted in a significant reduction in the implied heat rate in September 2008.

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 $87 per MWh in 2008 compared to an average of $86 per MWh for real-time prices. Although the day-ahead and real-time prices exhibited good average convergence in 2008, the average absolute price difference, which measures the volatility of the price differences, was large for several months in 2008, particularly in April, May and June. The price volatility in April, May and June 2008 was due in large part to the significant and unpredictable transmission congestion experienced in that timeframe. Relatively smaller spikes in the absolute price difference occurred in August and October 2008. These spikes were associated in part with certain *ex post* pricing revisions during the deployment of non-spinning reserves. As discussed in more detail in Sections II and III, the rules and procedures associated with both of these issues have since been revised. 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.

As discussed in prior reports, we continue to observe in 2008 a clear relationship between the net balancing energy deployments and the balancing energy prices. This is not expected in a well-functioning market. This relationship is partly due to the hourly scheduling patterns of most market participants. Energy schedules change by large amounts at the top of each hour while load increases and decreases smoothly over time. This creates extraordinary demands on the balancing energy market and erratic balancing energy prices, particularly in the morning when loads are increasing rapidly and in the evening when loads are decreasing rapidly.

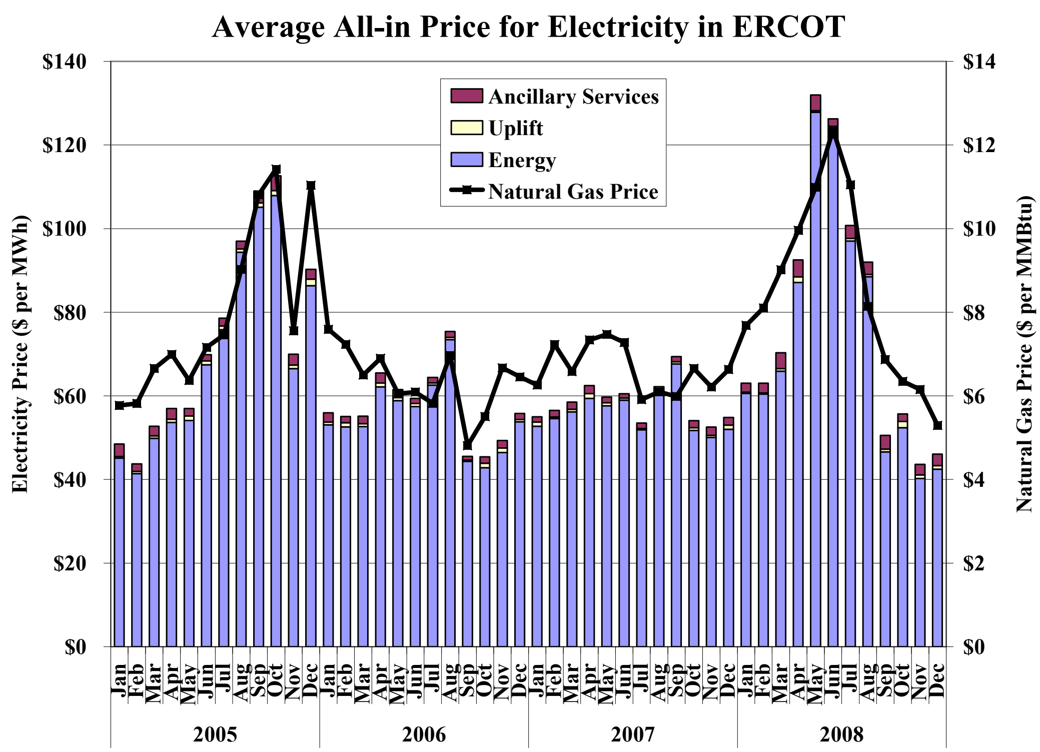




The previous two figures summarize these erratic price patterns by showing the balancing energy prices and actual load in each 15-minute interval during the morning “ramping up” hours and evening "ramping down" hours, with the red lines highlighting the transition from one hour to the next. These pricing patterns raise significant efficiency concerns regarding the operation of the balancing energy market. Moreover, this pattern has been consistently observed for several years and is likely to continue until changes are made to the market rules.[3] In prior reports, we have made several recommendations to address the issue under the current zonal design, although many have not been implemented because of the effort to timely implement the nodal market. The nodal market will provide for a comprehensive solution to the operational issues described in this and prior reports.

2. All-In Electricity Prices

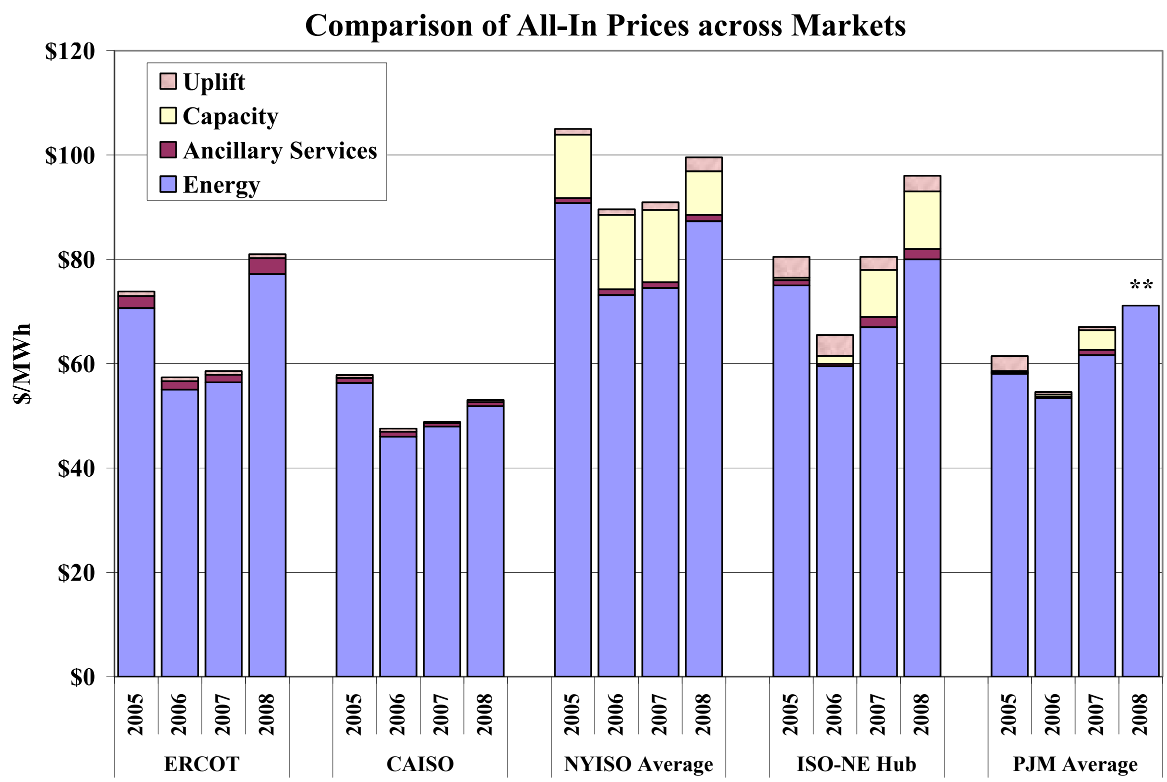
In addition to the costs of energy, loads incur costs associated with operating reserves, regulation, and uplift. The uplift costs include payments for out-of-merit capacity ("OOMC"), Replacement Reserve ("RPRS"), out-of-merit energy ("OOME"), and reliability must run agreements ("RMR"), but exclude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



3 *See* 2003 SOM Report, Assessment of Operations, 2004 SOM Report, 2005 SOM Report 2006 SOM Report and 2007 SOM Report.

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 2005 to 2008. Average natural gas prices increased in 2008 by 28 percent over 2007 levels, although gas prices in 2008 for the higher electricity demand months of May through September were 46% higher than the same months in 2007. The average all-in price for electricity was $58.47 in 2007 and $80.97 in 2008, an increase of 38.5 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.



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

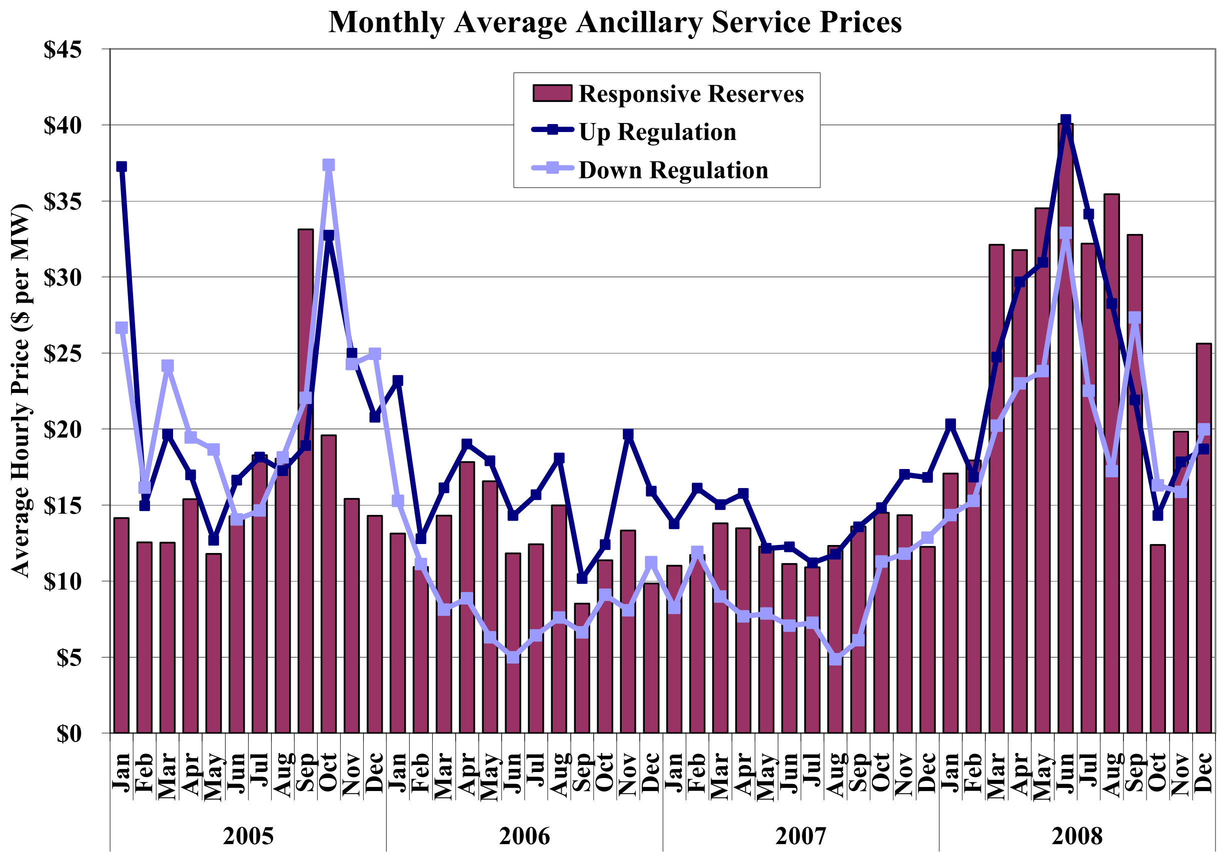
This figure shows that energy prices increased in wholesale electricity markets across the U.S. in 2008, primarily due to increases in fuel costs.

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 2008.

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 2005 to 2008.

This figure shows that after two years of relatively stability, 2008 experienced a significant increase in ancillary service capacity prices. These price movements can be primarily attributed to the variations in energy prices that occurred over the same timeframe.



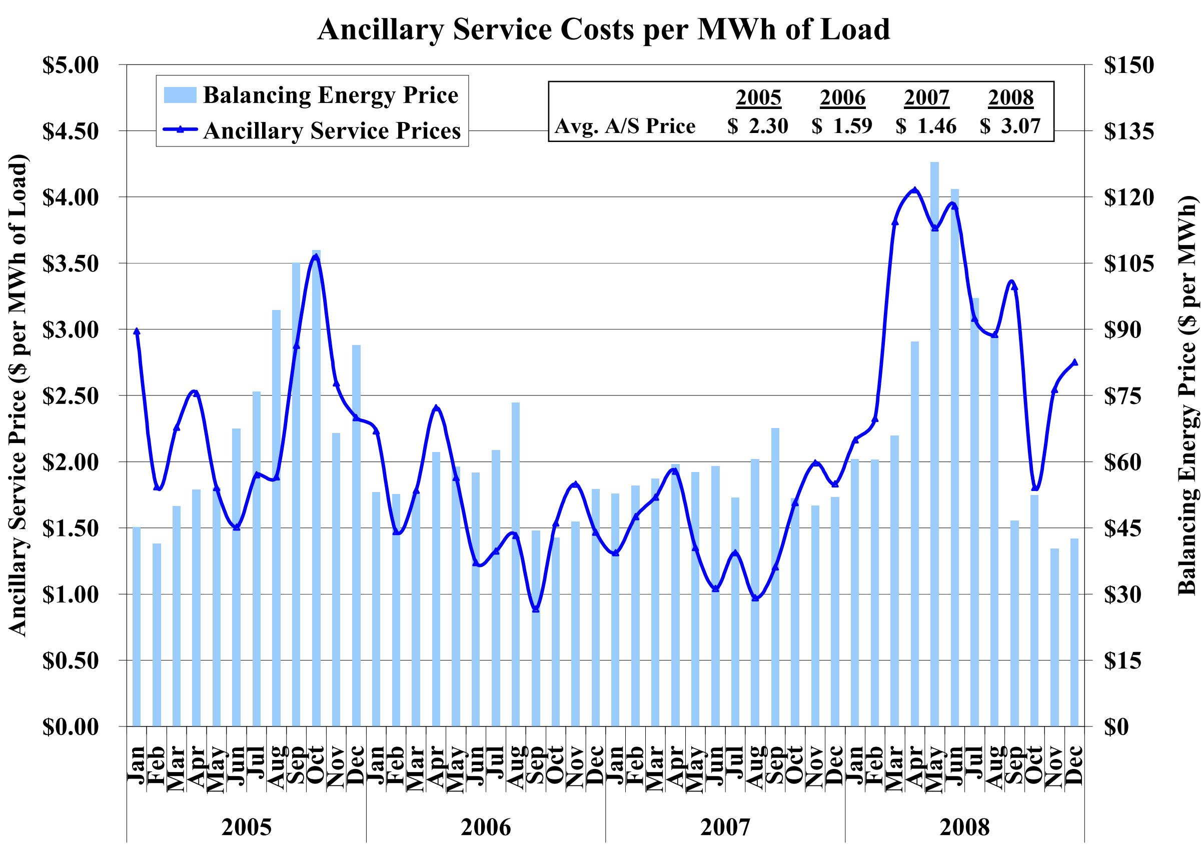
In addition to the effect of higher energy prices on ancillary service prices, the following factors had a significant effect on ancillary service prices in 2008:

- ERCOT increased its procurement of responsive reserve quantities from January through August 2008 from the historical constant quantity of 2,300 MW to as high as 2,800 MW during peak hours in the summer. Also, the required quantity of non-spinning reserves when procured was increased in most of 2008, and non-spinning reserves were procured more frequently in 2008 than in 2007.
- Significant transmission congestion materialized in April, May and June 2008 leading to significantly higher prices in the Houston and South Zones. These pricing outcomes had the effect of increasing the opportunity costs for providers of responsive reserve in these locations, thereby causing an upward shift in the supply curve for responsive reserve in these months.
- West to North congestion increased significantly in 2008, leading to over 1,100 hours of average negative prices in the West Zone. For providers of responsive reserves in the West Zone, exposure to negative prices significantly increases the cost to provide reserves because the resources must operate uneconomically at minimum load levels. Hence, in periods of expected high wind production, responsive reserve offers from suppliers in the West Zone are expected to be higher to reflect these economics risks.

- The quantity of Loads acting as Resources ("LaaRs") providing responsive reserves was moderately reduced in March through May, and experienced more significant reductions in September, part of October, November and December. The reduction in the provision of responsive reserves by LaaRs in these months resulted in a corresponding increase in the quantity of responsive reserve provided by generation resources, which are typically more expensive, thereby placing an upward pressure on responsive reserve prices.

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 it is 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 2005 through 2008.



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 increased to $3.07 per MWh in 2008 compared to $1.46 per MWh in 2007, an increase of more than 110 percent. However, while the all-in wholesale costs increased by more than 38 percent in 2008 compared to 2007, ancillary service costs accounted for only 2.8 percent of the increase in all-in wholesale power costs in 2008 over 2007.

B. Demand and Resource Adequacy

1. Installed Capacity and Peak Demand

Because electricity cannot be stored, the electricity market must ensure that generation matches load on a continuous basis. Thus, one critical issue for a wholesale electricity market is whether sufficient supplies exist to satisfy demand under peak conditions. In 2008, the load served by ERCOT reached a peak of over 62.2 GW, which was almost identical to the peak demand in 2007. Changes in the peak demand levels are very important because they are a key determinant of the probability and frequency of shortage conditions, although daily unit commitment practices, load uncertainty and unexpected resource outages are also contributing factors.

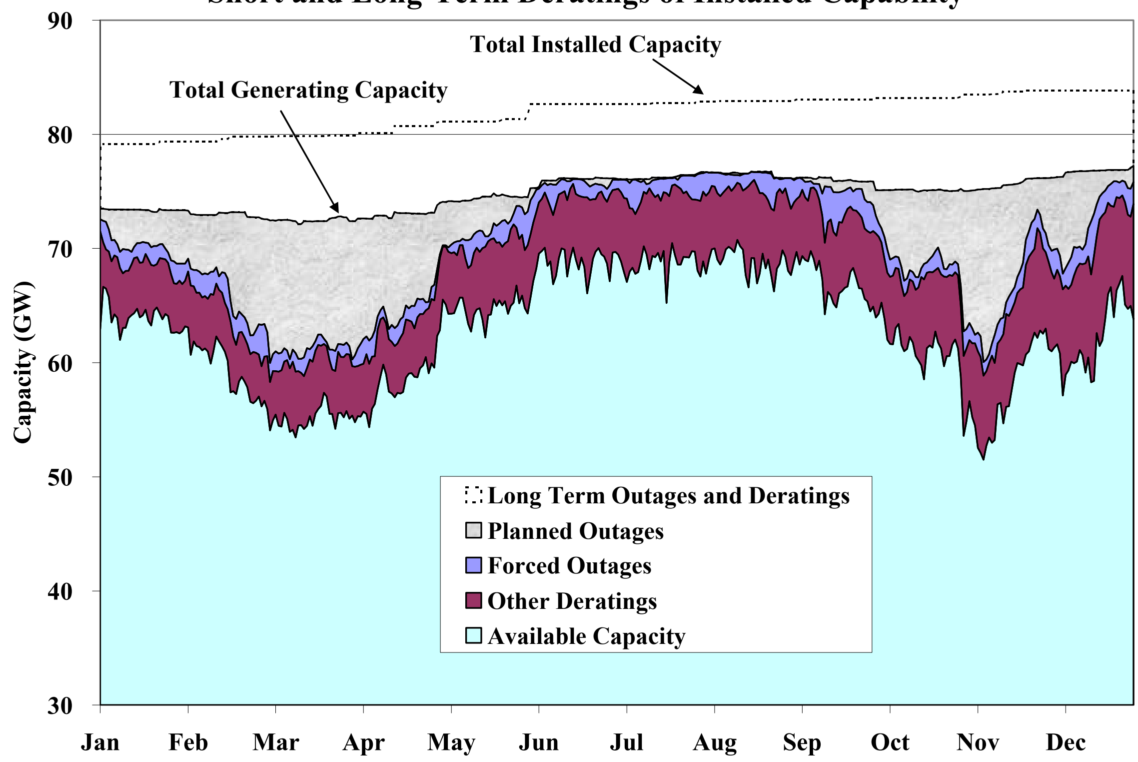
More broadly, peak demand levels and the capability of the transmission network are the primary factors that determine whether the existing generating resources are adequate to maintain reliability. The report provides an accounting of the current ERCOT generating capacity, which is dominated by natural gas-fired resources. These resources account for 70 percent of generation capacity in the ERCOT region while providing only 43 percent of total production in 2008.

ERCOT has more than 80 GW of installed capacity. This includes import capability, resources that can be switched to the SPP, and Loads acting as Resources ("LaaRs"). However, significant amounts of this are not kept constantly in service, with about 5 GW of mothballed capacity existing in 2008. Furthermore, ambient temperature restrictions increase during the summer months when demand is highest, leading to substantial deratings. Although ERCOT had sufficient capacity to meet load and ancillary services needs during the 2008 peak, it is important to consider that electricity demand will continue to grow and that a significant number of generating units in Texas will soon reach or are already exceeding their expected lifetimes. Without significant capacity additions, these factors may cause the resource margins in ERCOT to diminish over the next three to five years. Moreover, although several baseload facilities are currently under construction, the rapidly increasing penetration of intermittent resources such as wind and solar facilities will likely create the reliability need for additional operationally flexible resources, such as modern gas turbines. This reinforces the importance of ensuring that efficient economic signals are provided by the ERCOT market.

2. Generator Outages and Commitments

Despite adequate installed capacity, 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.

Short and Long-Term Deratings of Installed Capability



A derating is the difference between the installed capability of a generating resource and its maximum capability (or "rating") in a given hour. Generators can be fully derated (rating equals 0) due to a forced or planned outage. However, it is very common for a generator 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.*, ambient temperature conditions). The previous figure shows the daily available and derated capability of generation in ERCOT.

The figure shows that long-term outages and other deratings fluctuated between 9 and 18 GW. These outages and deratings reduce the effective resource margins in ERCOT from the levels reported above. With regard to short-term deratings and outages, the patterns of planned outages and forced outages were consistent with expectations in that: (1) forced outages occur randomly over the year and the forced outage rates were relatively low; and (2) planned outages were relatively large in the spring and fall and extremely small during the summer.

In addition to the generation outages and deratings, the report evaluates the results of the generator commitment process in ERCOT, which is decentralized and largely the responsibility of the QSEs. This evaluation includes analysis of the real-time excess capacity in ERCOT. We define excess capacity as the total online capacity plus quick-start units each day minus the daily peak demand for energy, responsive reserves provided by generation, and up regulation. Hence, it measures the total generation available for dispatch in excess of the electricity needs each day.

The report finds that the excess on-line capacity during daily peak hours on weekdays averaged 2,723 MW in 2008, which is approximately 8 percent of the average load in ERCOT. The overall trend in excess on-line capacity also 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 provide a market-wide optimal solution. Also contributing to the suboptimal results of the current unit commitment process is that the decentralized unit commitment is reported to ERCOT through 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. Consequently, ERCOT frequently takes additional actions to ensure reliability that may be more costly and less efficient than necessary. Under the nodal market design, the introduction of a day-ahead energy market with centralized Security Constrained Unit

Commitment ("SCUC") that is financially binding promises substantial efficiency improvements in the commitment of generating resources.

3. Load Participation in the ERCOT Markets

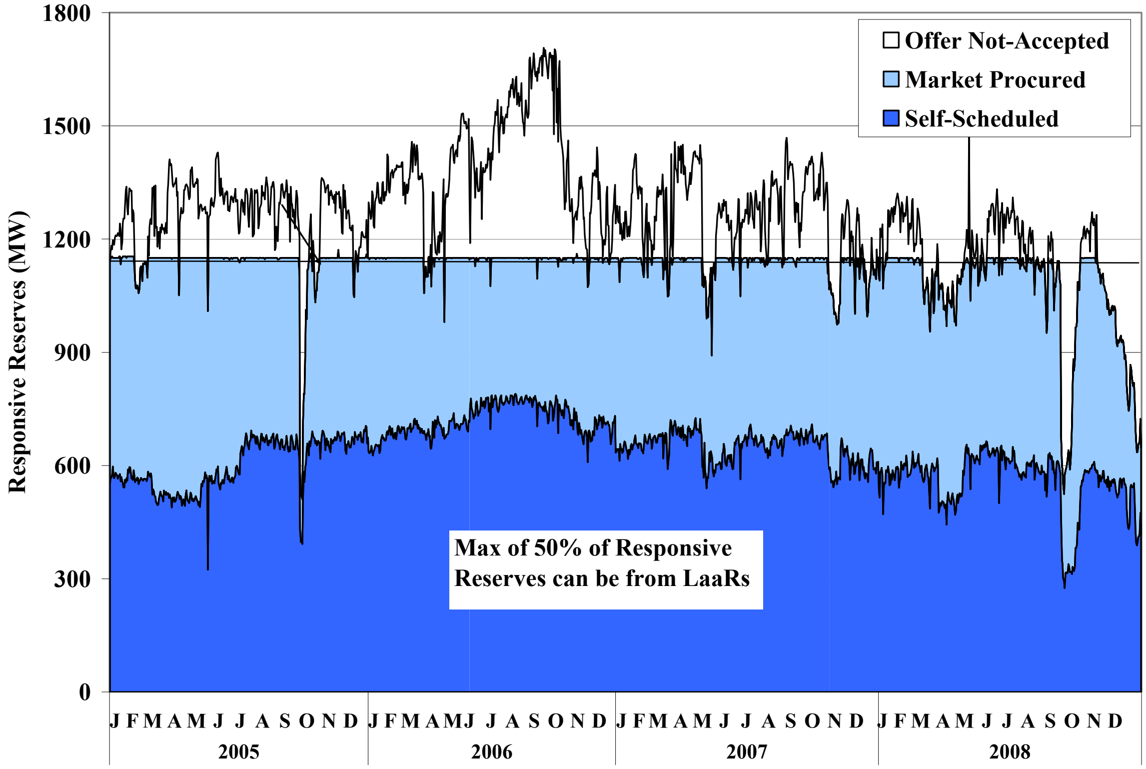
The ERCOT Protocols allow for loads to participate in the ERCOT-administered markets as either Load acting as Resources ("LaaRs") or Balancing Up Loads ("BULs"). LaaRs are loads that are qualified by ERCOT to offer responsive reserves, non-spinning reserves, or regulation into the day-ahead ancillary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

As of December 2008, 2,158 MW of capability were qualified as LaaRs. In 2008, 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 2008.[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 are 70 times more likely to be deployed. 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 2005 through 2008.

---

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.

Provision of Responsive Reserves by LaaRs
Daily Average



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 2005. Notable exceptions were a period in September/October 2005 corresponding to Hurricane Rita, and a more prolonged decrease in September/October of 2008 corresponding to the Texas landfall of Hurricane Ike. Of interest in late 2008 is the post-hurricane recovery of the quantity of LaaRs providing Responsive Reserve followed by a steady reduction for the remainder of the year, which was likely a product of the economic downturn and its effect on industrial operations.

4. Net Revenue Analysis

The next analysis of the outcomes in the ERCOT markets in 2008 is the analysis of "net revenue". 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:

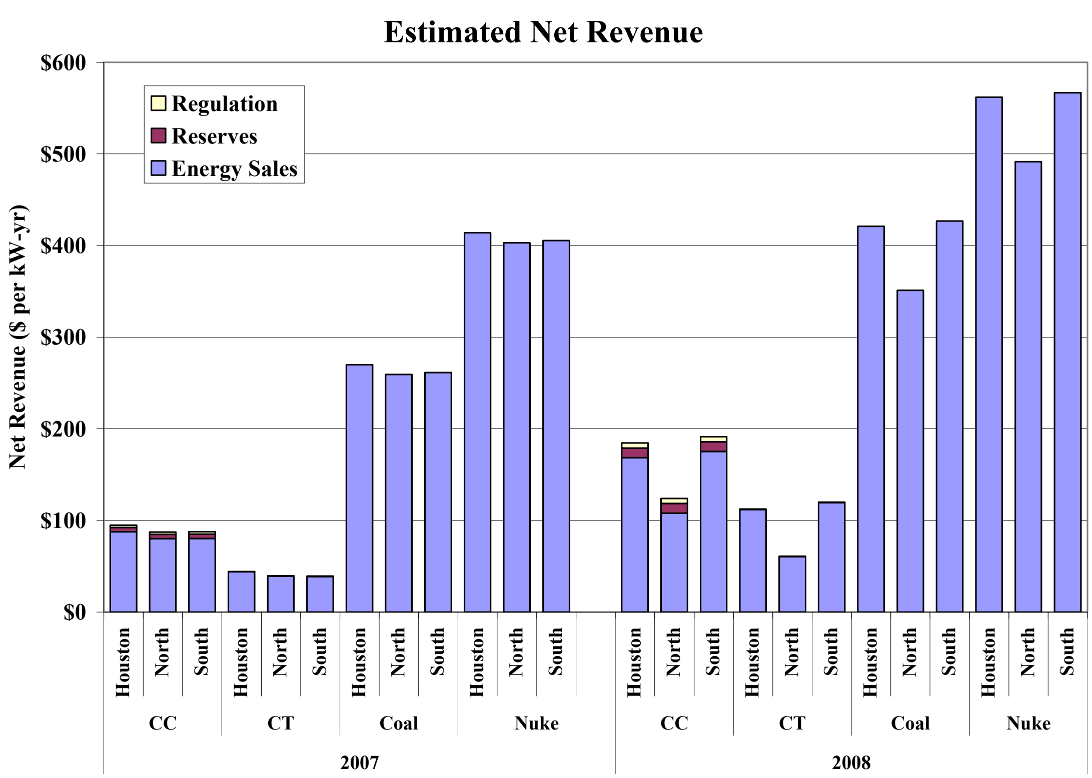
(i) New capacity is not currently needed because there is sufficient generation already available;

(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 2007 and 2008 for four types of units: a natural gas combined-cycle generator, a simple-cycle gas turbine, a coal-fired steam turbine with scrubbers, and a nuclear unit.



The figure above shows that the net revenue increased substantially in 2008 in each zone compared to 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 2008 for a new gas turbine was approximately $120, $113 and $61 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 2008 for a new combined cycle unit was approximately $191, $185 and $124 per kW-year in the South, Houston and North Zones, respectively. These values indicate that the estimated net revenue in 2008 was sufficient to support new entry for a new gas turbine in the South and Houston zones and for a combined cycle unit in the South, Houston and North zones. However, as discussed later in this subsection, significant portions of the net revenue results for gas turbine and combined cycle units in 2008 can be attributed to anomalous market design related inefficiencies rather than fundamentals that would support an investment decision for new gas turbines and combined-cycle units.

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 have 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. For a new coal unit, the estimated net revenue requirement is approximately $190 to $245 per kW-year. The estimated net revenue in 2008 for a new coal unit was approximately $427, $421 and $351 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 2008 for a new nuclear unit was approximately $567, $562 and $492 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 sufficient to support new entry in 2008, as was the case in 2005, 2006 and 2007. Thus, it is not surprising that some market participants are building new baseload facilities and that several others have initiated activities that may lead to the construction of additional baseload facilities in the ERCOT region.

Although estimated net revenue grew considerably in 2008 compared to prior 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. Finally, and most importantly in 2008, net revenues can be inflated when prices clear at high levels due to inefficiencies in the market design. Similar to the case of market power, a market participant may be deterred from investing in new capacity if it believes that prevailing net revenues are largely due to market design inefficiencies that will be corrected.

Such market design inefficiencies were apparent in 2008. As discussed in Section III, the vast majority of price excursions in 2008 – particularly in the South and Houston Zones – were not a function of market fundamentals; rather, the price excursions were driven by inefficient congestion management techniques that have since been corrected and are not expected to materialize in the future, especially upon implementation of the nodal market in 2010. In addition to these transmission congestion issues, in 2008 the ERCOT Protocols provided for *ex post* re-pricing provisions in intervals in which non-spinning reserve prices were deployed that frequently resulted in scarcity-level prices at times when ERCOT's operating reserve levels were not deficient. These rules were recently changed, thereby reducing the probability of scarcity-level prices during non-scarcity conditions going forward. Hence, a significant portion of the net revenue produced in 2008 is not reflective of fundamentals that would support an investment decision for new gas turbines and combined-cycle units.

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.

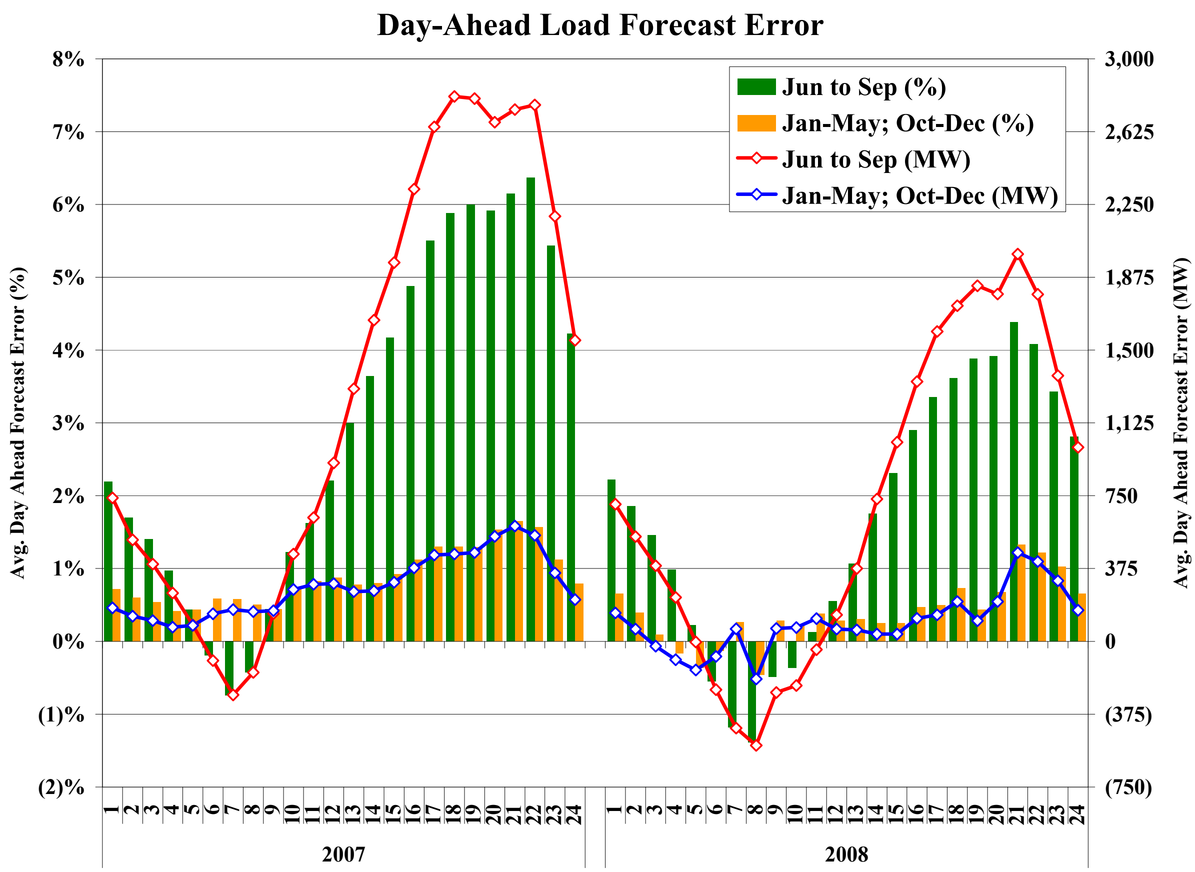
Unlike markets with a long-term capacity market where fixed capacity payments are made to resources across the entire year regardless of the relationship of supply and demand, the objective of the energy-only market design is to allow energy prices to rise significantly higher during legitimate shortage 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.

As noted in the net revenue analysis, 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. Both of these issues have been addressed in the zonal market and will be further improved with the implementation of the nodal market in 2010. Absent these inefficiencies, net revenues would not have been sufficient to support new peaker entry in 2008. Beyond these anomalies, there were three other factors that significantly influenced the effectiveness of the SPM in 2008:

- A substantial decrease in out-of-merit deployments by ERCOT during declared short-supply conditions;
- A continued strong positive bias in ERCOT's day-ahead load forecast that tended to regularly commit online resources in excess of the quantity required to meet expected demand and operating reserve requirements; and
- The dependence on market participants to submit offers at or near the offer cap to produce scarcity level prices during legitimate shortage conditions.

The first item relates to issues that occurred in 2007 and were successfully resolved through changes in ERCOT procedures and the implementation of PRR 750 during 2008. However, the other two items represent ongoing concerns with respect to the successful operation of the ERCOT energy-only market.

The second issue that can adversely affect the successful operation of the ERCOT energy-only market is ERCOT's day-ahead load forecast. The following figure shows the ERCOT day-ahead load forecast error by hour in 2007 and 2008, 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.



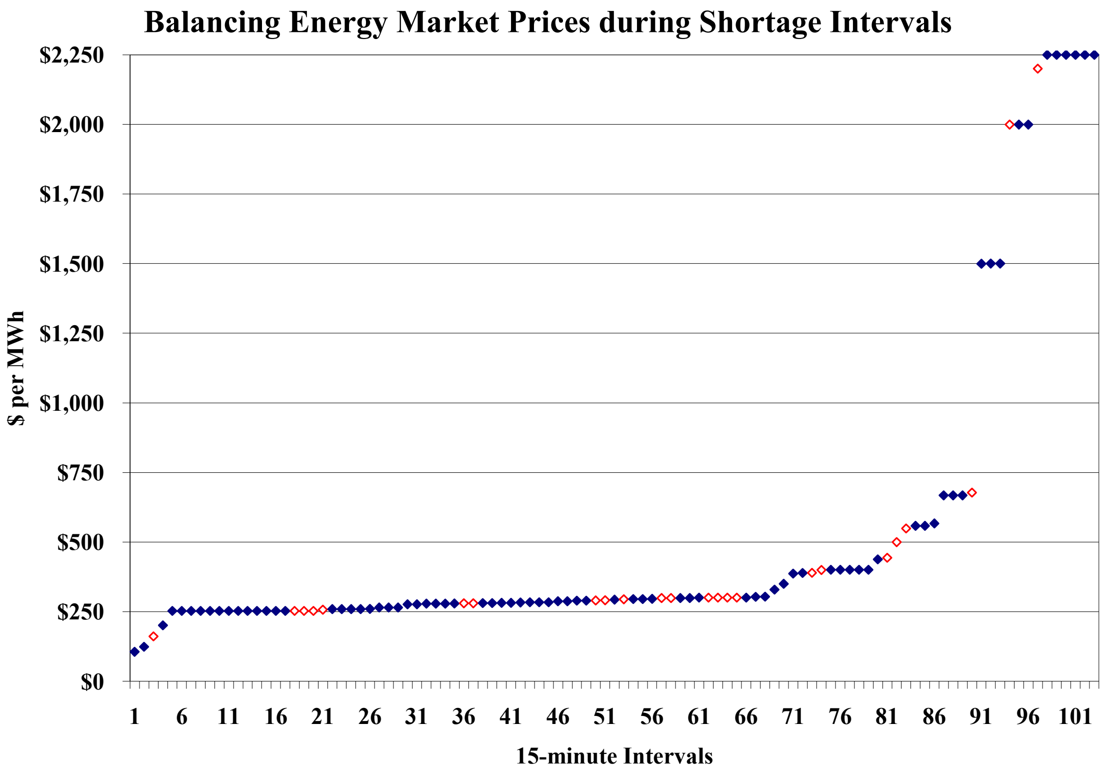
The existence of such a strong and persistent positive bias in the day-ahead load forecast will tend to lead to an inefficient over-commitment of resources and to the depression of real-time prices relative to a more optimal unit commitment. To the extent load uncertainty is driving the bias in the day-ahead 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 cost of the marginal action taken to satisfy the market's demand. In the vast majority of hours, the marginal action is the dispatch of the most expensive online generator. 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 103 15-minute shortage intervals in 2008.



The figure above shows that the prices during these 103 shortage intervals in 2008 ranged from $105 per MWh to the offer cap of $2,250 per MWh (prior to March 1, 2008, the offer cap was $1,500 per MWh), with an average price of $534 per MWh and a median price of $293 per MWh. The results in 2008 are similar to those in 2007 when there were 108 shortage intervals with an average price of $476 per MWh and a median price of $299 per MWh.

In this figure, the data are separated into solid blue and red outlined points. The blue points (79) represent true shortage conditions, whereas the red points (24) represent artificial shortage prices occurring as a result of large generation schedule reductions at the top of the hours from 10 PM to 1 AM. As discussed in more detail in Section I, the production of such artificial shortage prices under these conditions is the result of inefficiencies inherent to the current market design that will be significantly improved with the implementation of the nodal market.

Although each of the blue data points represents system conditions in which the market was in shortage, the pricing outcomes are widely varied, with the majority of prices reflecting the marginal cost 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 and 2008.

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 recent years. In contrast, private investment in mid-merit and peaking resources in ERCOT has been relatively thin. In an energy-only market, net revenue expectations for mid-merit and peaking resources are much more sensitive the effectiveness of the shortage pricing mechanism than to factors such as 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 small market participants to effectively withhold lower cost resources by offering at prices dramatically higher than their marginal cost.

At least for the pendency of the zonal market, shortage pricing will continue to remain dependent upon the existence of high-priced offers by market participants, and results such as those experienced in 2007 and 2008 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 shortage conditions. While important even in markets with a capacity market, 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.

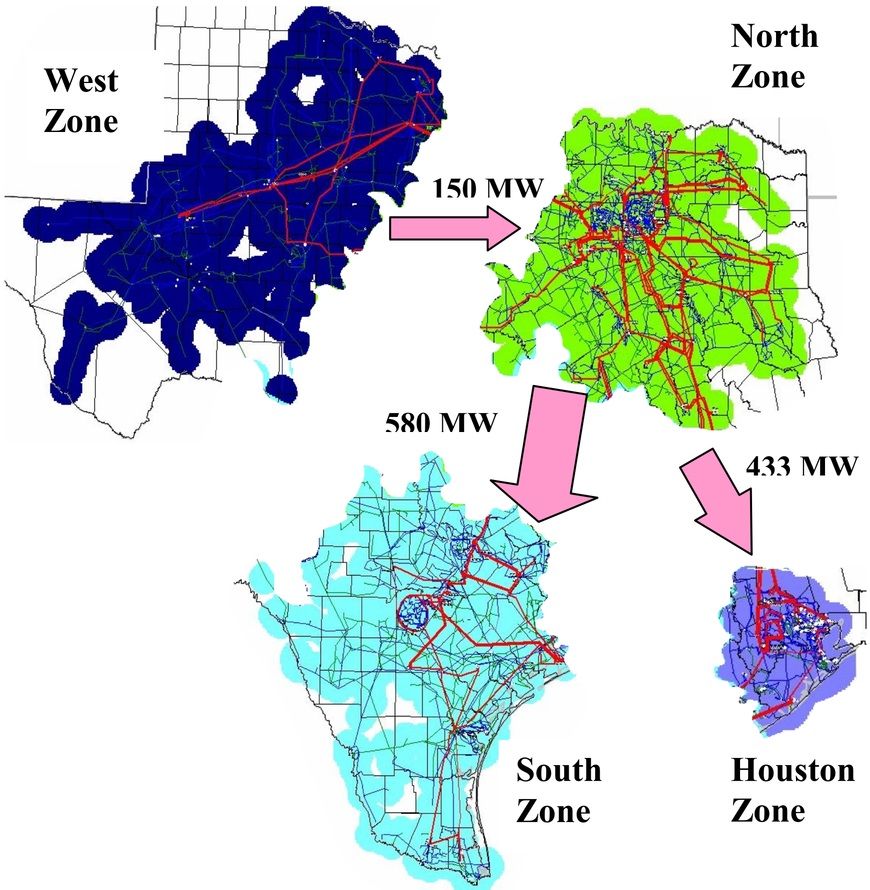
C. Transmission and Congestion

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 2008 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.

The levels of interzonal congestion increased considerably in 2008, particularly for the North to Houston and North to South CSCs (which was new in 2008), and for the West to North CSC.

Beginning in April and continuing into May 2008, the frequency of congestion on the North to Houston and North to South CSCs began to increase, at times becoming so significant that the constraint was unable to be resolved with available balancing energy in the zones where it was required. When congestion on a CSC cannot be resolved, maximum shadow prices are produced for the CSC that, in turn, produce balancing energy market prices in the deficient zones that can approach or even exceed the system-wide offer caps (the system-wide offer cap was $2,250 per MWh beginning March 1, 2008).[5] Historically, the inability to resolve a zonal constraint has been a relatively rare occurrence. In fact, excluding the North to Houston and North to South CSCs, the other CSCs together averaged only 15 intervals in 2008 with shadow prices that were greater than or equal to the current maximum CSC shadow price of $5,000 per MW. In contrast, the North to South and North to Houston CSCs experienced shadow prices greater than or equal to $5,000 per MW in 92 and 87 intervals, respectively.

The sharp increase in the frequency of occurrence of unresolved congestion on the North to Houston and North to South CSCs prompted the IMM, in consultation with ERCOT and the PUCT, to initiate in early May 2008 a detailed examination of ERCOT's congestion management procedures. This investigation quickly revealed that ERCOT rules permitted certain transmission elements to be managed with zonal balancing energy deployments when, in actuality, the congestion on these elements was neither effectively nor efficiently resolvable with zonal balancing energy deployments (the transmission elements that can be designated to be managed with zonal balancing energy deployments in the same manner as the CSC are referred to as "Closely Related Elements (CREs)" in the ERCOT Protocols).

Under the current zonal market model, transmission congestion is resolve through a bifurcated process that consists of either (1) zonal balancing energy deployments, or (2) local, unit-specific deployments. Because the pricing and incentives for both load and resources are better aligned with zonal congestion management techniques under the zonal market model, it is preferable to

---

5 A shadow price represents the marginal generation cost savings that would be achieved if one additional MW of capacity were available on a constraint.

manage transmission congestion by using zonal balancing energy, to the extent it is effective and efficient.

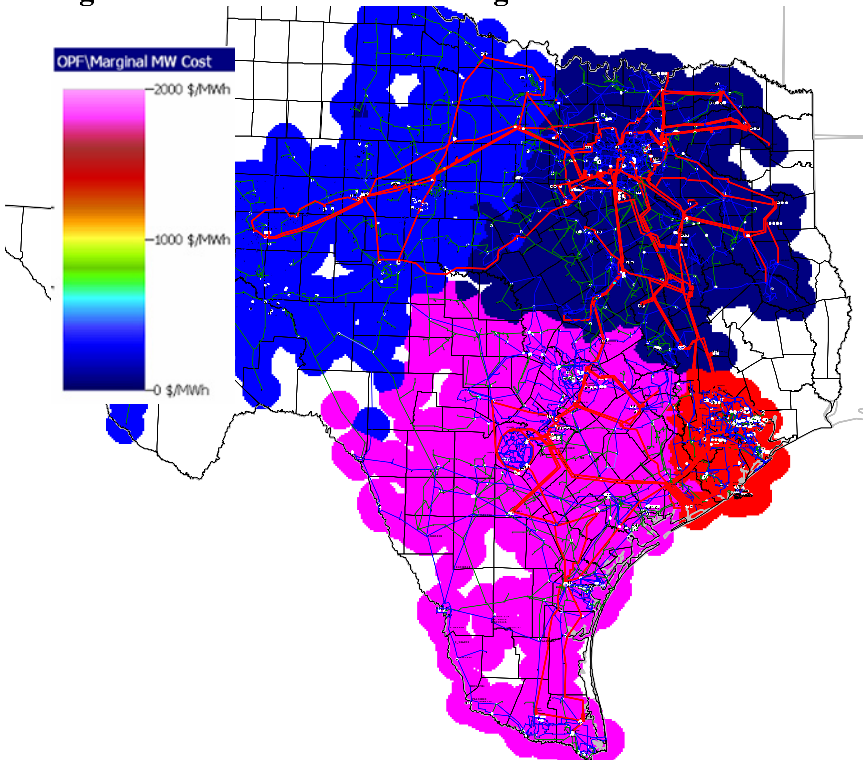
However, for the CREs in question related to the North to Houston and North to South CSCs, zonal balancing energy deployments were neither effective nor efficient in resolving the transmission congestion. The result was an increasing frequency of the deployment of substantial quantities of energy in both the Houston and South Zones up to the point of exhaustion, thereby triggering the maximum shadow prices for these CSCs and associated high balancing energy prices in the South and Houston Zones that approached or even exceeded the system-wide offer cap of $2,250 per MWh.

To address these market and reliability issues, in late May 2008, again in consultation with ERCOT and the PUCT, the IMM submitted PRR 764, which improved the definition of those transmission elements eligible to be designated as CREs. PRR 764 was processed through the ERCOT committees on an expedited basis and was implemented on June 9, 2008.

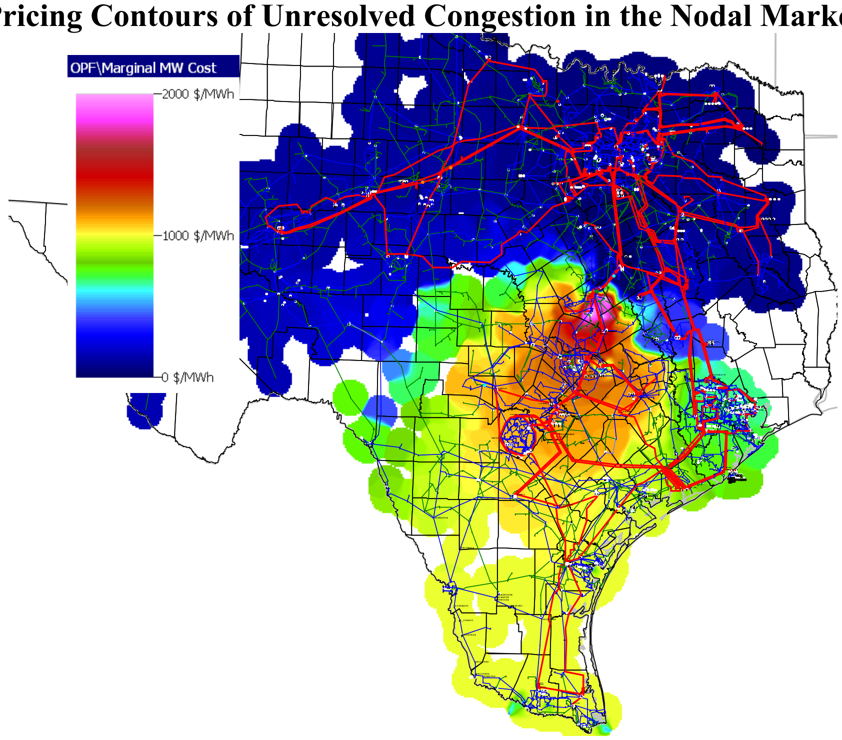
While PRR 764 effectively resolved the issues encountered during April through June 2008 within the context of the zonal market model framework, the implementation of the nodal market will eliminate the current bifurcated congestion management process by providing simultaneous, unit-specific solutions that will always presents the most effective and efficient congestion management alternatives to the system operator.

As a point of comparison, the following two figures show the pricing implications of unresolved North to South congestion under the zonal model and under a nodal model. These graphics show that an unresolved constraint produces extremely high market clearing prices are very widespread under the zonal model. In contrast, for the same unresolved constraint, the resulting high prices remain much more localized under the nodal model.

Pricing Contours of Unresolved Congestion in the Zonal Market



Pricing Contours of Unresolved Congestion in the Nodal Market



These differences in pricing outcomes are due to the use of zonal average shift factors under the zonal model compared to the use of location-specific shift factors under the nodal model.

Further, because the nodal market employs unit-specific offers and dispatch, the control of power flows on the system is much more flexible and precise than under the zonal model. Hence, it is much less likely under nodal dispatch to even encounter unresolvable constraints such as those experienced on the North to South and North to Houston interfaces in 2008.

In consideration of these differences, we have estimated the benefits that the nodal market would have produced by allowing more efficient resolution of the congestion on the North to South and North to Houston interfaces in 2008. This analysis indicates that the annual average balancing energy market price in the Houston and South Zones would have been reduced by approximately \$10.42 per MWh. Assuming that only 5 to 10 percent of customers in the South and Houston Zones were directly affected by the significant price increases in the balancing energy market and short-term bilateral markets associated with the North to Houston and North to South congestion, this analysis indicates that the efficiencies of the nodal market, had it been in place, could have reduced the annual costs for customers by \$87 to \$175 million in 2008. This analysis estimates only the savings that could have occurred through more efficient congestion management during the periods of acute North to Houston and North to South congestion, and does not include the benefits that the nodal market will provide more generally with respect to congestion management and other dispatch efficiency improvements.

The other CSC experiencing a significant increase in the level of congestion in 2008 was the West to North CSC, which was binding in 5,320 intervals (15 percent). This was more frequent than any other CSC in 2008 and 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 is due to the significant increase in installed wind generation relative to the load in the West Zone and limited transmission export capability to the broader market. The installed wind capacity in ERCOT grew from approximately 4.5 GW at the beginning of the year to approximately 8.1 GW by December 2008, with more than 90 percent of wind capacity located in the West Zone. Depending on load levels, transmission system topology and other system operating conditions, up to 4 GW of wind generation in the West Zone can be reliably produced before encountering a transmission export constraint on the West to North CSC.

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 1,100 hours during 2008.

Although plans exist through the Competitive Renewable Energy Zone ("CREZ") project to significantly increase the transmission export capability from the West Zone, it is likely given the current transmission infrastructure and the level of existing wind facilities in the West Zone that the quantity of wind production that can be reliably accommodated in the West Zone will continue to be significantly limited for several years until such transmission improvements can be completed.

2. Transmission Congestion Rights and Payments

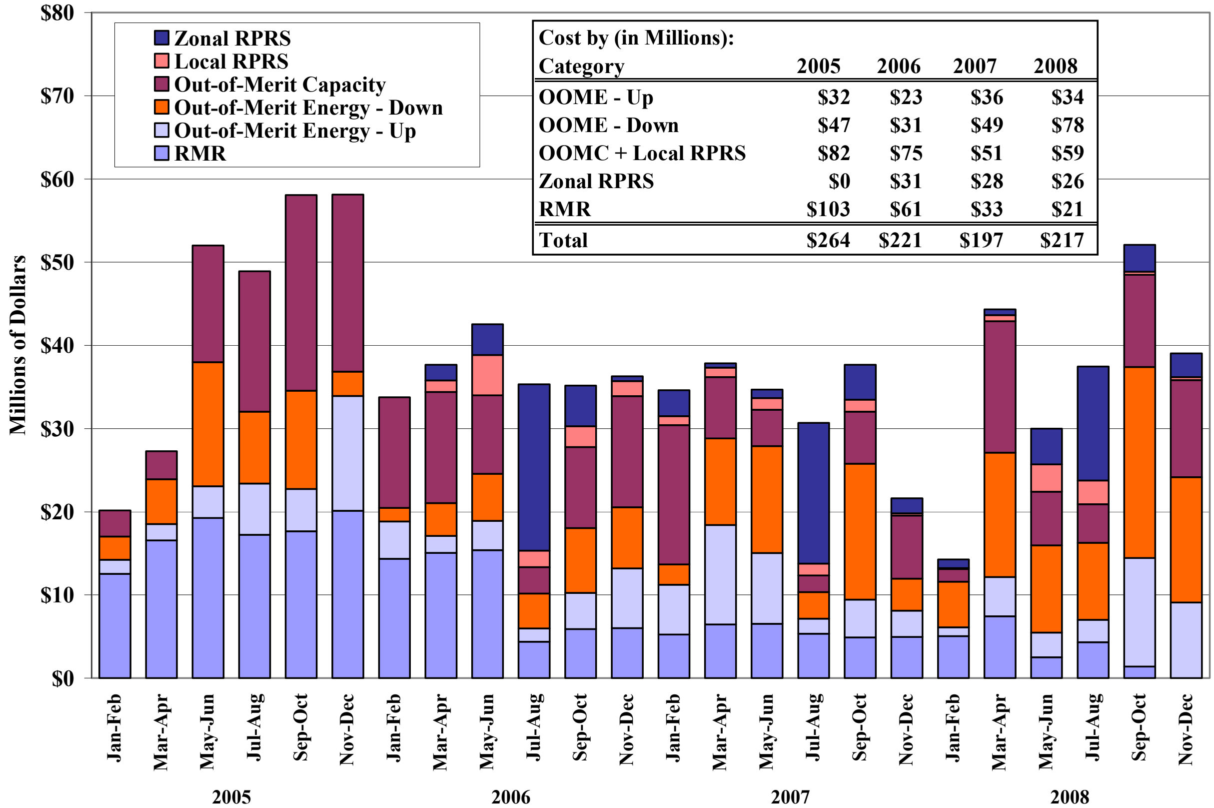
Participants in Texas can hedge against congestion in the balancing energy market by acquiring Transmission Congestion Rights ("TCRs") between zones, which entitle the holder to payments equal to the difference in zonal balancing energy prices. Because the modeled limits for the CSC interfaces vary substantially, 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. The aggregate shortfall increased to $99 million in 2008, up from $61 million in 2007. This increase was primarily due to increased interzonal congestion in 2008 and decreased accuracy in the quantity of TCRs sold in the monthly auction, especially for the West-to-North CSC.

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. In 2006, market participants over-estimated the value of congestion on the South to North, South to Houston, and North to Houston CSCs. In 2007, market participants still over-estimated the value of congestion on the South to North and South to Houston CSCs, but significantly under-estimated the value of congestion on the North to Houston, North to West and West to North CSCs. However, 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.

3. Local Congestion and Local Capacity Requirements

ERCOT manages local (intrazonal) congestion 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 these units are called out-of-merit order, 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 2005 to 2008.

Expenses for Out-of-Merit Capacity and Energy

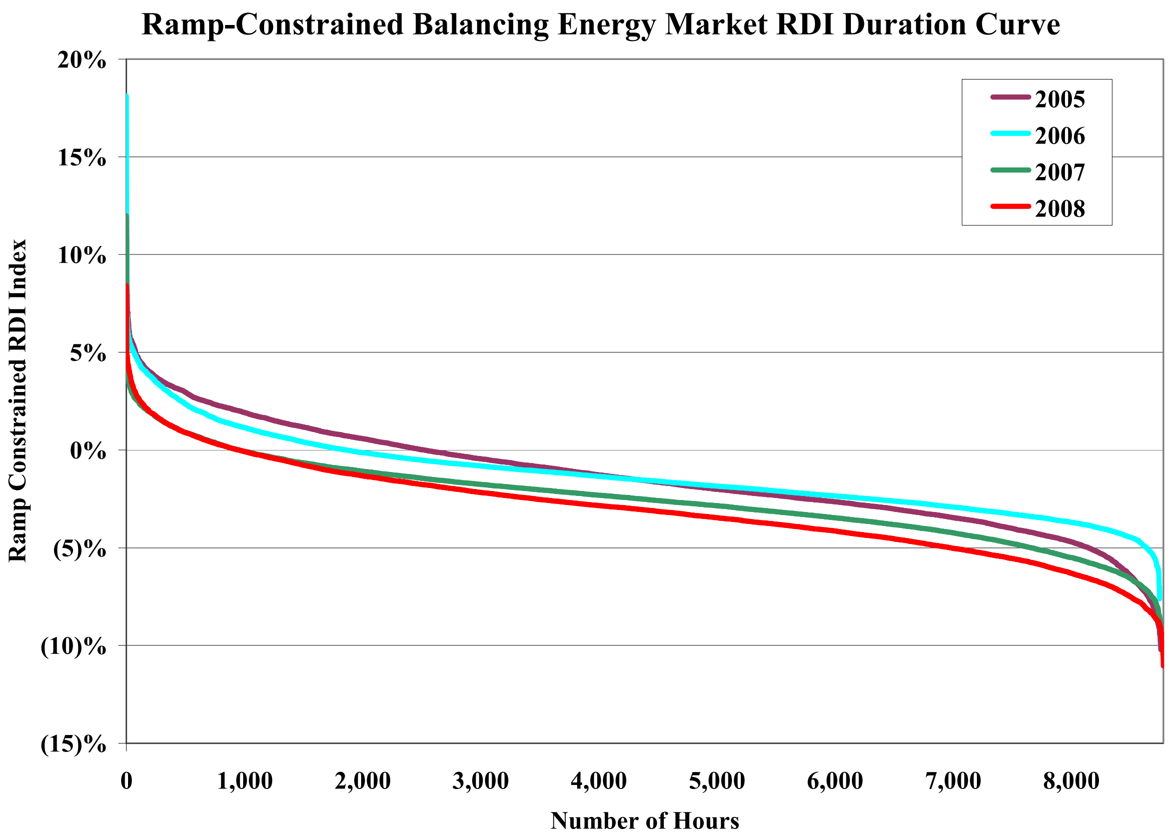


The results in the figure above show that overall uplift costs for RMR units, OOME units, OOMC/Local RPRS and Zonal RPRS units were $217 million in 2008, which is a $20 million increase over the $197 million in 2007. OOME Down costs accounted for the most significant portion of the change in 2008, increasing from $49 million in 2007 to $78 million in 2008. OOMC/Local RPRS costs increased from $50 million in 2007 to $60 million in 2008, and RMR costs decreased from $33 million in 2007 to $20 million in 2008. 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.

D. 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 needed to serve the ERCOT load and satisfy the ancillary services requirements.

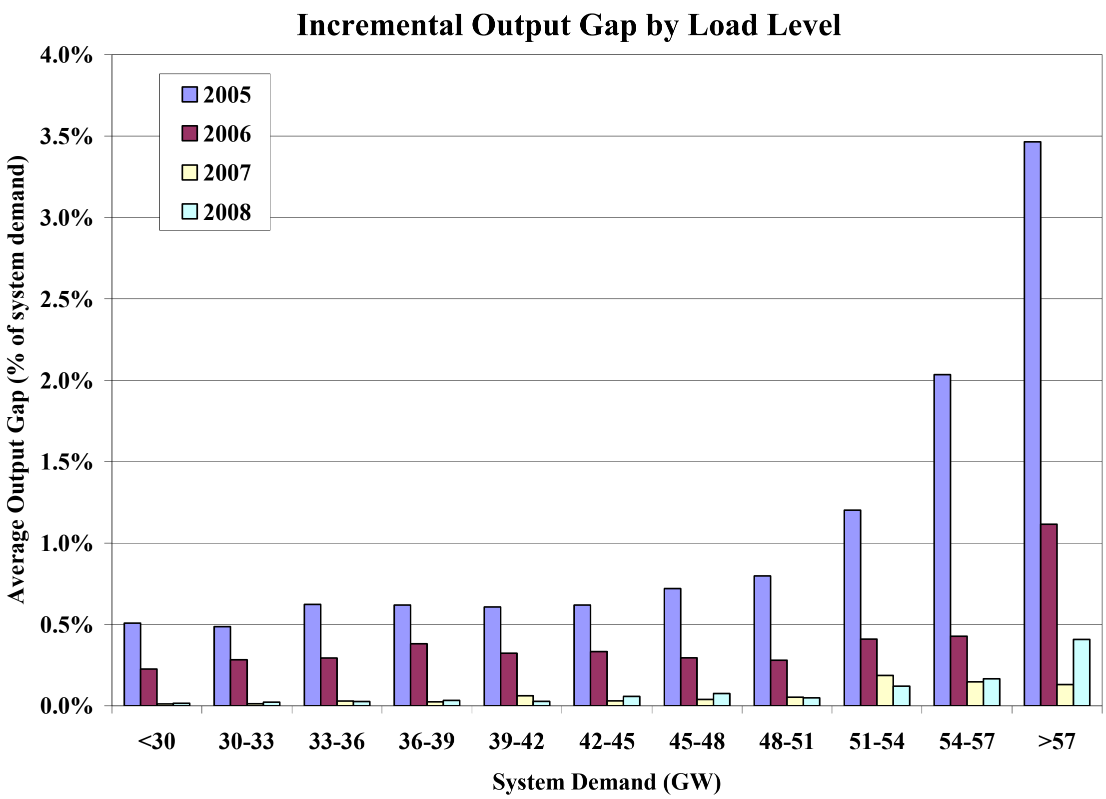
The pivotal supplier analysis indicates that the frequency with which a supplier was pivotal in the balancing energy market remained relatively constant in 2008 compared to 2007. The following figure shows the ramp-constrained balancing energy market Residual Demand Index ("RDI") duration curves for 2005 through 2008. 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.



The frequency with which at least one supplier was pivotal (*i.e.*, an RDI greater than zero) has fallen over the last four years from 29 percent of hours in 2005 to 21 percent of the hours in 2006 and less than 11 percent of hours in 2007 and 2008. These results indicate that the structural competitiveness of the balancing energy market in 2008 maintained the improvement exhibited in 2007 compared to prior 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.

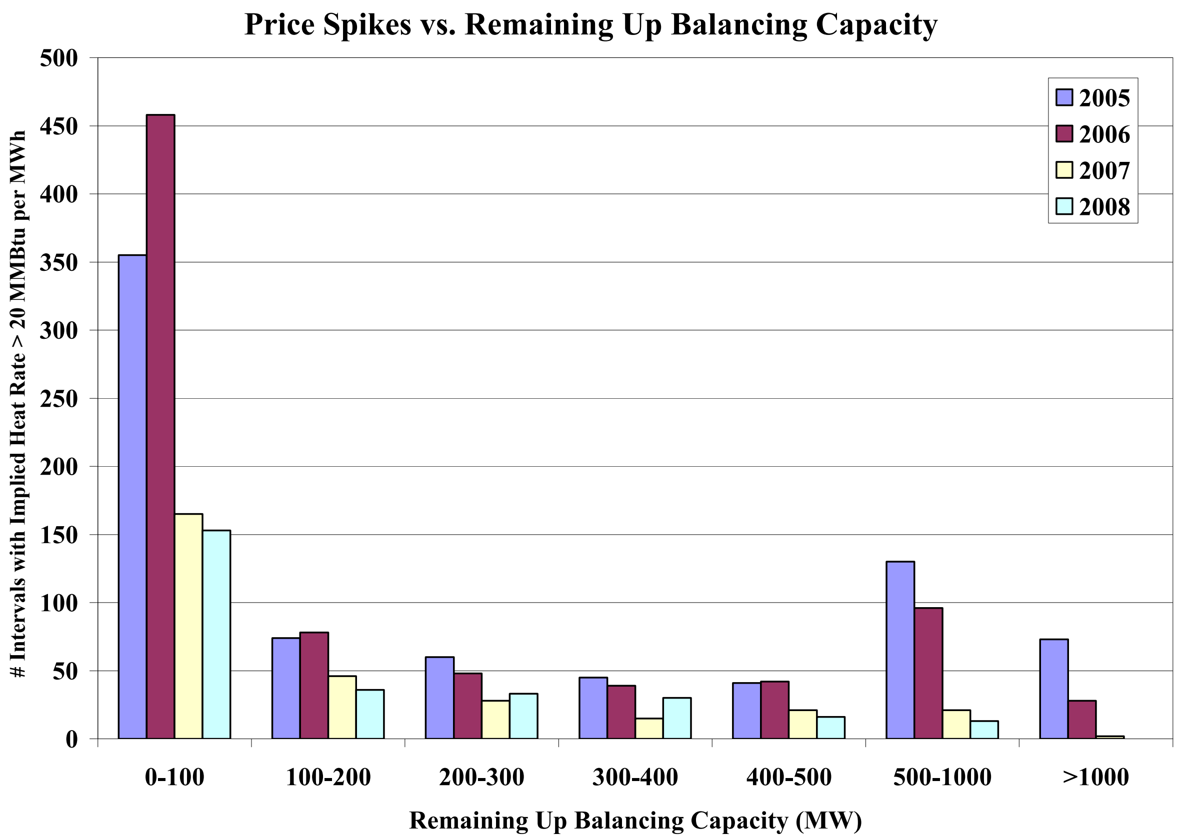
The figure below compares the real-time load to the average incremental output gap for all market participants as a percentage of the real-time system demand from 2005 through 2008.



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 and 2008. Although 2008 exhibited a higher average incremental output gap at the highest load levels, the overall magnitude remains small and does not raise significant economic withholding concerns.

A final measure used to evaluate the competitiveness of the market outcomes in 2008 analyzes the number of balancing energy market price spikes compared to the quantity of remaining Up Balancing capacity. If the market is operating competitively, price spikes should occur during shortage and near shortage conditions, and the number of price spikes should reduce significantly as the amount of available surplus capacity increases.

For the purpose of this analysis, a price spike is measured as an interval in which the balancing energy market price exceeded an implied heat rate of 20 MMBtu per MWh, which is greater than the marginal costs of most online generating units. However, the marginal cost of offline quick start units is often greater than this threshold. Thus, some of the price spikes in this figure are indicative of the deployment of quick start gas turbines, particularly in 2007 and 2008 when several market participants had well over 1,000 MW of quick start capability qualified to provide balancing energy. In contrast, in 2005 only one market participant had quick start unit qualified to provide balancing energy (Austin Energy; 7 units and approximately 330 MW), and in 2006 one additional market participant had qualified quick start gas turbines (CPS Energy; 4 units and approximately 200 MW).



The results in the figure above indicate very competitive market outcomes in 2008, with over 95 percent of the price spikes occurring during intervals with less than 500 MW of Up Balancing capacity remaining. These results show significant improvement over 2005 and 2006 when only 74 and 84 percent, respectively, of the price spikes occurred during intervals with less than 500 MW of Up Balancing capacity remaining.

The changes in the market outcomes from 2005 through 2008 shown in the figure above are consistent with expectations given the improvements in structural and supplier conduct competitiveness over this timeframe that are highlighted in the previous two figures.

Overall, based upon the analyses in this section, we find that the ERCOT wholesale market performed competitively in 2008.

I. REVIEW OF MARKET OUTCOMES

A. Balancing Energy Market

1. Balancing Energy Prices During 2008

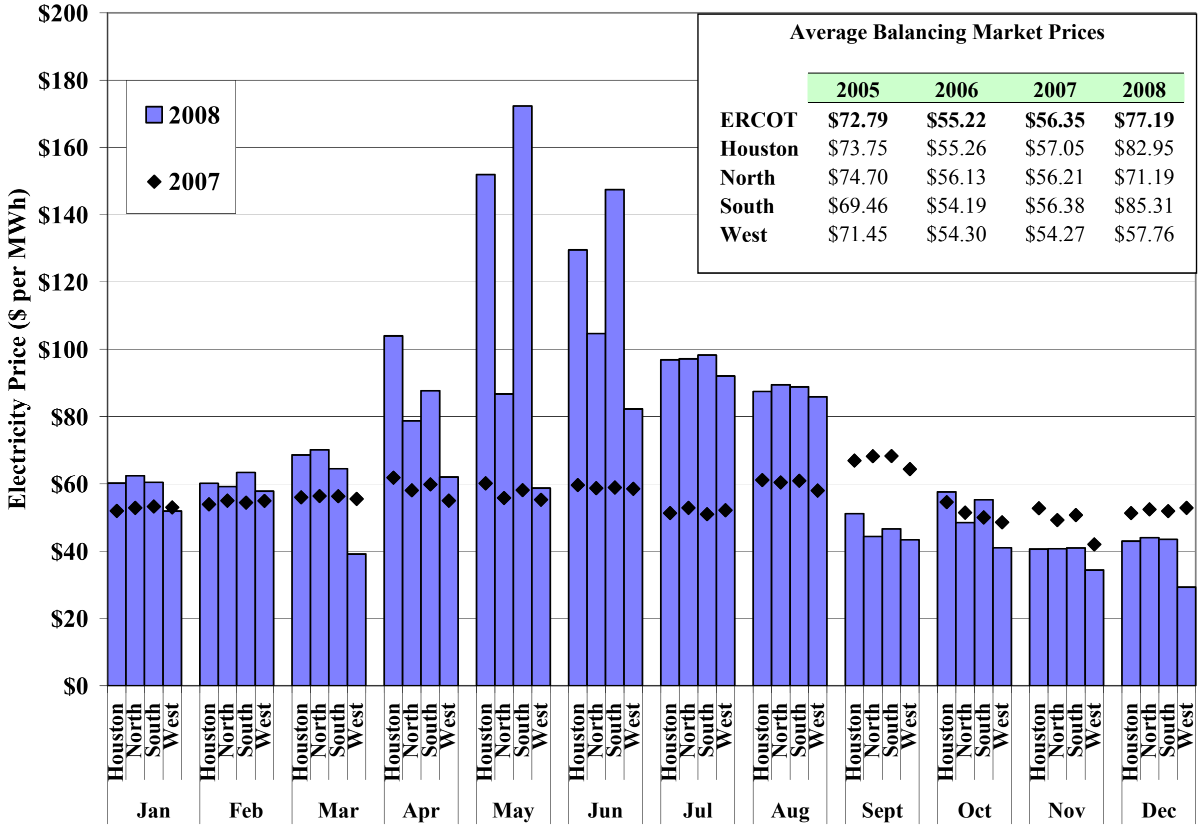
The balancing energy market is the spot market for electricity in ERCOT. As is typical in other wholesale markets, only a small share of the power produced in ERCOT is transacted in the spot market, although at times such transactions can exceed 10 percent of total demand. Although most power is purchased 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 2008.

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 2007 and 2008, with annual summary data for 2005 and 2006.[6]

[6] 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. This is not consistent with average prices reported elsewhere in this report that are weighted by the balancing energy procured in the 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

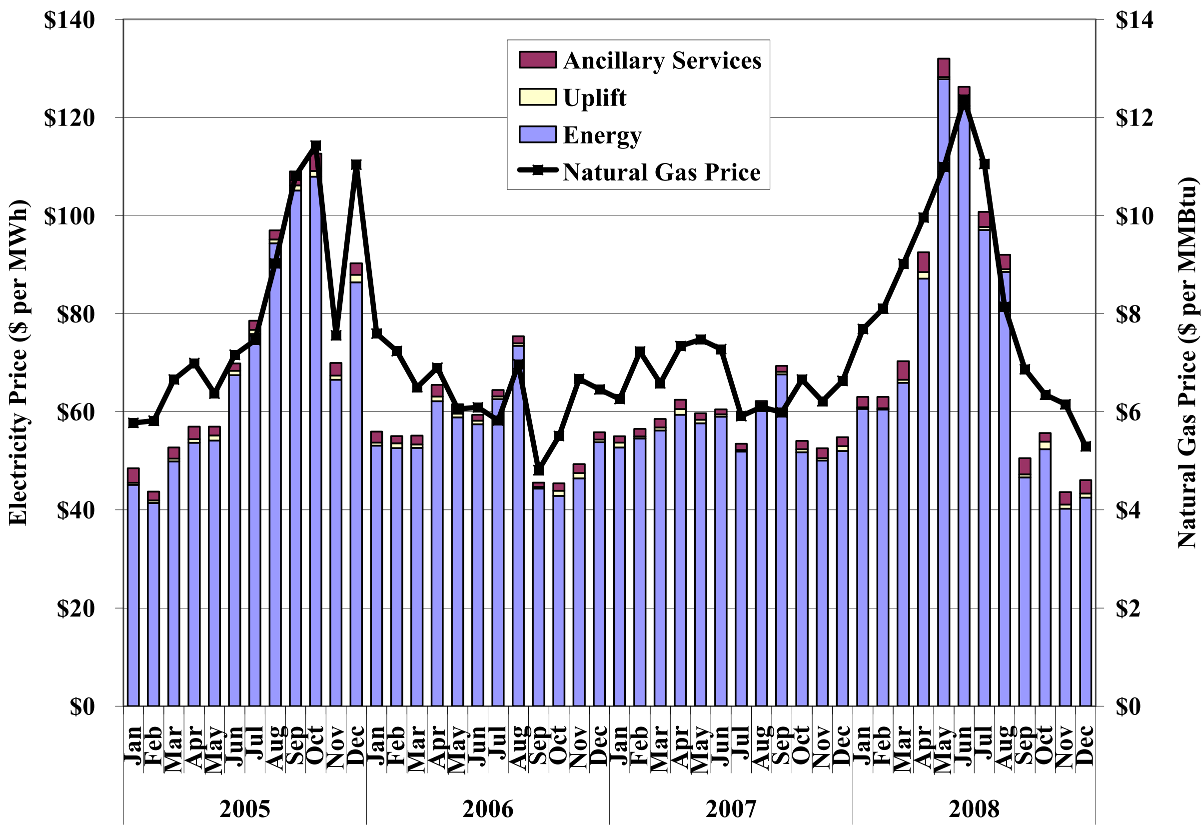


ERCOT average balancing energy market prices were 37 percent higher in 2008 than in 2007. May, June and July experienced the highest balancing energy market price increases in 2008 at 222, 206 and 187 percent, respectively, of the prices in the same months in 2007. In contrast, average balancing energy market prices in September, November and December 2008 decreased to 69, 80 and 82 percent, respectively, of the prices in the same months in 2007.

The average natural gas price increased 28 percent in 2008 over 2007. May, June and July experienced the highest natural gas price increases in 2008 at 147, 170 and 187 percent, respectively, of the prices in the same months in 2007. In contrast, average natural gas prices in October, November and December 2008 decreased to 95, 99 and 80 percent, respectively, of the prices in the same months in 2007. Natural gas is typically the marginal fuel in the ERCOT market. Hence, the changes in energy prices from 2007 to 2008 were largely a function of natural gas price movements, although significant transmission congestion during April, May and June was also responsible for approximately one-quarter of the average balancing energy market price increase in 2008.

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".[7] 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 2005 to 2008 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.

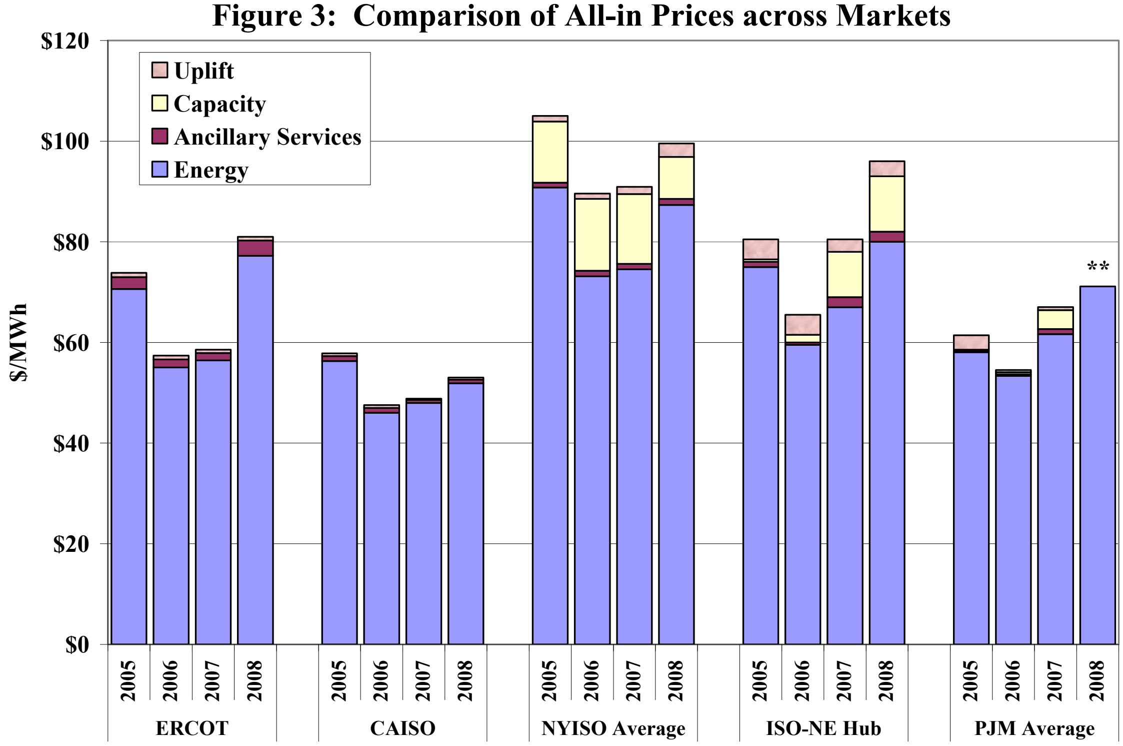
7 As discussed in more detail in Section III, uplift costs are costs that are allocated to load that pay for out-of-merit dispatch, out-of-merit commitment, and Reliability Must Run contracts.

- 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, and Reliability Must Run contracts.

Figure 2 indicates that natural gas prices were a primary driver of the trends in electricity prices from 2005 to 2008. Again, this is not surprising given that natural gas is the predominant fuel 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

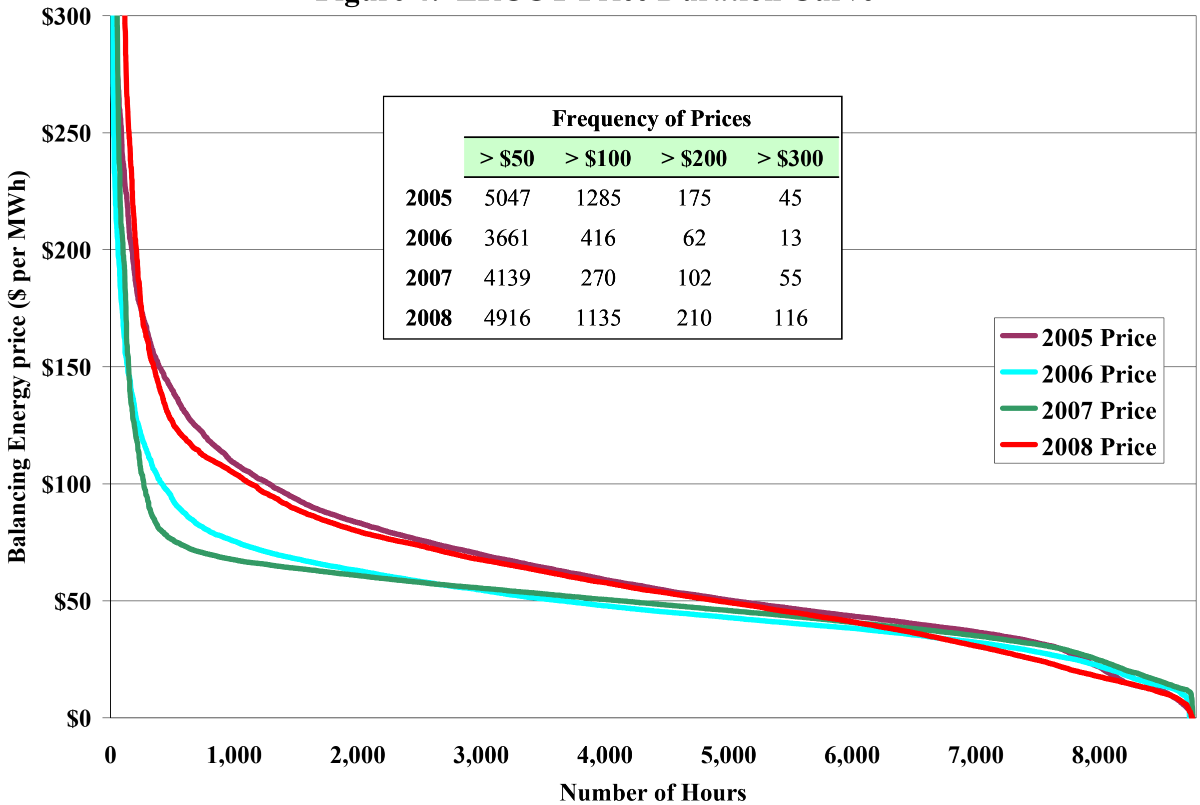


** 2008 Capacity, Ancillary Services and Uplift data unavailable for PJM

Figure 3 shows that energy prices increased in wholesale electricity markets across the U.S. in 2008, primarily due to increases in fuel costs.

Figure 4 presents price duration curves for the ERCOT balancing energy market in each year from 2005 to 2008. 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 in more than 4,900 hours in 2008 compared to more than 4,100 hours in 2007. These year-to-year changes reflect the effects of higher fuel prices in 2008 which impact electricity prices in a broad range of hours.

Figure 5 shows the hourly average price duration curve for each of the four ERCOT zones in 2008.

Figure 5: Zonal Price Duration Curves

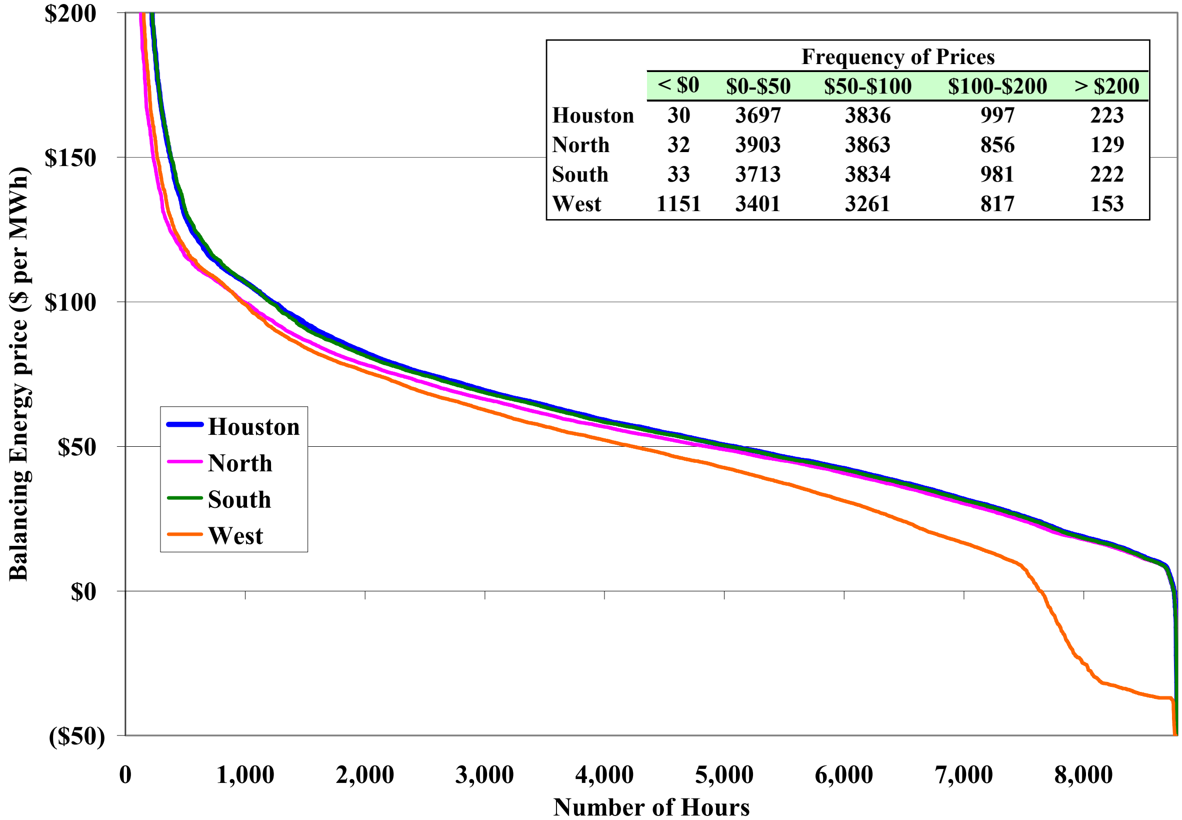
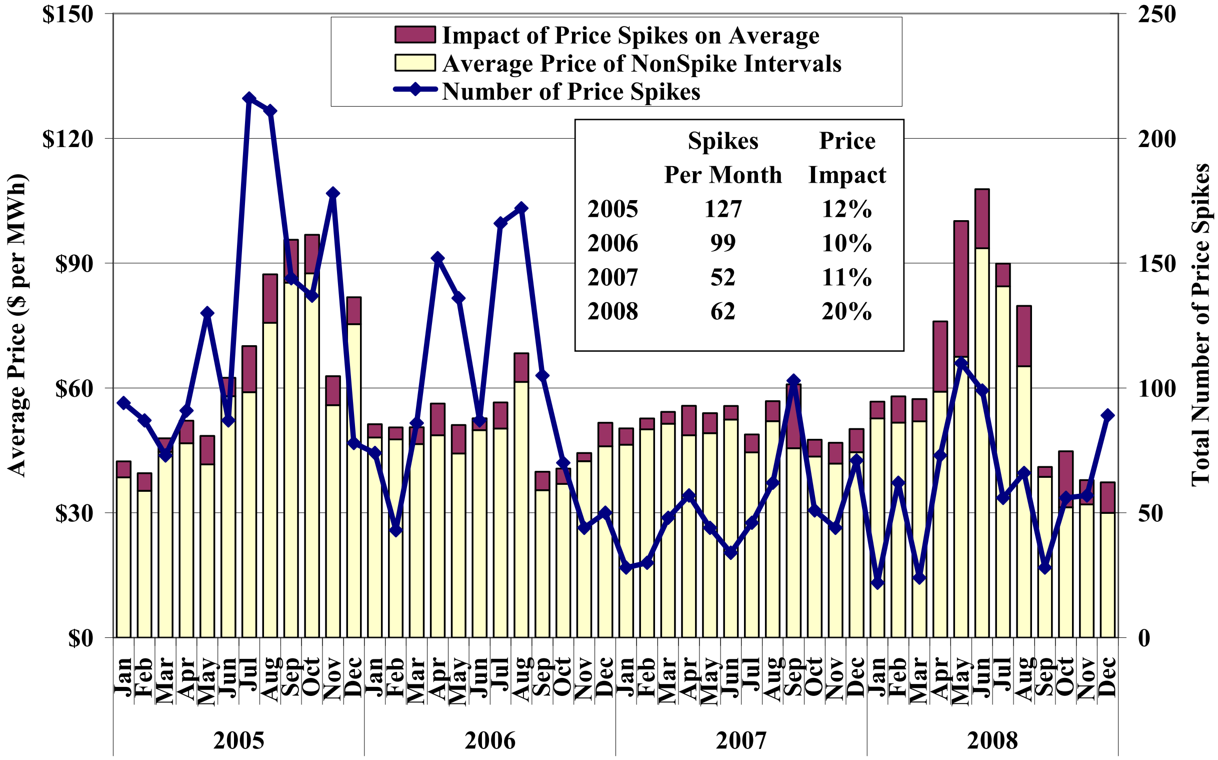


Figure 5 shows that the Houston, North and South Zones had similar prices over the majority of hours in 2008, but that the Houston and South Zones each experienced significantly more hours in which the price exceeded $200. The price duration curve for the West Zone is generally lower than all other zones, with over 1,100 hours when the average hourly price was less than zero. These zonal price differences are caused by zonal transmission congestion, the details of which are discussed 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 2004 and 2008 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 2005 to 2008. Figure 6 shows average prices and the number of price spikes in each month of 2005 to 2008. 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 generators in ERCOT).

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



The number of price spike intervals was 99 per month during 2006. The number decreased in 2007 to 52 per month, and increased to 62 per month in 2008. In 2008, the highest frequency of price spikes occurred in May with 110 price spikes, which was caused by significant transmission congestion and is discussed in more detail in Section III. 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 approximately $6.98, $4.68, $5.30 and $10.71 per MWh during 2005, 2006, 2007, and 2008, 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. At least five other factors provided a meaningful contribution to price outcomes in 2008.

First, as discussed in Section II, ERCOT peak demand and installed capacity were relatively flat in 2008, and energy production increased only slightly in 2008 compared to 2007. These results were similar to 2007 compared to 2006. In contrast to years prior to 2007 that experienced increasing demand and decreasing supply, the static supply and demand characteristics from 2007 to 2008 contributed to comparable wholesale pricing outcomes over the course of these two years, with the exception of the second factor, which is transmission congestion.

As discussed in Section III, chronic and severe transmission congestion from North to South and North to Houston materialized in April, May and June 2008 that had a significant effect on balancing energy pricing outcomes, particularly in the Houston and South zones. In addition, significant increases in installed wind generation in the West Zone led to an increase in West to North congestion, in turn producing a load-weighted average price in the West Zone that was approximately 26 percent below the ERCOT average price in 2008, with wind resources frequently being the marginal generation source in the West Zone.

Third, aside from the effect of wind generation on the West Zone prices, the continued increase in wind production in 2008 served to displace more costly generation resources when the wind was producing. This will tend to lower average prices across the market, but the intermittent nature of wind can also lead to transitory price spikes as other generation resources may be required on short notice to fill the gap left by significantly lower than expected or rapidly declining wind output.

Fourth, the balancing energy offer cap increased to $2,250 per MWh on March 1, 2008, consistent with Commission rule. Prior to March 1, the rule had set the offer cap at $1,500 per MWh. The increased offer cap is intended to produce higher prices during system shortage conditions as a part of the PUCT's rules that rely upon energy prices exclusively to ensure generation resource adequacy as opposed to the reliance on both capacity and energy prices used in most other domestic organized electricity markets. As discussed in Section II, this mechanism was not always effective in achieving this intended outcome, and some of ERCOT's reliability-based actions can often disrupt the market-based balance of supply and demand, thereby frustrating the long-term success of the energy-only market.

Finally, the overall competitive performance of the market exhibited continued improvement in 2008, which will tend to lower prices. We examine competitive performance in detail in Section IV. Analyses in the next sub-section 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 sub-section 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 corrected for 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*.[8] The second chart shows the same duration curves for the five percent of hours in each year with the highest implied heat rate. The figure shows duration curves for the implied marginal heat rate for 2005 to 2008.

In contrast to Figure 4, Figure 7 shows that the implied marginal heat rates were relatively consistent across the majority of hours from 2005 to 2008. The rise in energy prices from 2007 to 2008 is much less dramatic when we explicitly control for fuel price changes, 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 was 73 in 2006, 103 in 2007, and increased to 145 in 2008. This indicates that there are price differences that are due to factors other than changes in natural gas prices.

Figure 8 shows the implied marginal heat rates for the top five percent of hours in 2004 through 2008. These data reveal that the frequency of price spikes with an implied marginal heat rate greater than 30 increased significantly in 2008 compared to prior years.

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

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

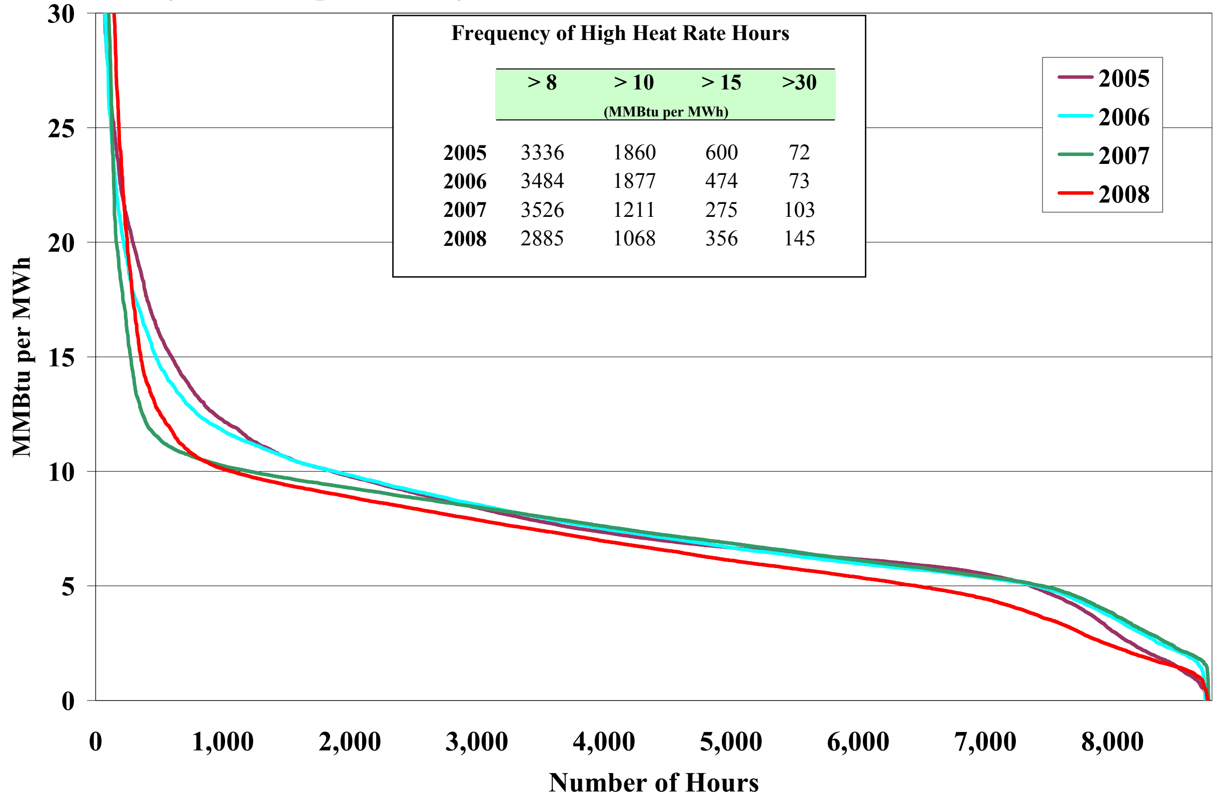
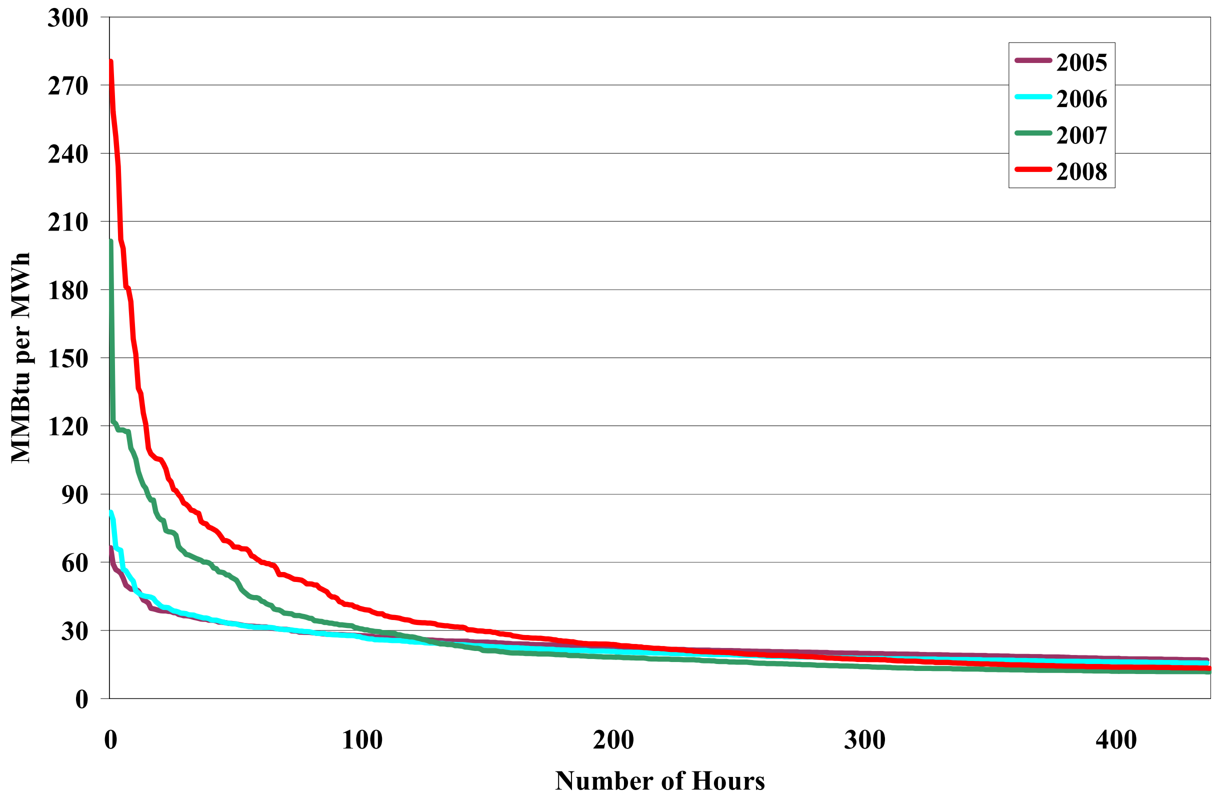
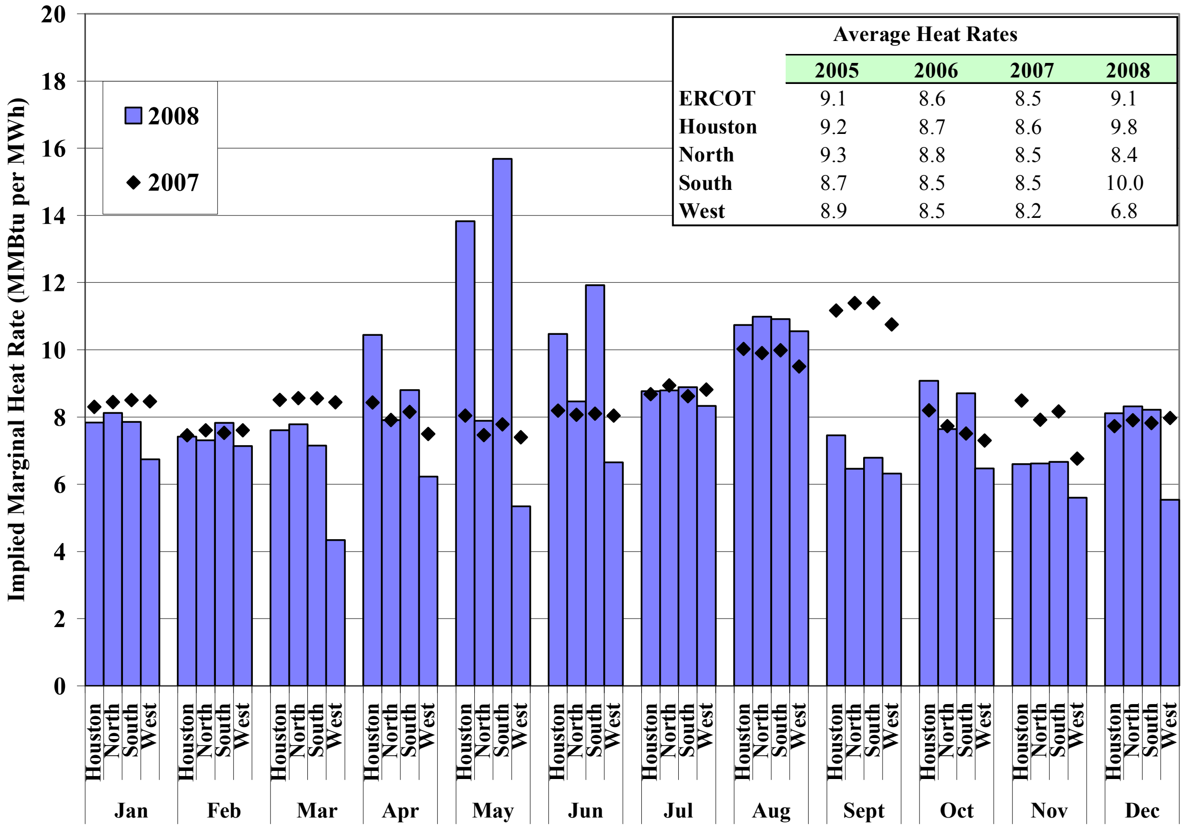


Figure 8: Implied Marginal Heat Rate Duration Curve – Top 5% of Hours



To better understand these differences, the next figure shows the implied marginal heat rates on a monthly basis in each of the ERCOT zones in 2007 and 2008, with annual summary data for 2005 and 2006. 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 increased by 6.8 percent from 8.50 in 2007 to 9.08 in 2008.

Figure 9: Monthly Average Implied Marginal Heat Rates



	2005	2006	2007	2008
ERCOT	9.1	8.6	8.5	9.1
Houston	9.2	8.7	8.6	9.8
North	9.3	8.8	8.5	8.4
South	8.7	8.5	8.5	10.0
West	8.9	8.5	8.2	6.8

The average implied heat rate was higher in 2008 than in 2007 for the months of April, May, June, August and October. The increases in implied heat rates during April through June compared to 2007 are explained primarily by significant transmission congestion that affected the Houston and South Zones most significantly, and is discussed in more detail in Section III. The increase in the implied heat rate in August and October was due to a greater number of shortage intervals in these two months in 2008 compared to 2007, as well as the effects of *ex post* pricing adjustments during the deployment of non-spinning reserves applied under then-existing ERCOT Protocols, which is discussed in more detail in Section II. In contrast, the implied heat rate in September 2008 was significantly lower than in September 2007. This is explained by two factors. First, September 2007 experienced more shortage intervals than September 2008, which led to an increase in the implied heat rate in September 2007. Second, demand in the ERCOT region was significantly reduced in September 2008 because of the landfall of Hurricane Ike causing widespread and prolonged outages in the Houston area. This suppressed demand and in turn resulted in a significant reduction in the implied heat rate in September 2008.

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 2005 and 2008.[9] 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 indicates 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 2005 and 2008.

9 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, we use the average of the North, South and Houston Zone products.

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

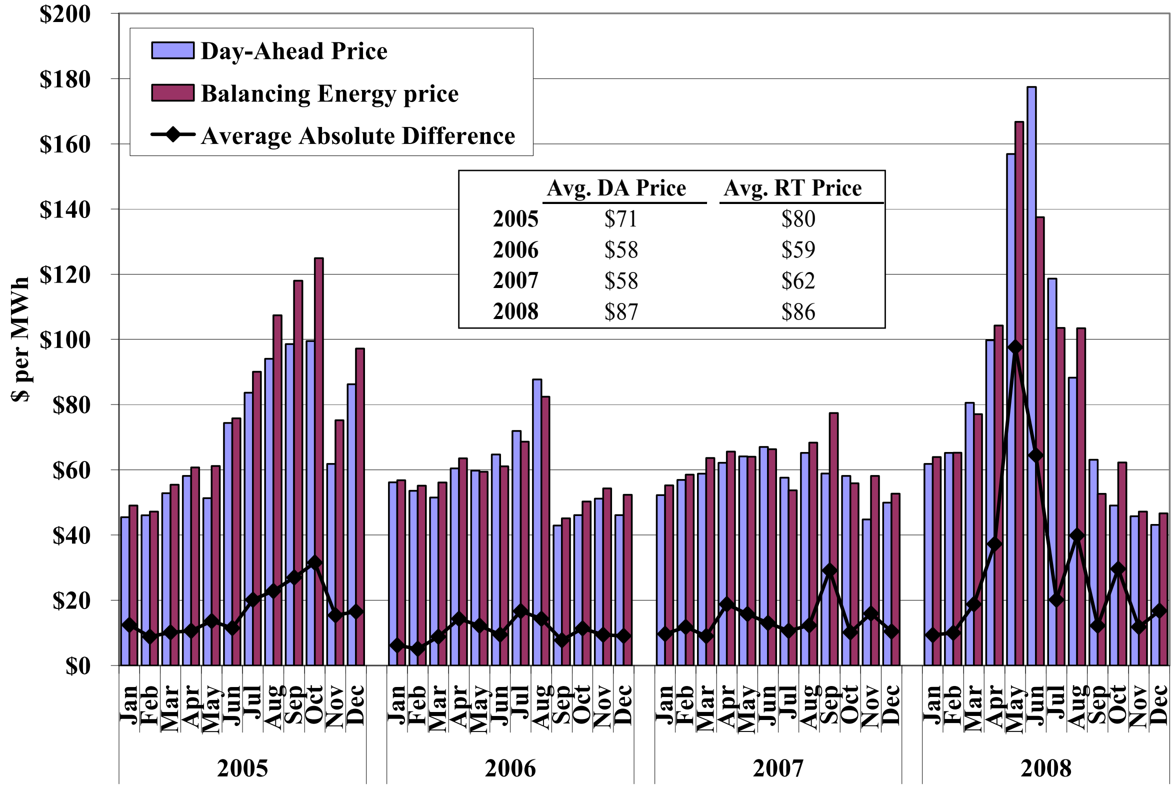


Figure 10 shows price convergence during peak periods (*i.e.*, weekdays between 6 AM and 10 PM). Day-ahead prices averaged $87 per MWh in 2008 compared to an average of $86 per MWh for real-time prices. Although the day-ahead and real-time prices exhibited good average convergence in 2008, Figure 10 also shows that the average absolute price difference was large for several months in 2008, particularly in April, May and June.

The average absolute difference was $17 in 2005, $10 in 2006, $14 in 2007 and $31 in 2008. As noted above, the average absolute difference measures the volatility of the price differences. The price volatility in April, May and June 2008 due in large part to the significant and unpredictable transmission congestion experienced in that timeframe. As discussed in Section III, ERCOT procedures were modified in June 2008 to improve the efficiency of congestion management within the context of the zonal market model, thereby leading to a reduction in the volatility of price differences. Relatively smaller spikes in the absolute price difference occurred in August and October 2008. These spikes were associated in part with certain *ex post* pricing revisions during the deployment of non-spinning reserves. As discussed in more detail in

Section II, changes were developed in 2008 and implemented in 2009 to improve the efficiency of pricing outcomes during the deployment of non-spinning reserves.

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

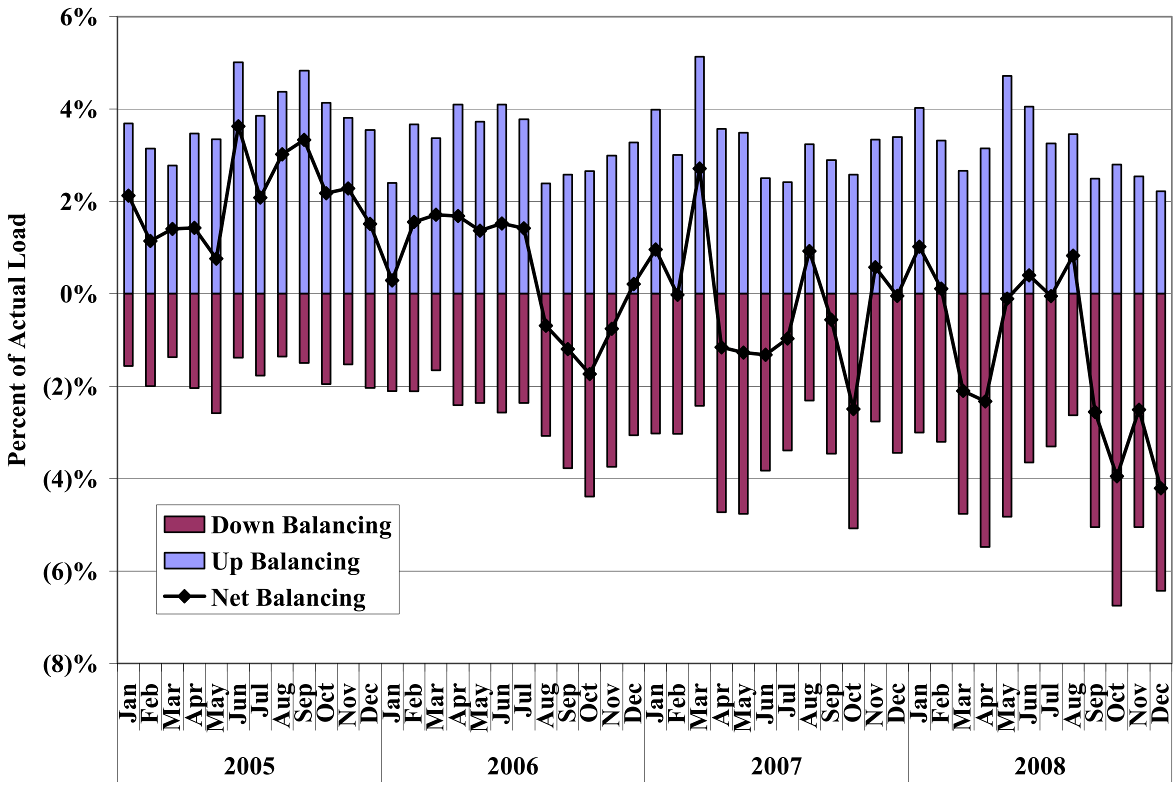
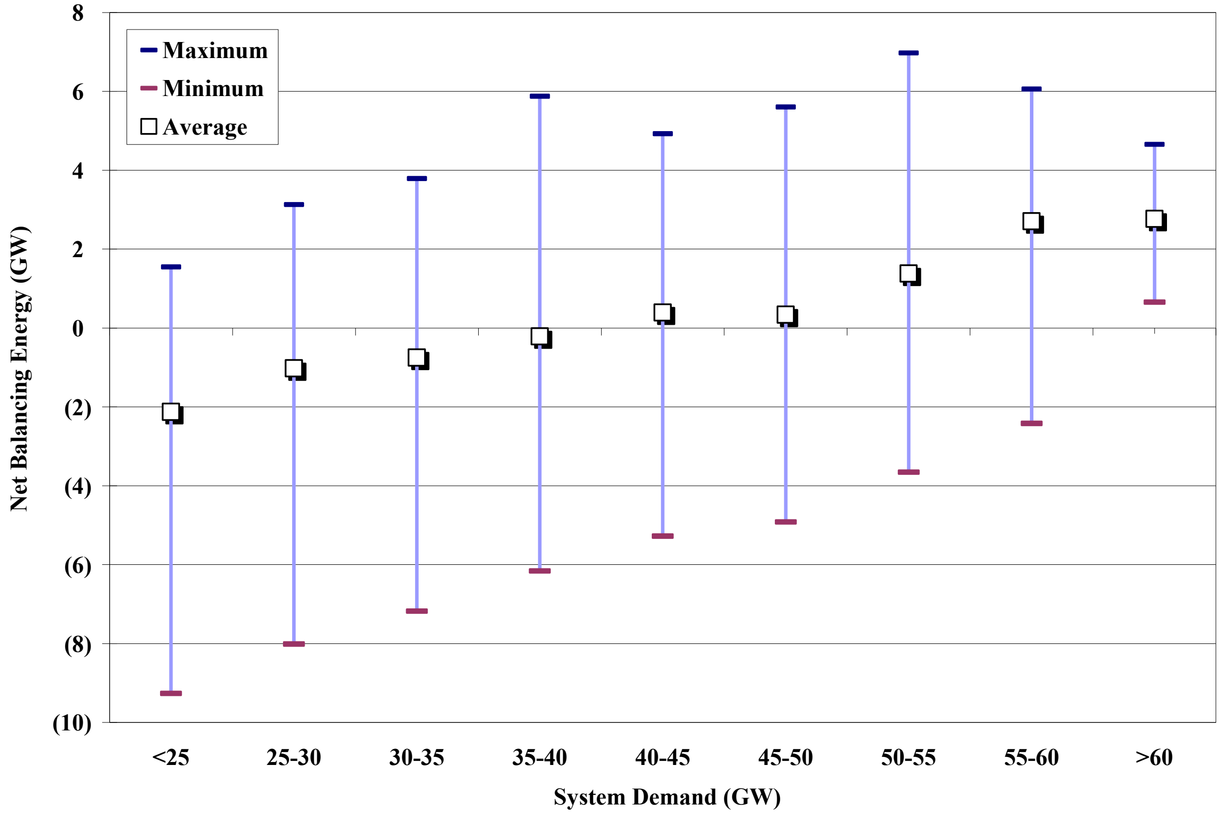


Figure 11 shows that the total volume of up balancing and down balancing energy as a share of actual load increased from an average of 6.8 percent in 2007 to 7.7 percent in 2008. Starting in August 2006, the average volume of down balancing energy began to increase. In 2008, for the first time the average amount of down balancing energy was greater than up balancing energy.

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 2008.

Figure 12: 2008 Net Balancing Energy by Load Level



While Figure 11 shows average net down balancing energy deployments in 2008, 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 40 GW, and average net up balancing deployments for load levels greater than 40 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 has allowed 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. Indeed, the tendency toward net up balancing energy purchases at times outside the summer months helps to explain the prevalence of price spikes during off-peak months. 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 12 percent of the hours in 2008.

Figure 13: Magnitude of Net Balancing Energy and Corresponding Price

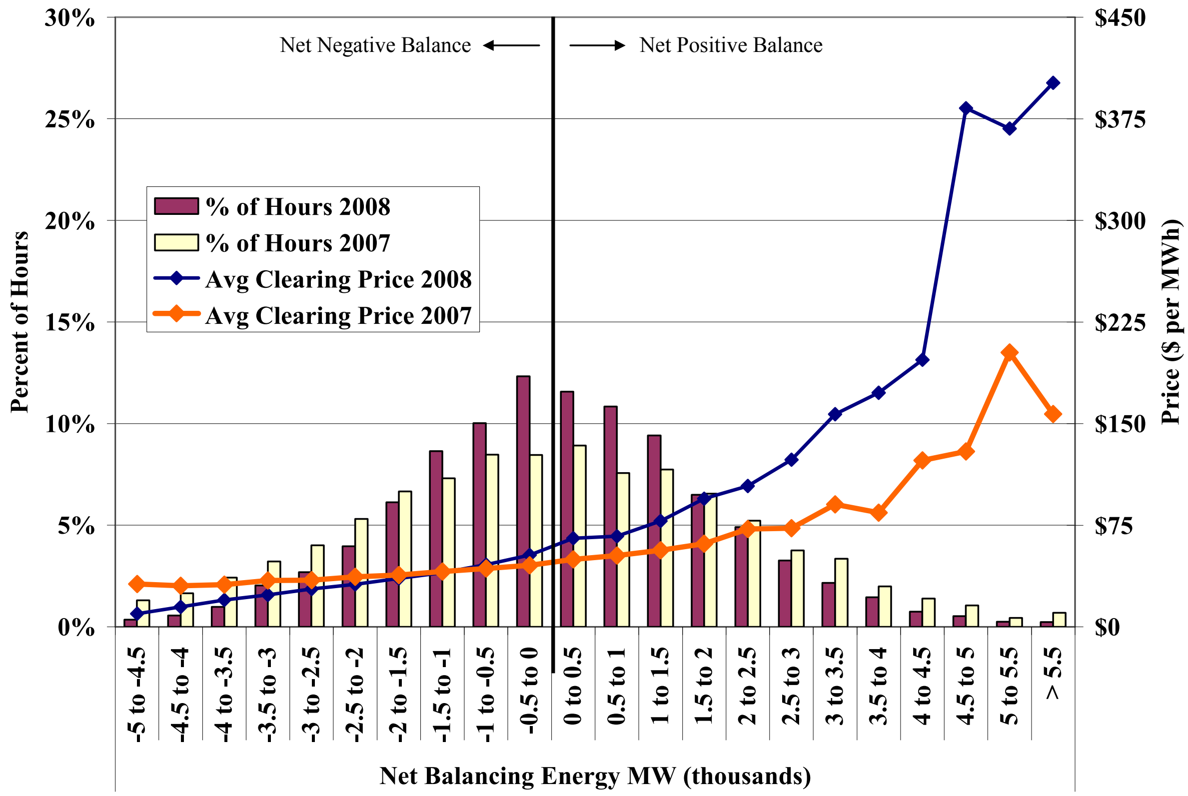


Figure 13 shows a relatively symmetrical distribution of net balancing energy purchases in 2007 centered around zero gigawatts, but 2008 is skewed more toward the left, meaning more down than up balancing energy was deployed. This is consistent with Figure 11 which showed that on average the quantity of net down balancing exceeded that of net up balancing during 2008. In approximately 45 percent of the hourly observations shown, net balancing energy schedules averaged between -1.0 and 1.0 GW. Hence, there were many hours when the net balancing energy traded was relatively low, because the total scheduled energy was frequently close to the actual load.

The lines plotted in Figure 13 shows the average balancing energy prices corresponding to each level of balancing energy volumes for 2007 and 2008. 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.

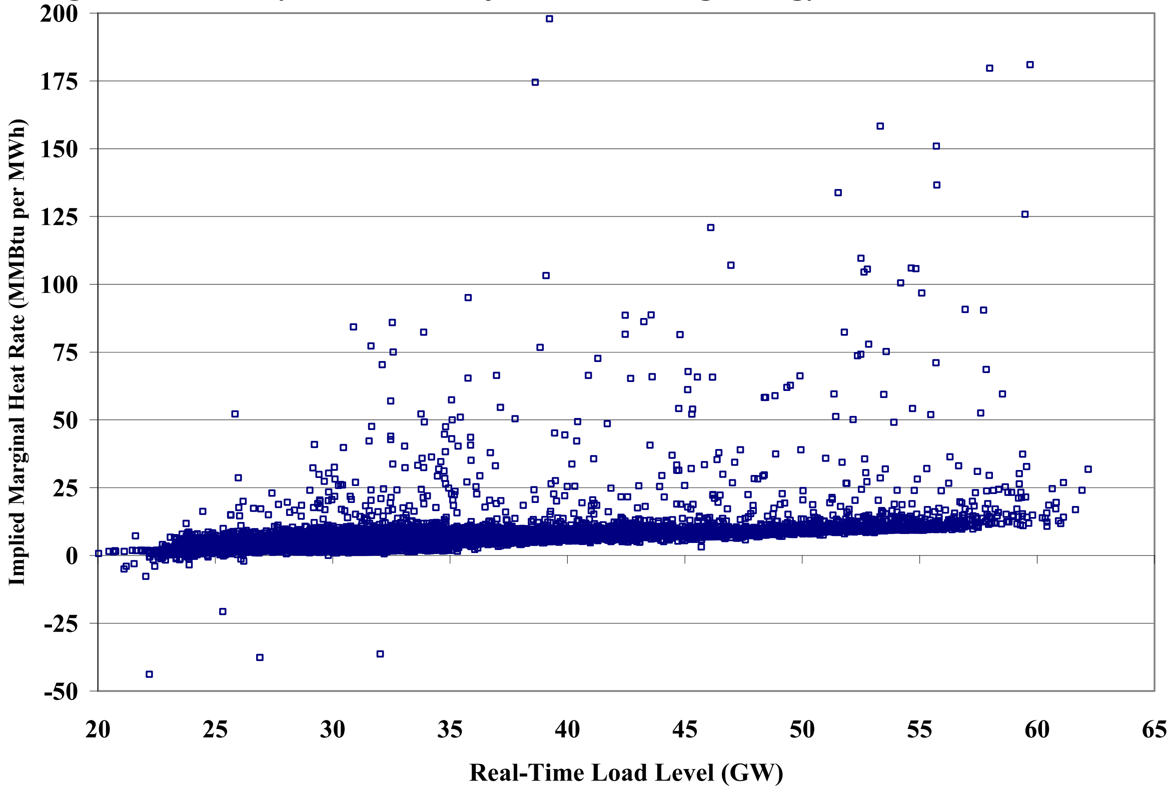
5. Determinants of Balancing Energy Prices

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



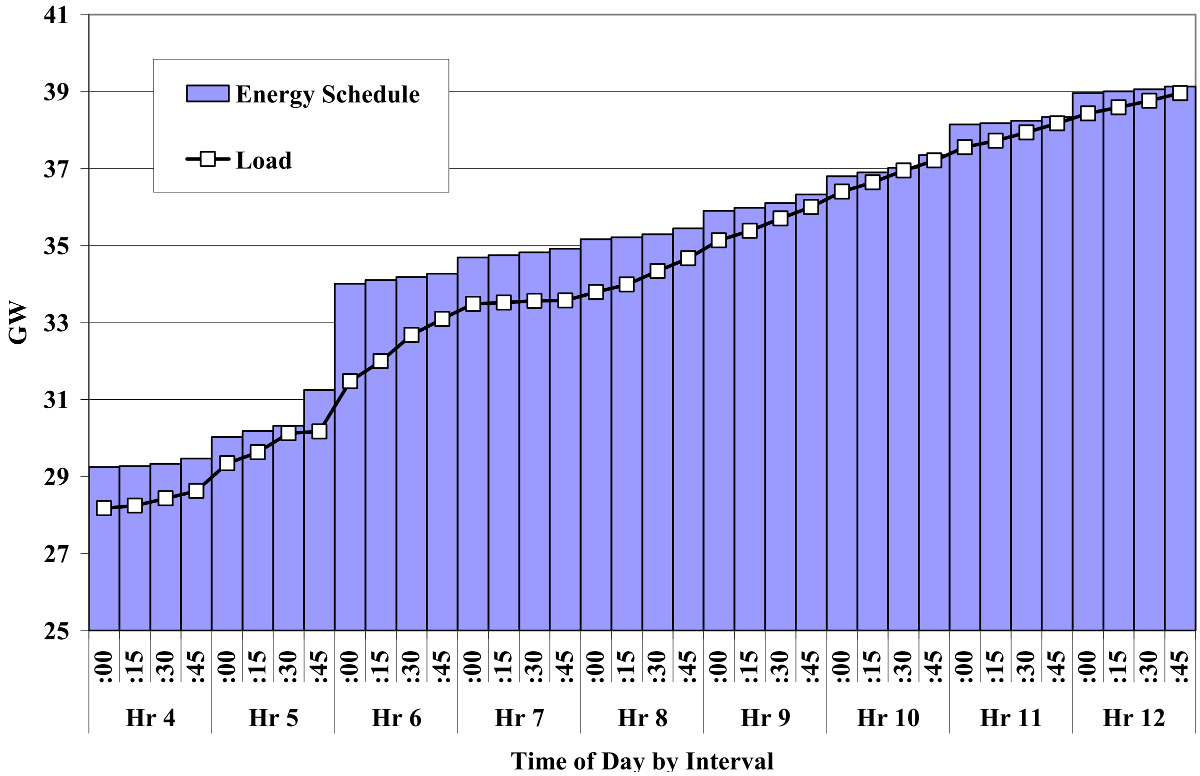
The figure indicates a positive correlation between real-time load and the clearing price in the balancing market. Although prices were generally higher at higher load levels, the analysis shown in Figure 13 indicates that the net volume of energy purchased in the balancing energy market is often a stronger 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 2008.

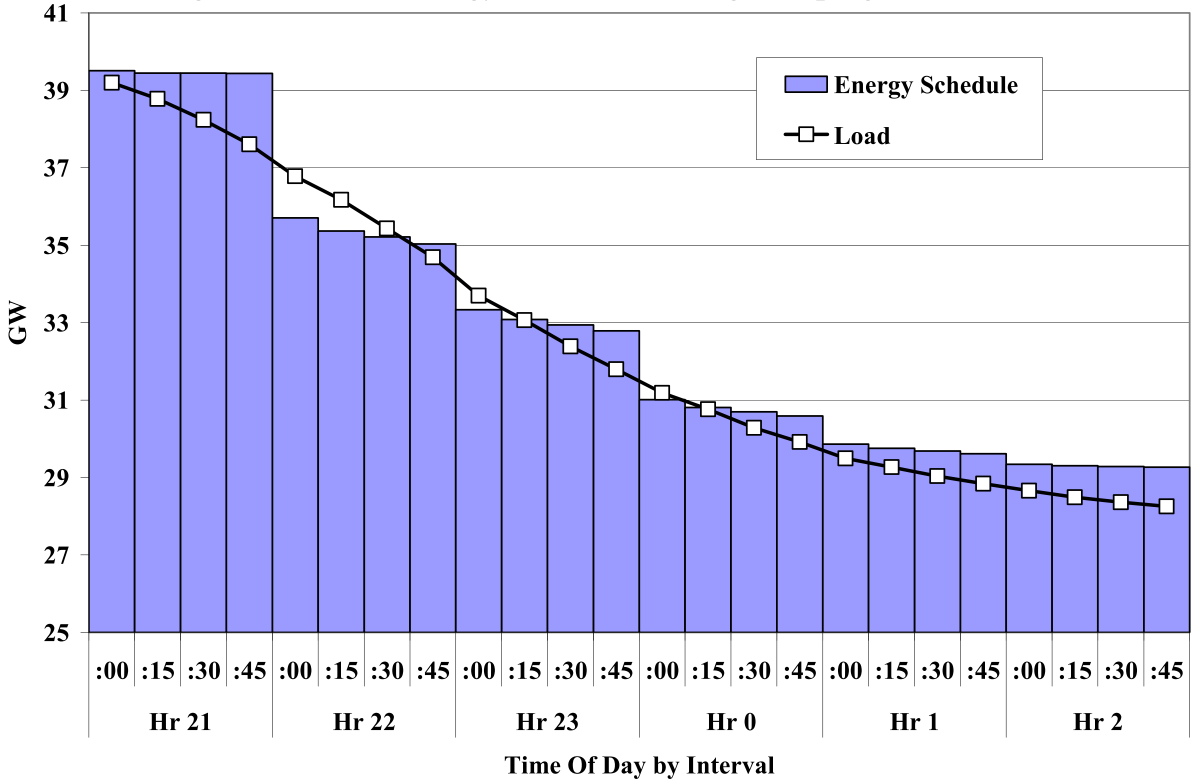
Figure 15: Final Energy Schedules during Ramping Up Hours



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 2008.

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

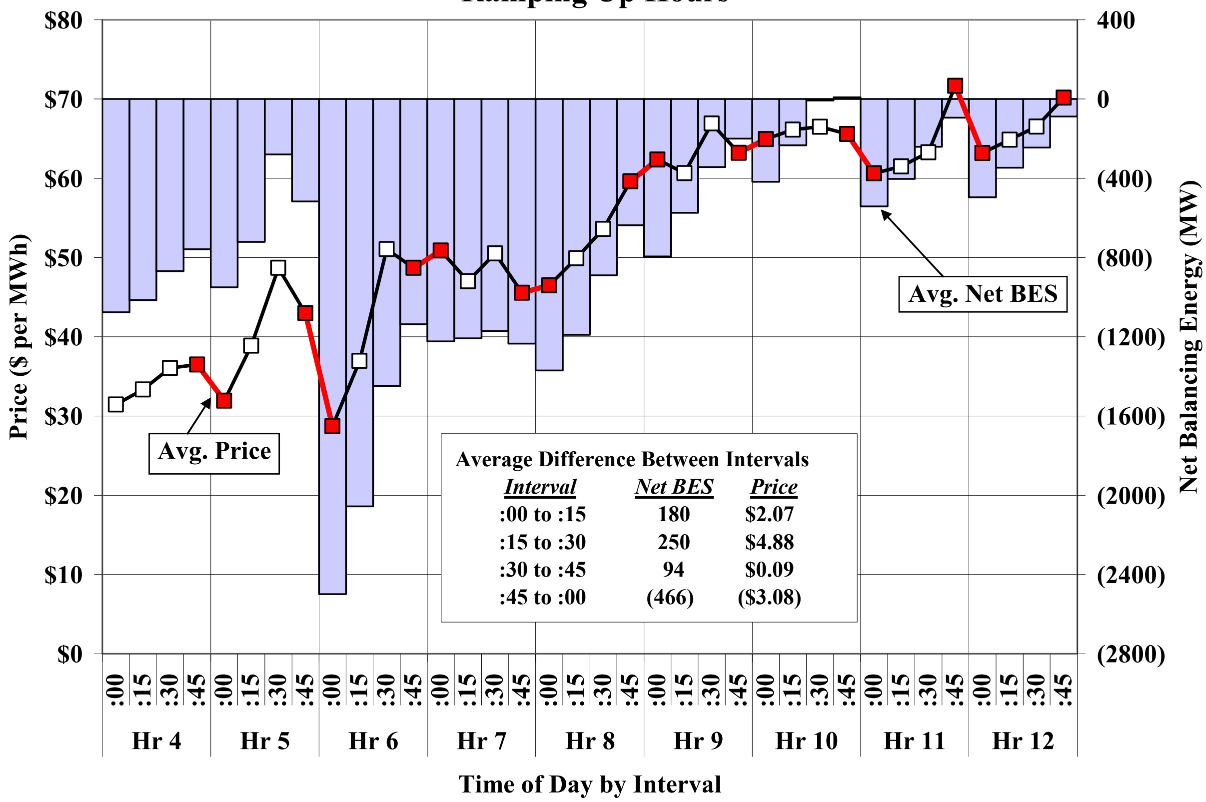


Figure 17 reveals two key aspects of the balancing energy market. First, as discussed above, balancing energy prices are highly correlated with balancing energy deployments. Second, with the exception of hour 7, there is a distinct pattern of increasing net balancing energy deployments during the hour. 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 prices and could result in erratic dispatch signals to the generators.

Figure 18: Balancing Energy Prices and Volumes Ramping Down Hours

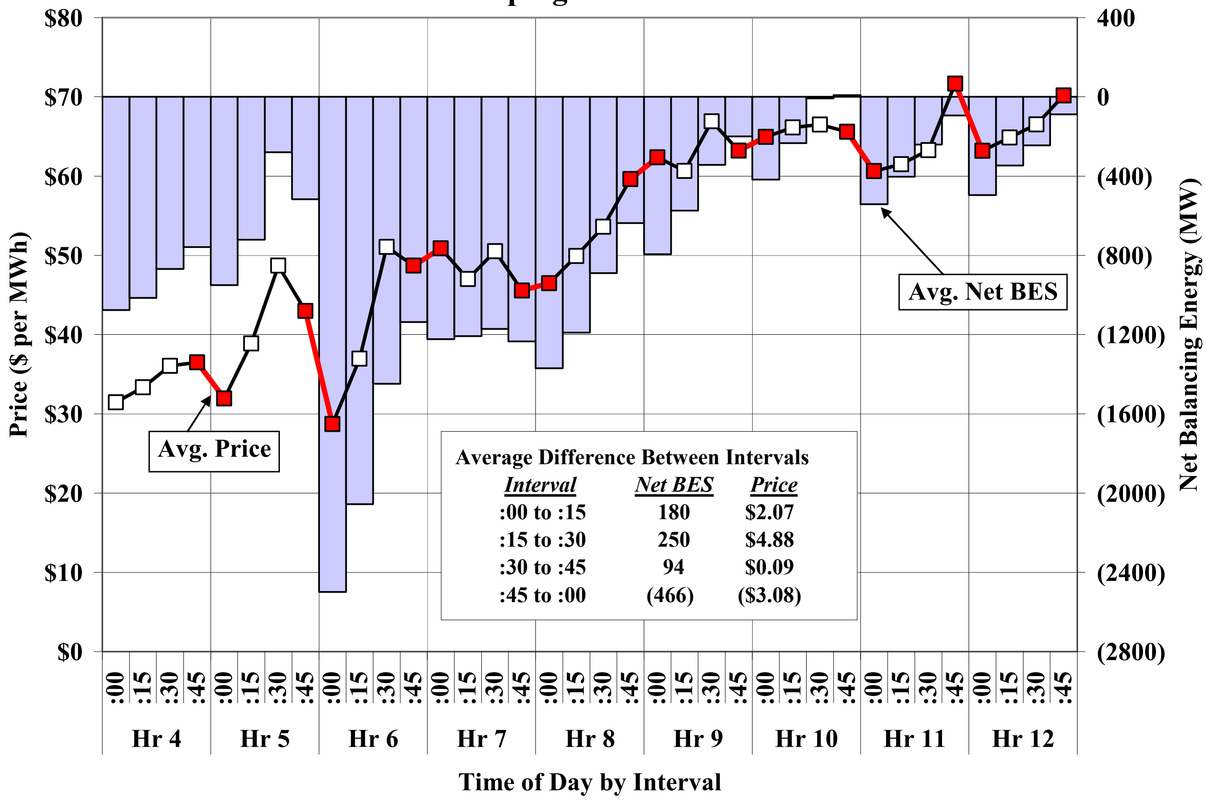


Figure 18 shows the same analysis for the ramping down hours. During ramping down hours, at the beginning of the hour, actual load tends to be higher than energy schedules, resulting in substantial balancing energy purchases, particularly in hour 22. At the end of the hour actual load tends to be lower 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 2008.

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

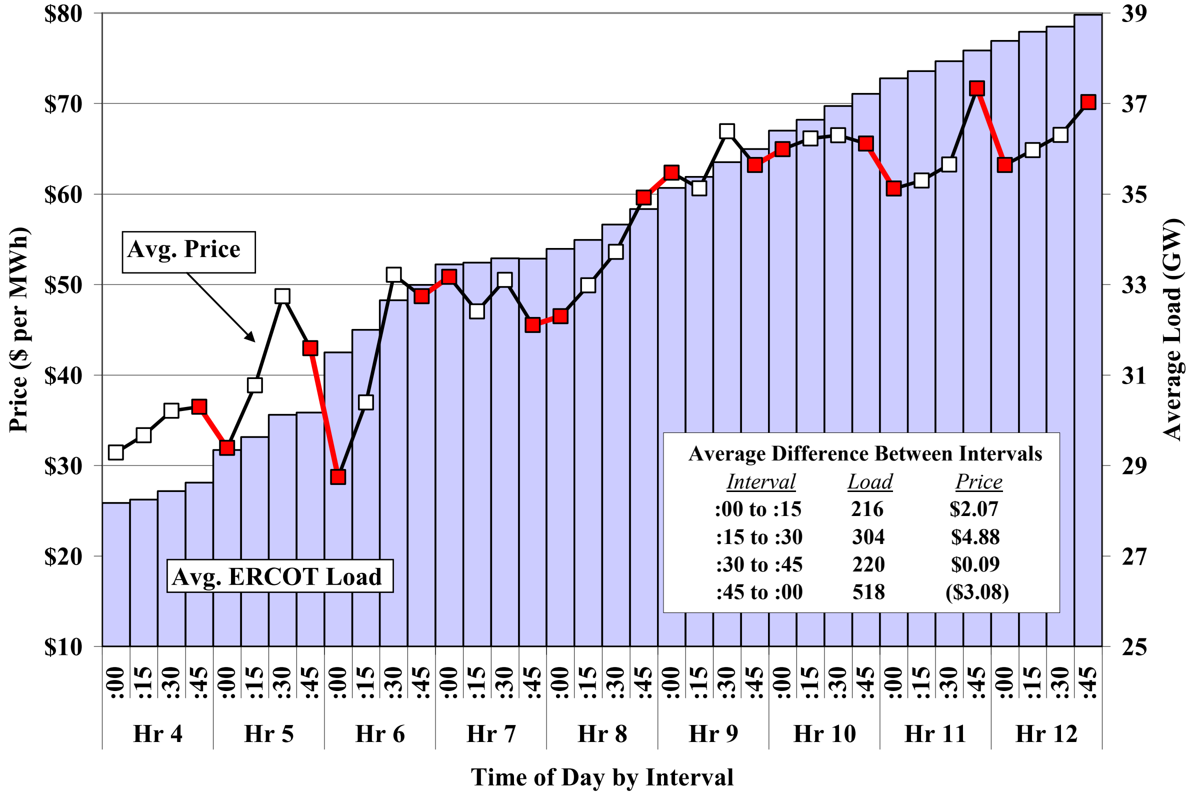


Figure 19 shows that, with the exception of hour 7, load steadily increases in every interval and prices generally move upward from an average of $31 per MWh at 4:00 AM to $70 per MWh at 12:45 PM. If actual load were the primary determinant of energy prices, the balancing energy prices would rise gradually as the actual load rises. However, Figure 19 shows this is not the case. In hours 5, 6, 11, and 12 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 -$3.08 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

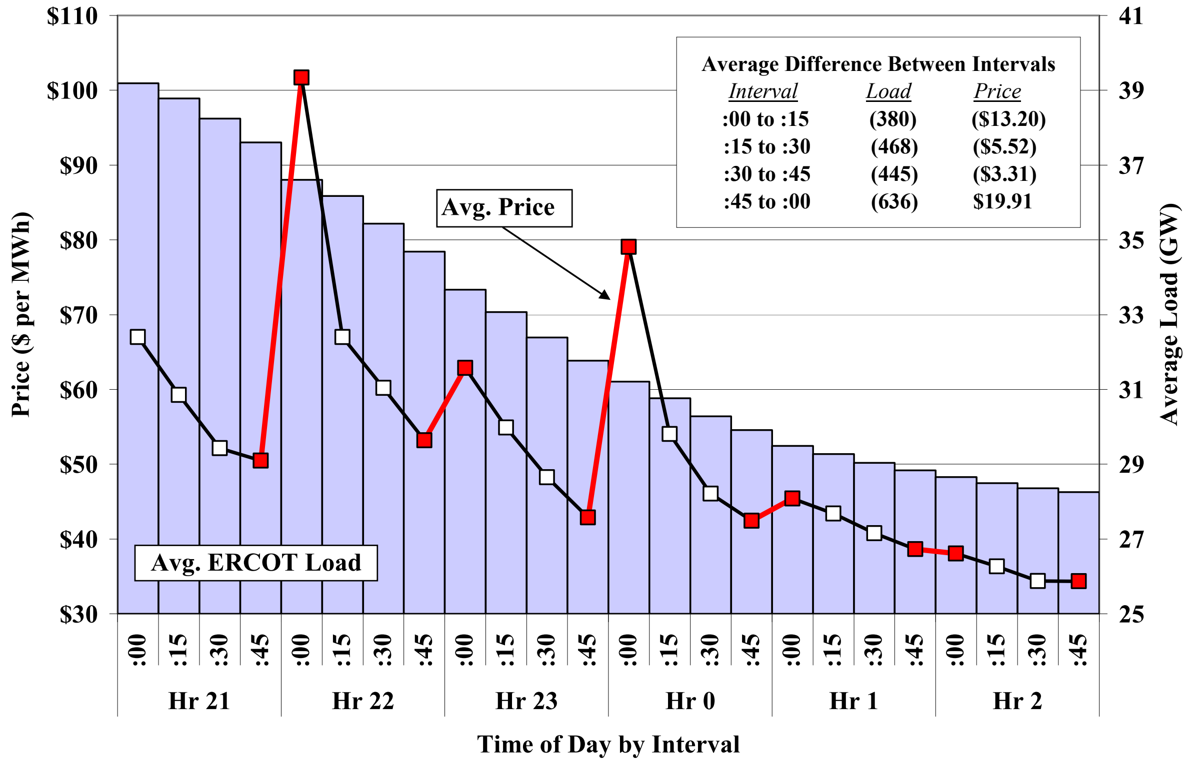


Figure 20 shows that while balancing energy prices decrease over these intervals, the pattern is similar to that exhibited in the ramping up hours. The balancing energy price decreases in each interval of the hour before rising substantially in the first interval of the following hour. The balancing energy price increases by an average of $19.91 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 and 2007.[10]

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

10 See 2006 and 2007 SOM Reports.

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, as more fully described in the 2005 SOM Report.[11] The efficiency implications associated with these issues continued in 2008 and will likely continue until the current zonal market design is replaced. However, ERCOT is implementing 14 minute ramp rates in 2009 that should 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.

B. Ancillary Services Market

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 2008.

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.

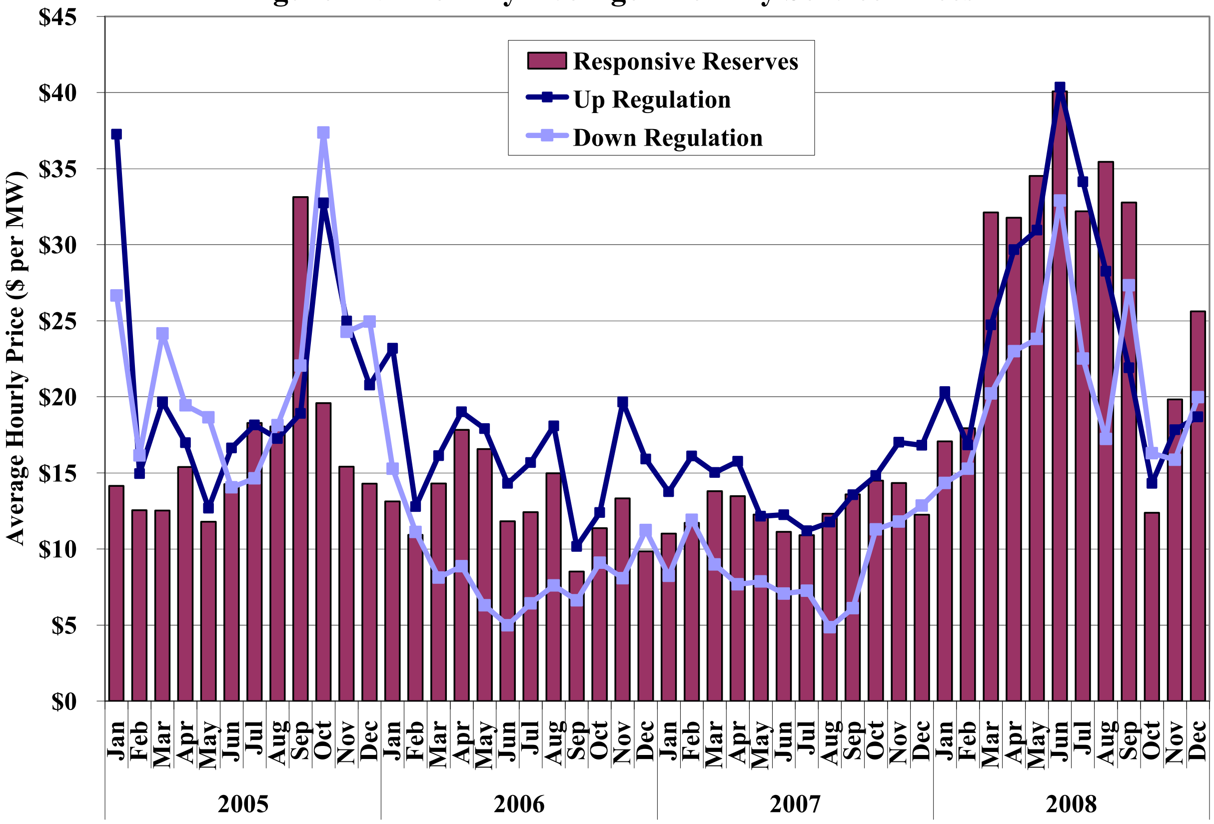
1. Reserves and Regulation Prices

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 2005 and

[11] 2005 SOM Report at 68-76.

2008. 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 after two years of relatively stability, 2008 experienced a significant increase in ancillary service capacity prices. These price movements can be primarily attributed to the variations in energy prices that occurred over the same timeframe. In addition to the effect of higher energy prices on ancillary service prices, ERCOT increased its procurement of responsive reserve quantities from January through August 2008 from the historical constant quantity of 2,300 MW to as high as 2,800 MW during peak hours in the summer. Also, the required quantity of non-spinning reserves when procured was increased in most of 2008, and non-spinning reserves were procured more frequently in 2008 than in 2007.

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. Exceptions to this pattern occurred in 2005 when down regulation prices averaged 4 percent higher than up regulation prices, and in the third quarter of 2008 when prices for up and down regulation services were at comparable levels.

Figure 21 also shows that the prices for up regulation generally exceeded prices for responsive reserves from 2005 to 2007. 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.

As discussed in Section III, significant transmission congestion materialized in April, May and June 2008 leading to significantly higher prices in the Houston and South Zones. These pricing outcomes had the effect of increasing the opportunity costs for providers of responsive reserve in these locations, thereby causing an upward shift in the supply curve for responsive reserve in these months.

Also discussed in Section III is the significant increase in West to North congestion in 2008 which led to over 1,100 hours of average negative prices in the West Zone. For providers of responsive reserves in the West Zone, exposure to negative prices significantly increases the cost of the provision of reserves because the resources must operate uneconomically at minimum load levels. Hence, in periods of expected high wind production, responsive reserve offers from suppliers in the West Zone would be expected to increase to reflect these economics risks.

A final factor affecting responsive reserve pricing outcomes in 2008 was the provision of responsive reserves by Loads acting as Resources ("LaaRs"). As described in more detail in Section II and shown in Figure 38, the quantity of LaaRs providing responsive reserves was moderately reduced in March through May, and experienced more significant reductions in September, part of October, and in November and December. The reduction in the provision of responsive reserves by LaaRs in these months resulted in a corresponding increase in the quantity of responsive reserve provided by generation resources, which are typically more expensive, thereby placing an upward pressure on responsive reserve prices.

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 shows such an analysis, comparing the average prices for responsive reserves and non-spinning reserves over the past five years in those hours when ERCOT procured non-spinning reserves. Non-spinning reserves were purchased in approximately 23, 20 and 14 percent of hours in 2005, 2006 and 2007, respectively, but increased to 51 percent of the hours in 2008. Part of the increased frequency of the procurement of non-spinning reserves in 2008 was associated with ERCOT's official change in procedures in November 2008 to procure non-spinning reserves 24-hours per day, although ERCOT has discretion in the decision to purchase of non-spinning based on its assessment of reliability risks and had been moving toward more frequent purchases of non-spinning reserves prior to November.

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

	2005	2006	2007	2008
Non-Spin Reserve Price	$25.10	$21.75	$6.07	$7.97
Responsive Reserve Price	$28.16	$25.55	$16.74	$36.39

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 in 2007 and 2008 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 2005 through 2008.

Figure 22: Ancillary Service Costs per MWh of Load

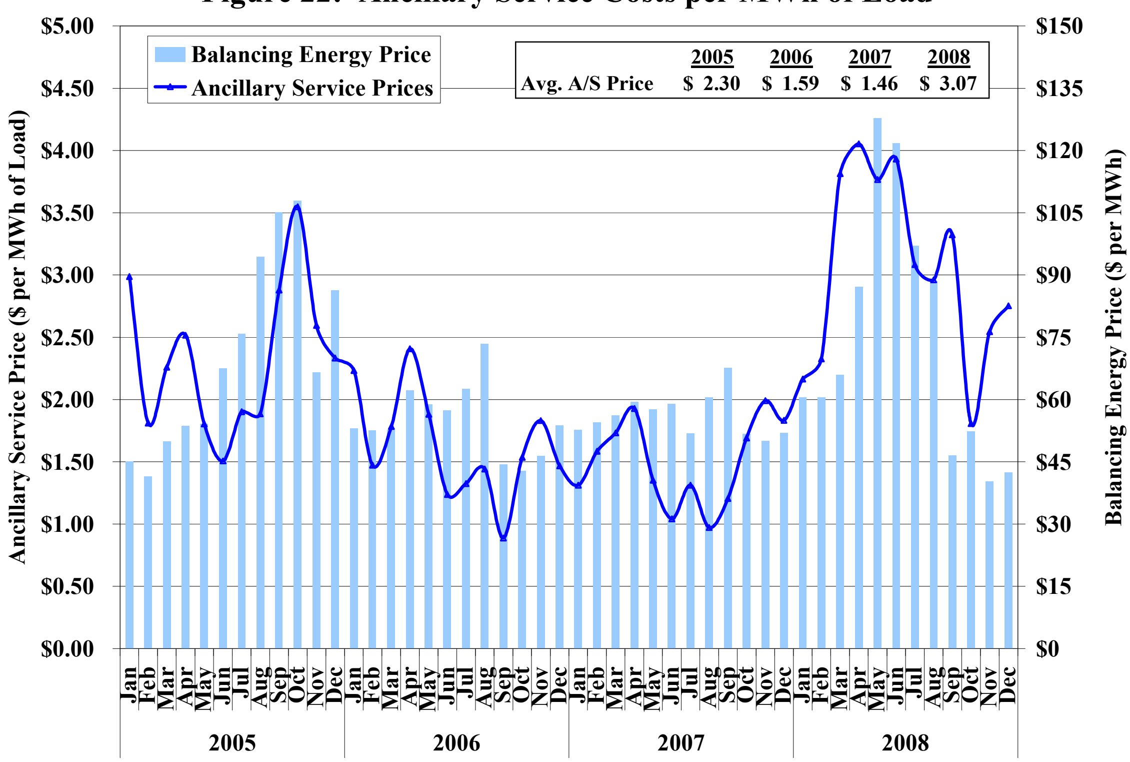


Figure 22 shows that total ancillary service costs are generally correlated with balancing energy price movements which, as previously discussed, are highly correlated with natural gas price movements. The average ancillary service cost per MWh of load increased to $3.07 per MWh in 2008 compared to $1.46 per MWh in 2007, an increase of more than 110 percent. However, while the all-in wholesale costs shown in Figure 2 increased by more than 38 percent in 2008 compared to 2007, ancillary service costs accounted for only 2.8 percent of the increase in all-in wholesale power costs in 2008 over 2007.

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 1,340 MW of regulation capability prior to the initial ramping period (beginning at 6 AM). The requirement then increases to more than 2,000 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

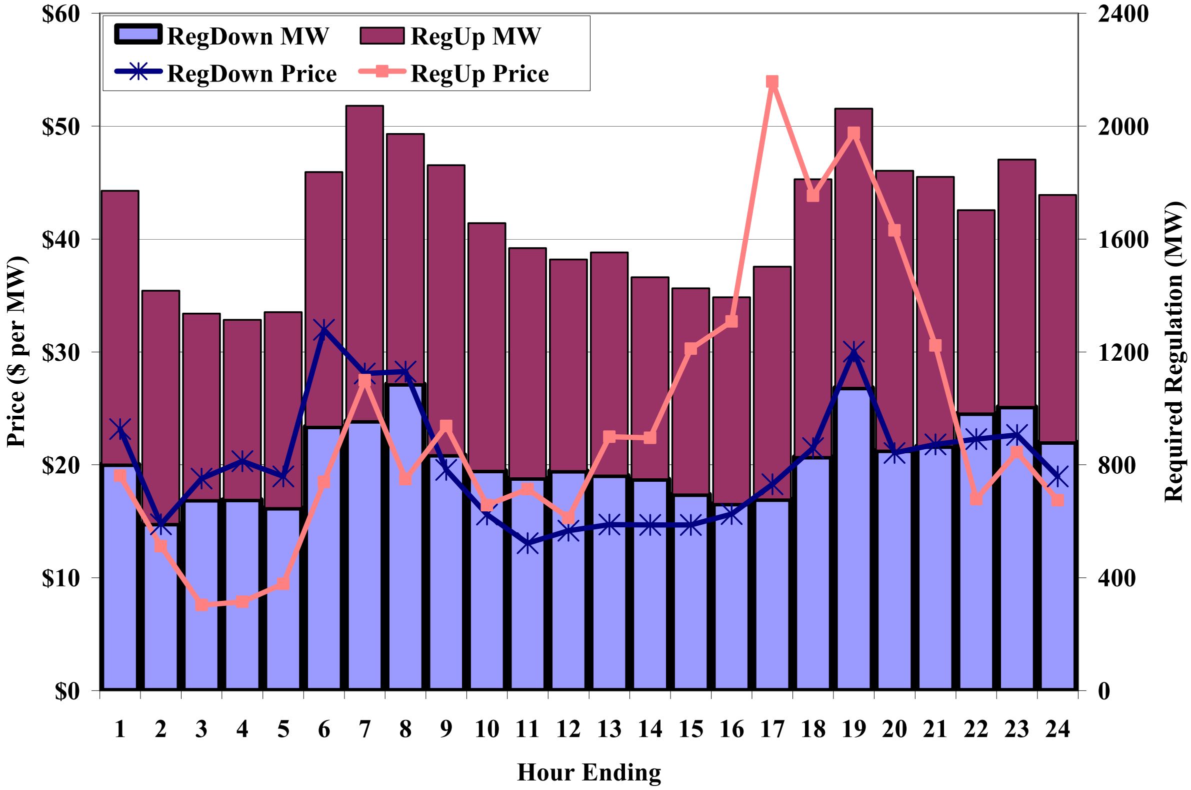


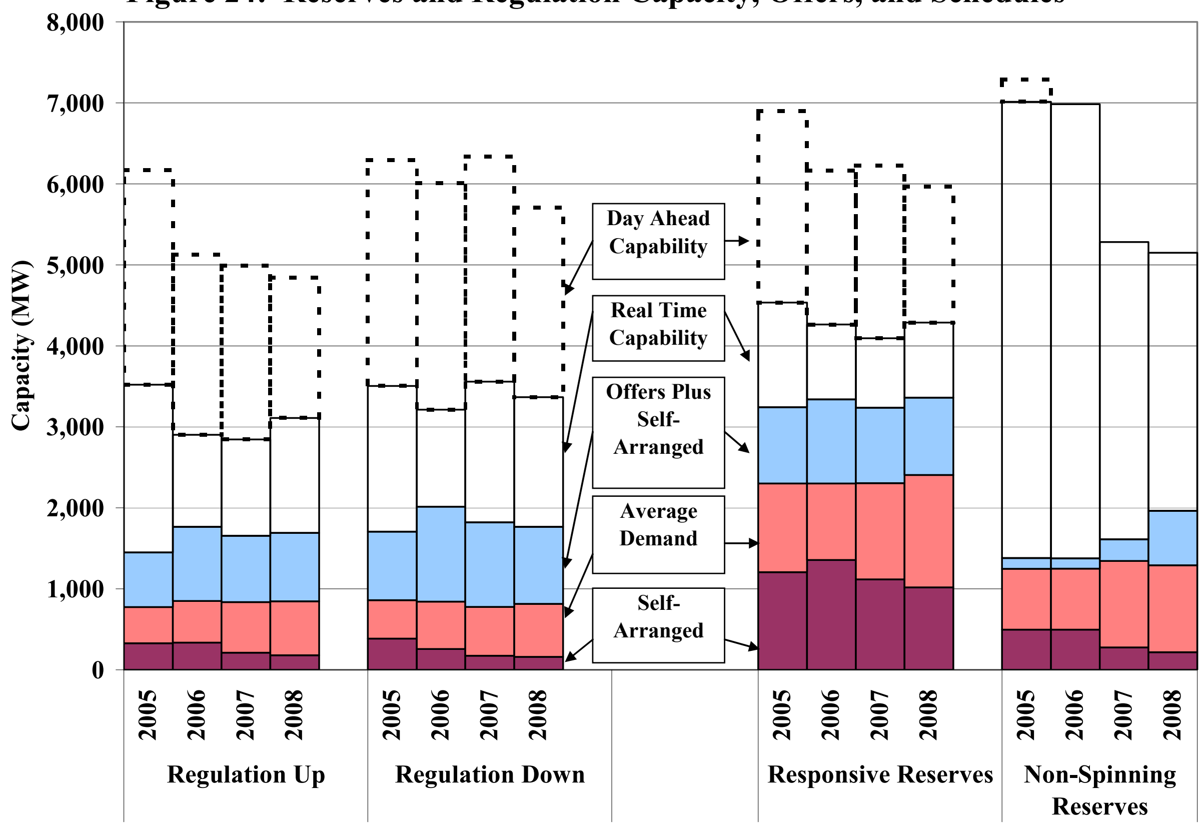
Figure 23 indicates that average regulation prices are generally correlated with the regulation quantity purchased and the typical load pattern in ERCOT. During non-ramping hours, such as overnight and late morning, up and down regulation prices are at their lowest levels. During the ramping hours in early morning average up and down regulation prices reached approximately $30 per MW. During evening ramping hours, down regulation prices also reached $30 per MW, while up regulation prices topped out at almost $55 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.

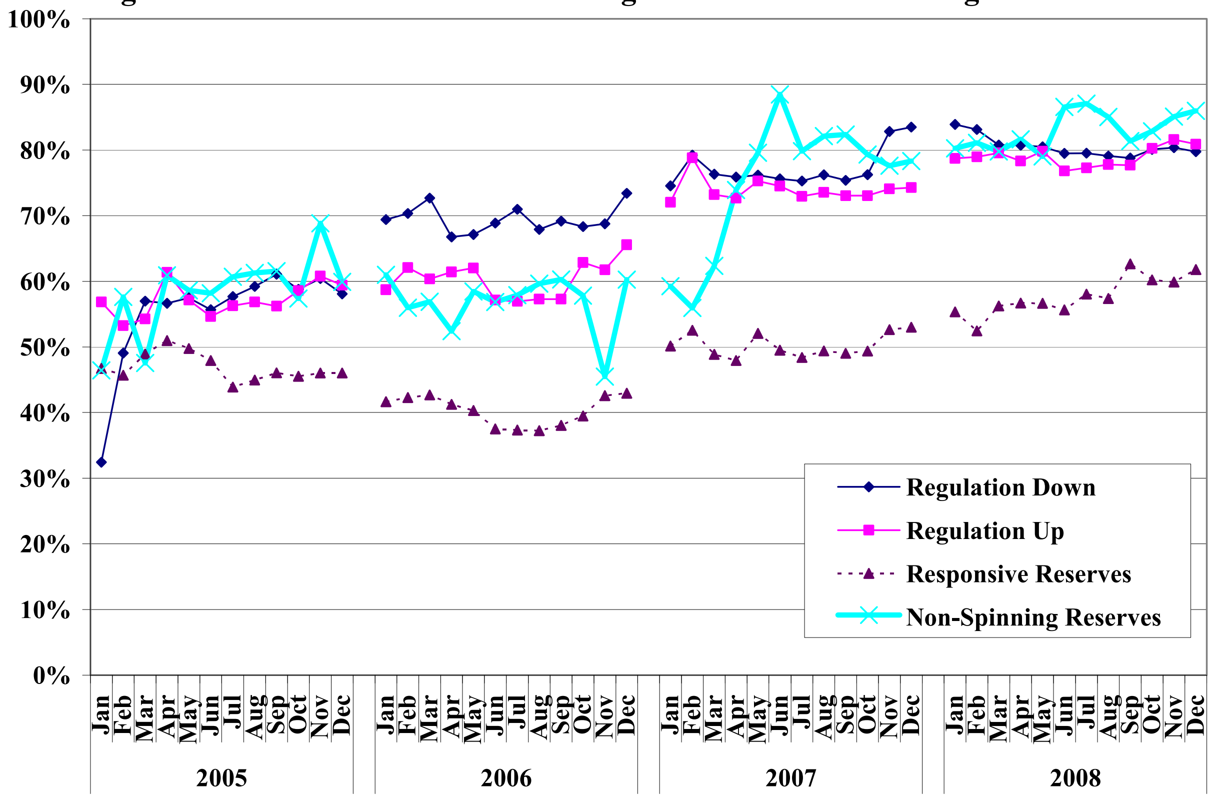
For non-spinning reserves, Figure 24 includes the capability of units that QSEs indicate are able to ramp-up in thirty minutes and able to start-up on short notice. The total capability shown in this figure does not account for capacity of online resources. However, it should be noted that any on-line 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 reserve in 2006, 2007 and 2008, in which about 54 percent, 52 percent and 56 percent, respectively, of available responsive reserve capacity was offered, less than one-half of each type of ancillary services capability was offered during the year from 2005 to 2008. 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

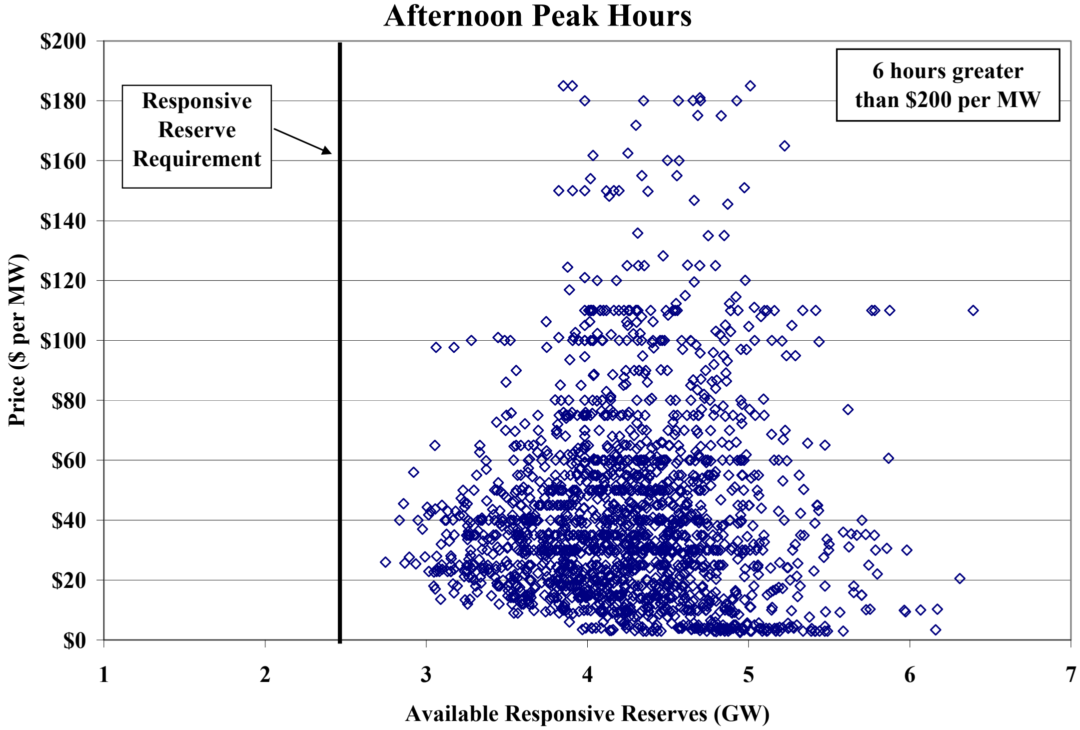


As market participants have gained more experience with the ERCOT markets, larger portions of the available reserves and regulation capability have been offered into the market, thereby increasing the market's liquidity. Nevertheless, Figure 25 shows that a fair share of these services is still self-supplied,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

The final analyses in this section evaluate the prices prevailing in the responsive reserve and the non-spinning reserve markets in 2008. Prices in the ERCOT responsive reserve market are significantly higher than in other markets that co-optimize the procurement and dispatch of energy and responsive reserves. 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 approximately 2,450 MW in 2008 to show the amount of the surplus in each hour.

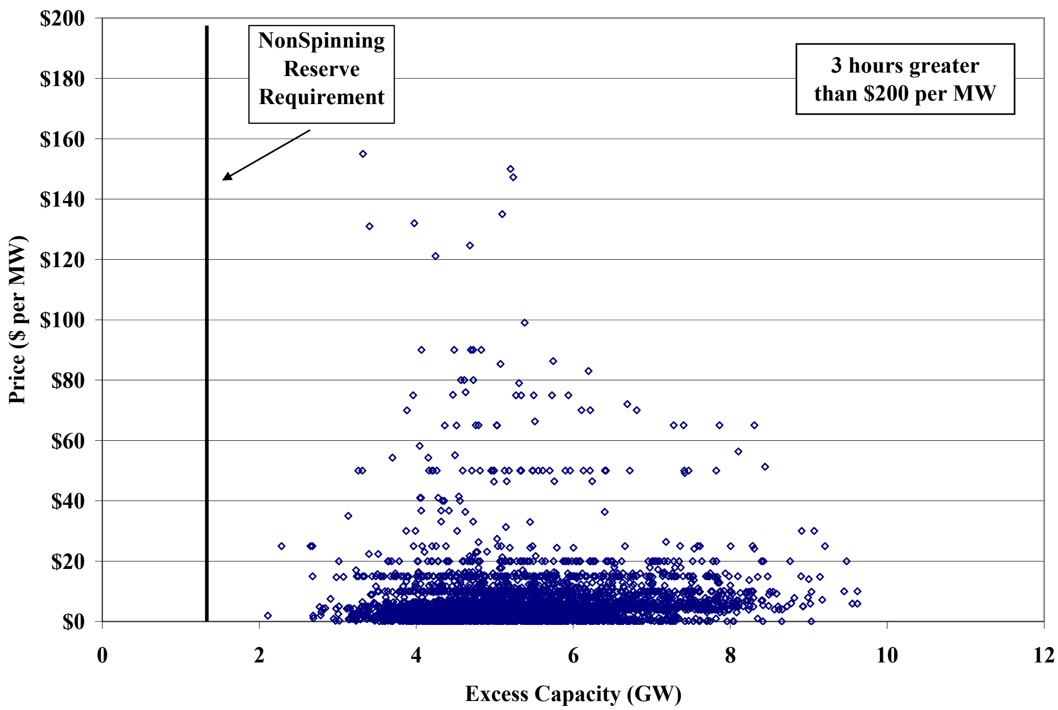
Figure 26: Hourly Responsive Reserves Capability vs. Market Clearing Price Afternoon Peak Hours



This figure indicates a somewhat random pattern of responsive reserve prices in relation to the hourly available capability. In a well functioning-market for responsive reserves, we would expect excess capacity to be negatively correlated with the clearing prices. 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.

In 2008 non-spinning reserves were purchased on a day-ahead basis primarily during defined times of extreme or unpredictable demand. Non-spinning reserves are resources that can be deployed within 30 minutes. Thus, off-line quick-start units can provide non-spinning reserves. In addition, any resource that plans to be on-line with capacity not already scheduled for energy, regulation, or responsive reserves can also provide non-spinning reserves. Figure 27 shows the relationship between excess available non-spinning reserves capability and the market clearing price in the non-spinning reserves auction for all the hours in 2008.

Figure 27: Hourly Non-Spinning Reserves Capability vs. Market Clearing Price
All Hours



Like the previous analysis of responsive reserves, the results shown in Figure 27 indicate a somewhat random pattern of prices compared to excess capacity capable of providing non-spinning reserves. Again, the lack of co-optimized markets for energy and reserves may be a primary contributor to the high prices for non-spinning reserves when there are large quantities of excess capacity available.

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 2008 and the existing generating capacity available to satisfy the load and operating reserve requirements.

A. ERCOT Loads in 2008

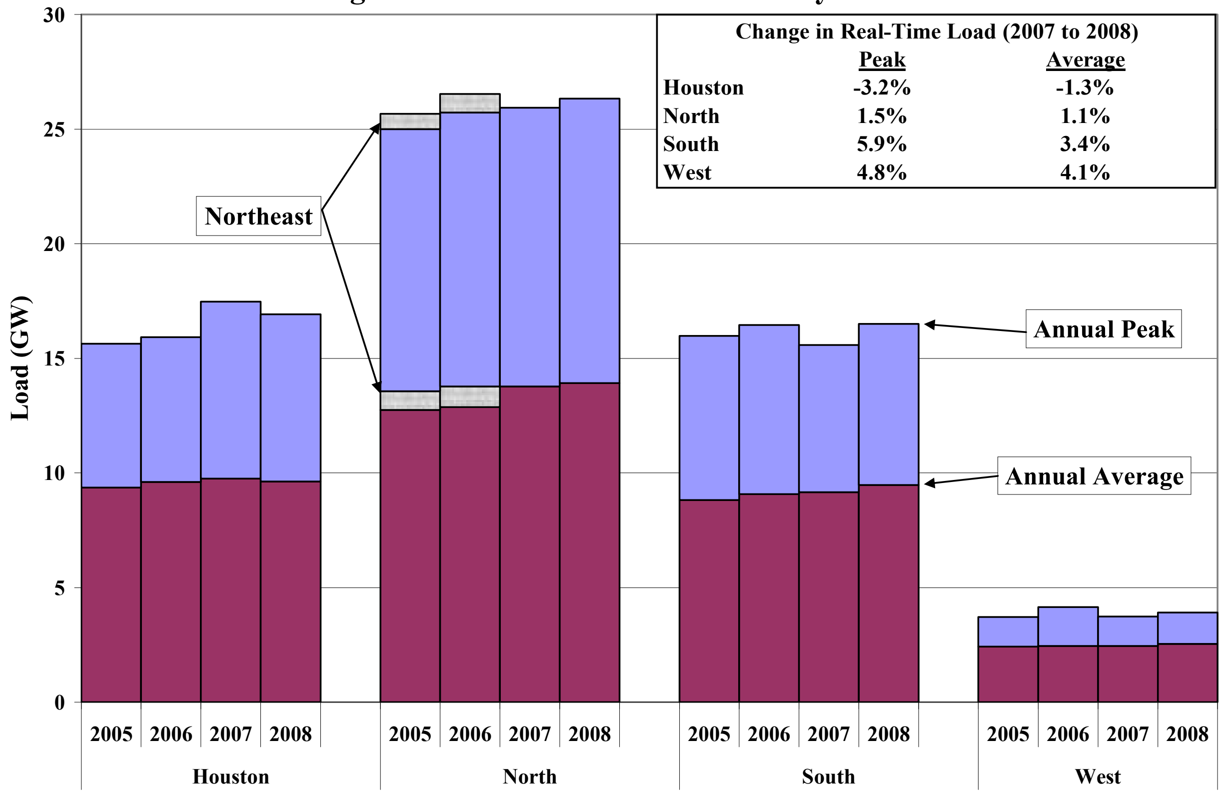
There are two important dimensions of load that should be evaluated separately. First, the changes in overall load levels from year to year can be shown by tracking the changes in average load levels. This metric will tend to capture changes in load over a large portion of the hours during the year. 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 2008 are examined in this subsection and summarized in Figure 28.

This figure shows peak load and average load in each of the ERCOT zones from 2005 to 2008. 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 40 percent of the total ERCOT load);[12] the South and Houston Zones are comparable (with about 26 percent and 28 percent, respectively) while the West Zone is the smallest (with about 7 percent of the total ERCOT load). Figure 28 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

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

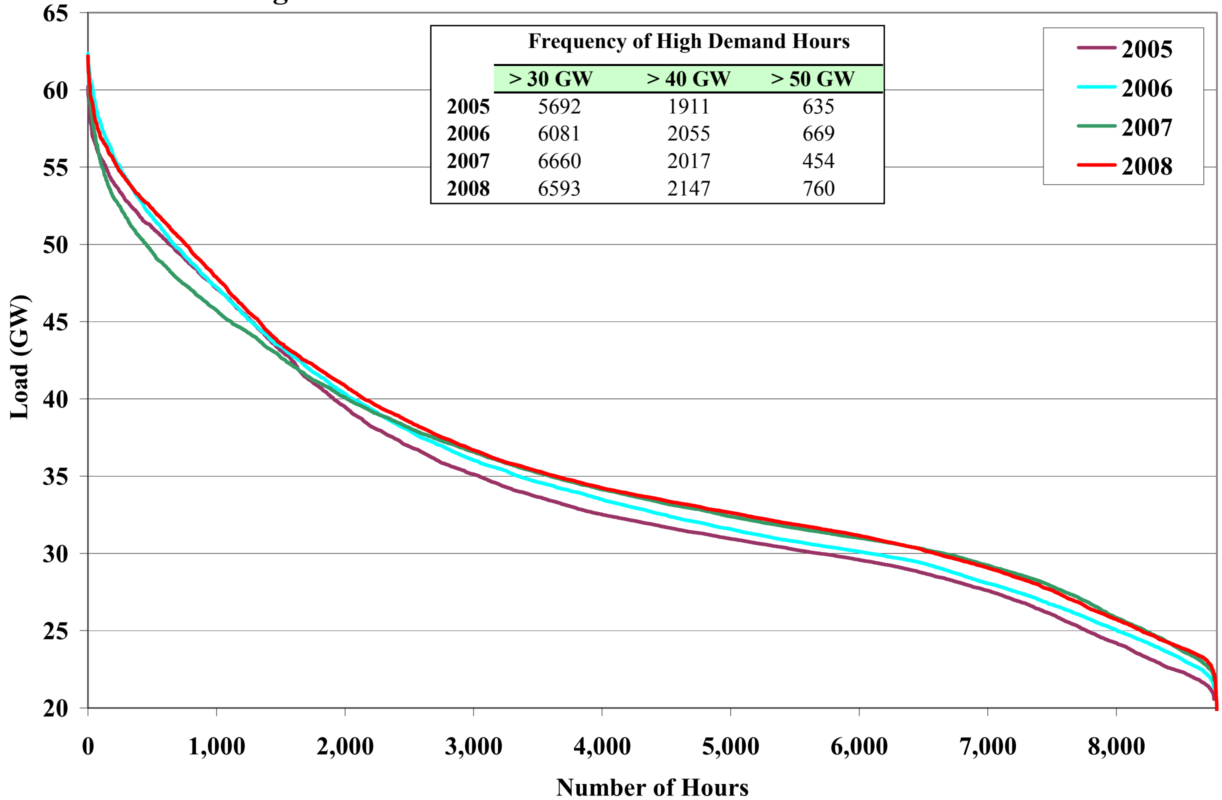
in different hours for different zones. As a result, the sum of the non-coincident peaks for the zones was greater than the annual ERCOT peak load.

Figure 28: Annual Load Statistics by Zone



To provide a more detailed analysis of load at the hourly level, Figure 29 compares load duration curves for each year from 2005 to 2008. 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.

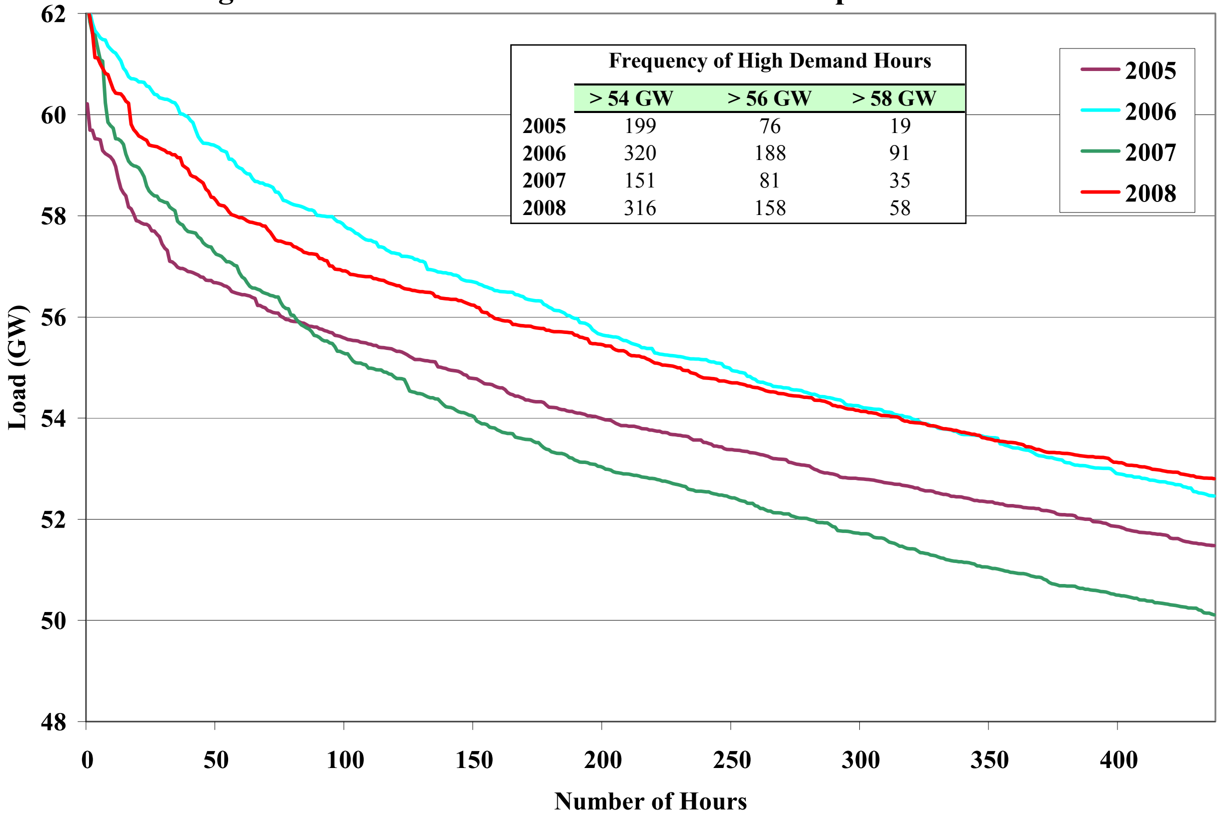
Figure 29: ERCOT Load Duration Curve – All Hours



As shown in Figure 29, the load duration curve for 2008 is comparable to 2007 at load levels less than 40 GW. Load increased about 1.5 percent from 2007 to 2008. In 2008, more than 8 percent of the hours were high load (greater than 50 GW) compared to 5 percent of the hours in 2007.

To better show the differences in the highest-demand periods between years, Figure 30 shows the load duration curve for the five percent of hours with the highest loads. It shows that while load increased in each year from 2005 to 2008, the frequency of high-demand hours in 2008 also increased compared with year 2007. Load exceeded 58 GW in 58 hours in 2008 and 35 hours in 2007. 2007 and 2008 both had higher average loads than 2006, although the number of hours that the load exceeded 58 GW in 2006 was significantly higher (91 hours) than 2007 or 2008.

Figure 30: ERCOT Load Duration Curve – Top 5% of Hours



	> 54 GW	> 56 GW	> 58 GW
2005	199	76	19
2006	320	188	91
2007	151	81	35
2008	316	158	58

This figure also shows that the peak load in each year was roughly 17 to 24 percent greater than the load at the 95th percentile of hourly load. For instance, in 2008, the peak load value was over 62 GW while the 95th percentile was about 53 GW. This is typical of, and even somewhat flatter than, the load patterns in most electricity markets. These load characteristics imply that a substantial amount of capacity – as much as 12 GW – is needed to supply energy in less than 5 percent of the hours. This load pattern serves to emphasize the importance of efficient pricing during peak demand conditions to send accurate economic signals for the investment in and retention of these resources.

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 2008. 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

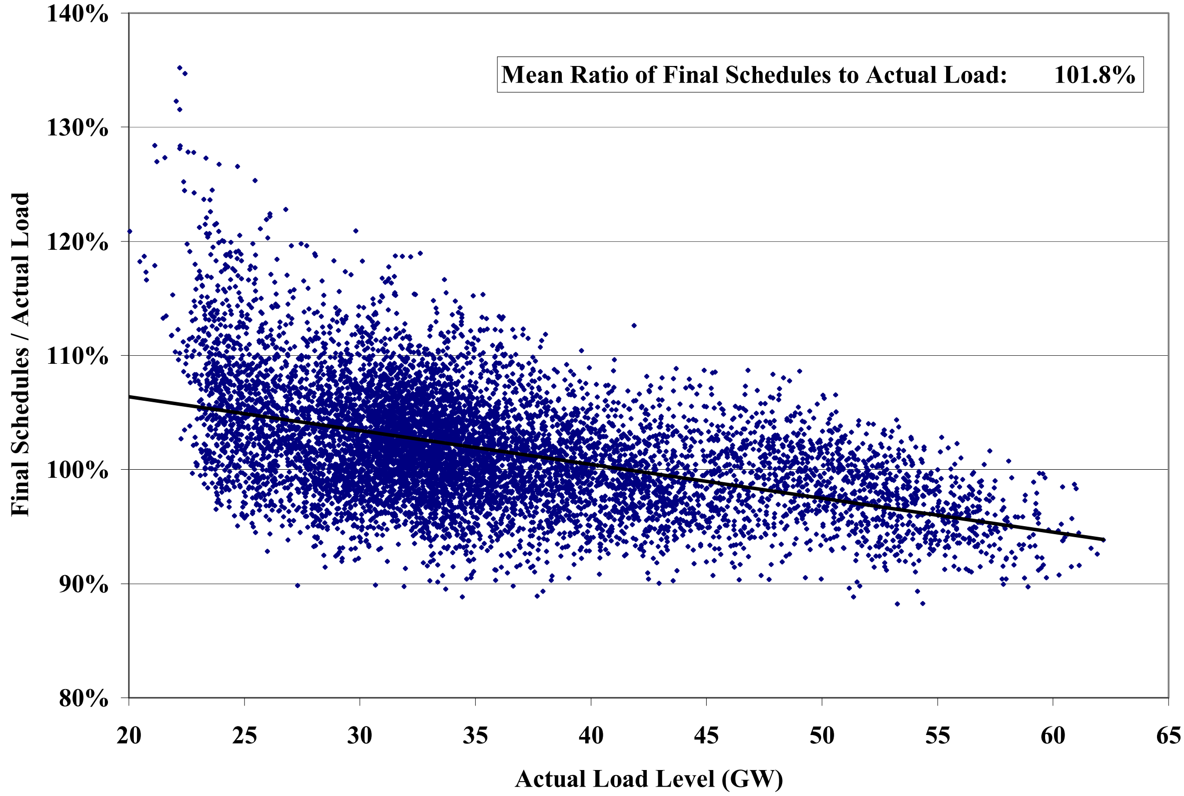


Figure 31 shows that final load schedules generally come very close to actual load in the aggregate, as indicated by an average ratio of the final load schedules to actual load of 101.8 percent. However, the figure also includes a trend line indicating that the ratio of final load schedules to actual load tends to decrease as load rises. In particular, the ratio given by the trend line is above 100 percent for loads under 40 GW and declines to 97 percent at higher load levels. The overall pattern shown in the figure above is similar to 2007, 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 suggests 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

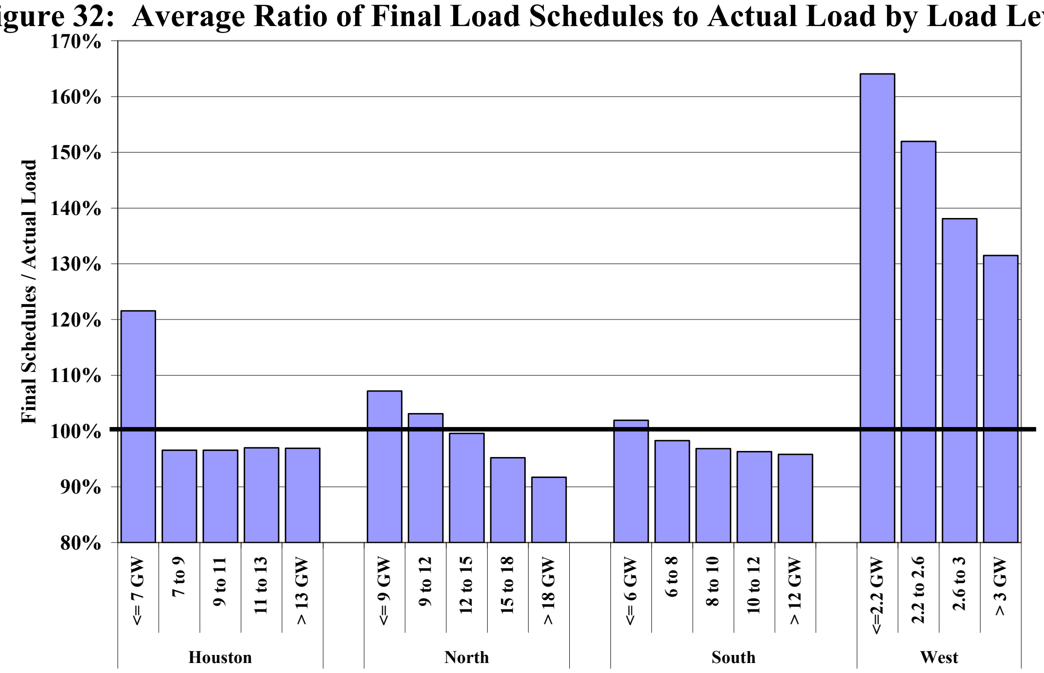


Figure 32 shows the ratio of final load schedules to actual load evaluated at five different load levels for each of the ERCOT zones. Figure 32 shows that:

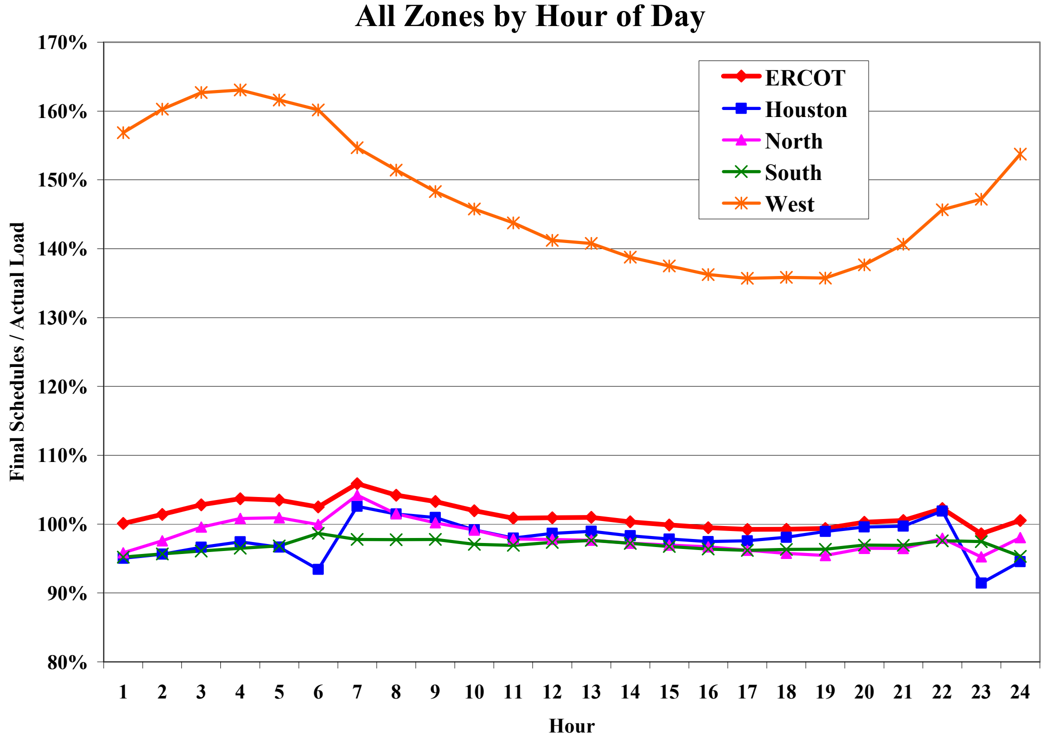
- The final schedule quantity decreases in three of the four zones as actual load increases.
- The West Zone is generally over-scheduled, although the ratio declines as load increases.

- The Houston and South Zones are under-scheduled at most load levels.
- The North Zone is under-scheduled at the highest load levels.
- The Houston Zone was significantly over-scheduled at the lowest load levels, which is a result of the significant reduction in loads in September 2008 due to Hurricane Ike.

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-of-day for each of the four zones in ERCOT as well as for ERCOT as a whole.

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



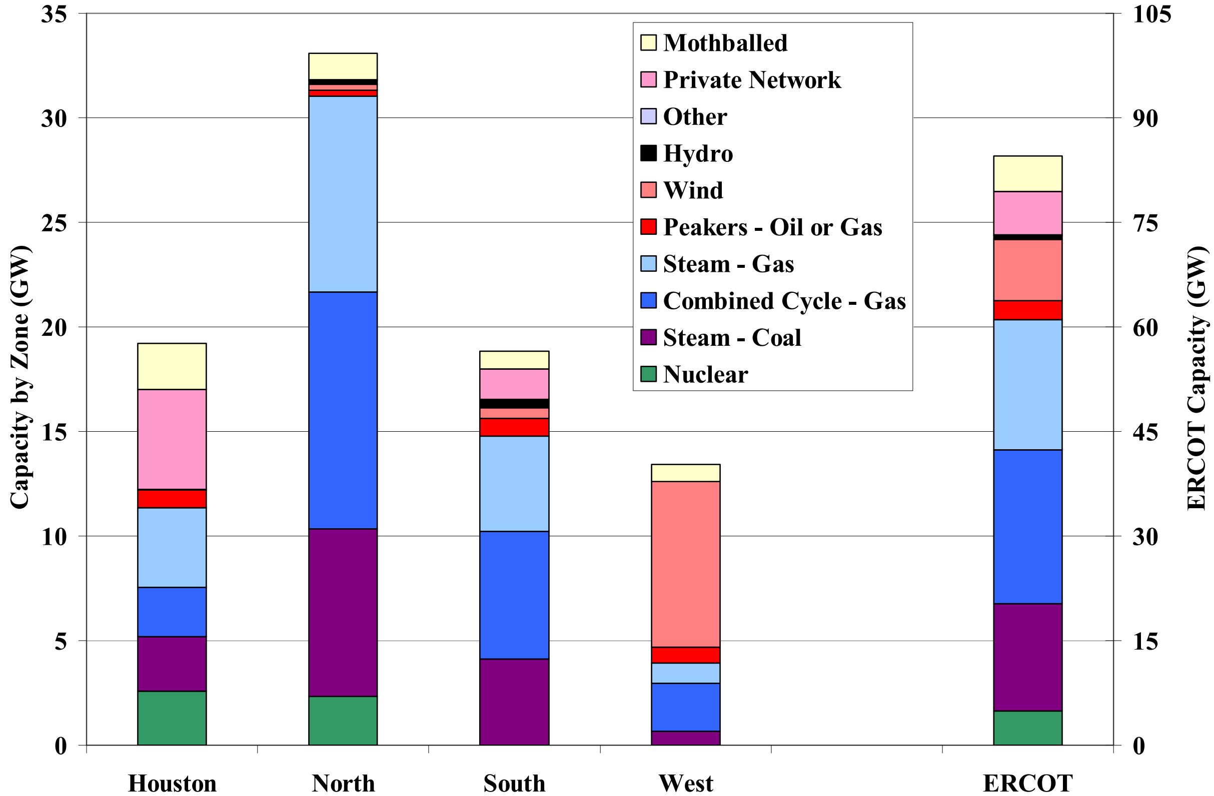
This figure shows that on an ERCOT-wide basis, final schedules are close to actual load in most of the hours during the day. The ERCOT-wide ratio increases to 106 percent at hour ending 7 and decreases to 99 percent and at hour ending 23. In the other hours, the ERCOT-wide ratio ranges between 95 and 101 percent, excluding the West Zone. The higher ratio in the West Zone is most likely explained by the increases in wind capacity in 2008 where the wind is scheduled in all hours that the resource is available, regardless of actual load levels in that zone.

Hour ending 7 and hour ending 22 represent start and end points of the 16 hour block of peak hours commonly used in bilateral contracts. Hence, a logical explanation for the patterns shown in Figure 33 is that participants tend to submit schedules consistent with their bilateral transaction positions. This is not irrational if the market participants also submit balancing energy offers to optimize the energy that is actually deployed. In addition, market participants bear additional price risk in ramping hours (as shown in the prior section), explaining their propensity to schedule a larger portion of their needs during these periods.

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 South Zones, the mix of generating capacity is relatively uniform in ERCOT. Figure 34 shows the installed generating capacity by type in each of the ERCOT zones.

Figure 34: 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 60 percent of generation capacity in ERCOT as a whole and almost 70 percent in the South Zone. Approximately one-half of this natural gas-fired capacity represents relatively new combined-cycle units than 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.

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 20.3 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 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.

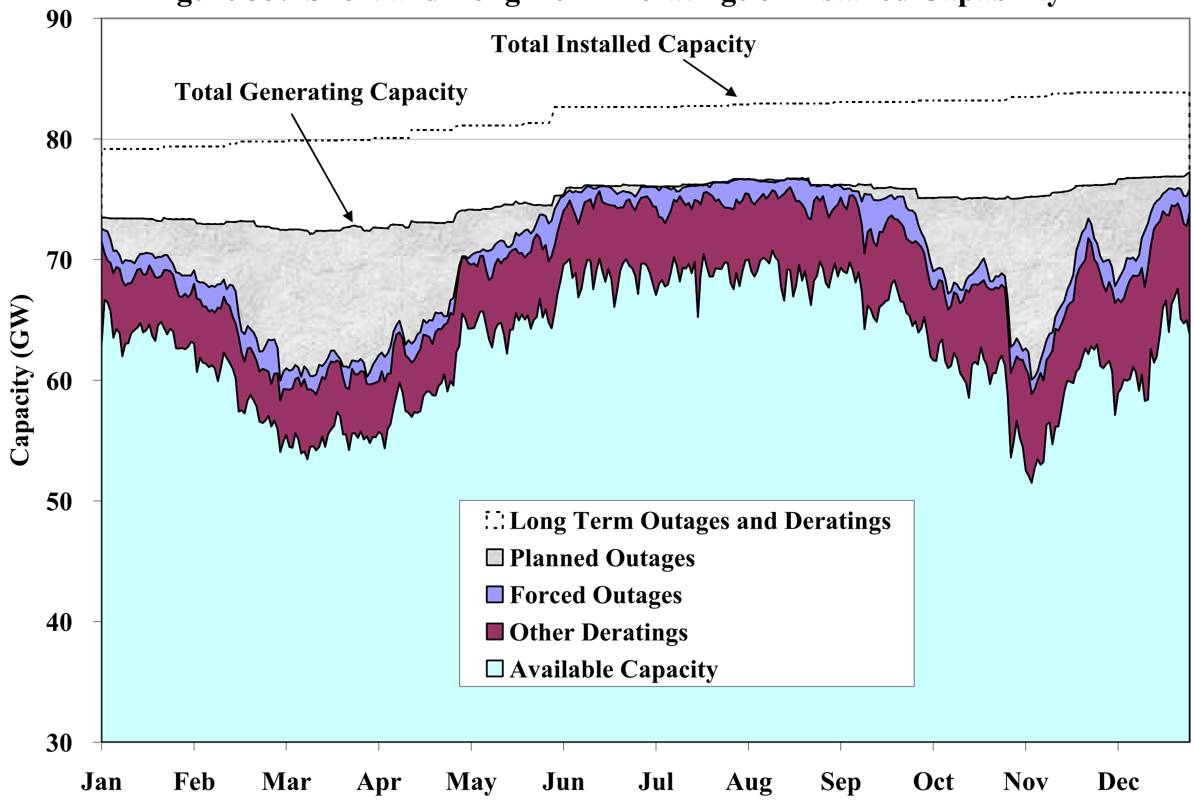
The distribution of capacity among the ERCOT zones is similar to the distribution of demand with the exception of the large amount of wind capacity in the West Zone. The North Zone accounts for approximately 39 percent of capacity, the South Zone 22 percent, the Houston Zone 23 percent, and the West Zone 16 percent. The Houston is typically an importer of 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 44 percent of capacity, the South Zone 25 percent, the Houston Zone 24 percent, and the West Zone 8 percent.

1. Generation Outages and Deratings

Figure 34 in the prior subsection shows that installed capacity is approximately 84 GW including mothballed units and all wind capacity, and approximately 73 GW excluding mothballed capacity and including only 8.7 percent of wind capacity. Hence, the installed capacity is well in excess of the capacity required to meet annual peak load plus ancillary services requirements of 65 to 66 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 35 shows a breakdown of total installed capability for ERCOT on a daily basis during 2008. 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 35: 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 35 shows that long-term outages and other deratings fluctuated between 9 and 18 GW. These outages and deratings reduce the effective resource margins in ERCOT from the levels reported above. Most 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);
- Output ranges on available generating resources that are not capable of producing up to the full installed capability level (*e.g.*, wind resources); or

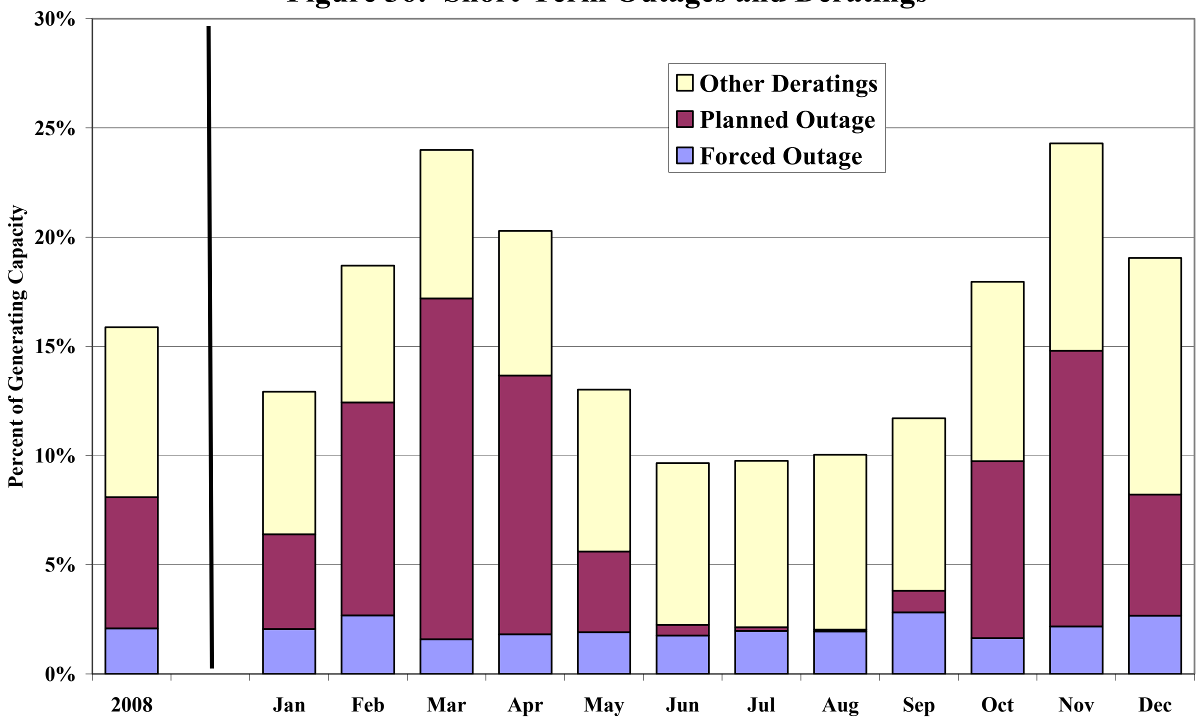
- Resources out-of-service for extended periods due to maintenance requirements.

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.

The next analysis focuses specifically on the short-term forced outages and other short-term deratings. Figure 36 shows the average magnitude of the outages and deratings lasting less than 60 days for the year and for each month during 2008.

Figure 36: Short-Term Outages and Deratings*



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

Figure 36 shows that total short-term deratings and outages were as large as 24 percent of installed capacity in the spring and fall, and dropping below 10 percent for the summer. Most of this fluctuation was due to anticipated planned outages, which ranged as high as 8 to 15 percent of installed capacity during March, April, October, and November. Short-term forced outages occurred more randomly, as would be expected, ranging between 1.6 percent and 2.6 percent of total capacity on a monthly average basis during 2008. These rates are relatively low in comparison to other operating markets, which can be attributed to a number of factors described below.

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 36 can be expected to be lower than equivalent forced outage rates of other markets. Second, we were not confident that the forced outage logs received from ERCOT included all forced outages that actually occurred.

The largest category of short-term deratings was the "other deratings", which 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 7 percent on average during the summer in 2008 and as high as 10 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[13] units minus

13 For the purposes of this analysis, "quick-start" includes simple cycle gas turbines that are qualified to

the demand for energy, responsive reserve, up regulation and non-spinning reserve provided from online capacity or quick-start units. To evaluate the commitment of resources in ERCOT, Figure 37 plots the excess capacity in ERCOT during 2008. The figure shows the excess capacity in only the peak hour of each weekday because largest amount of additional generation commitment usually occurs at the peak hour. Hence, one would expect larger quantities of excess capacity in other hours.

Figure 37: Excess On-Line and Quick Start Capacity During Weekday Daily Peaks

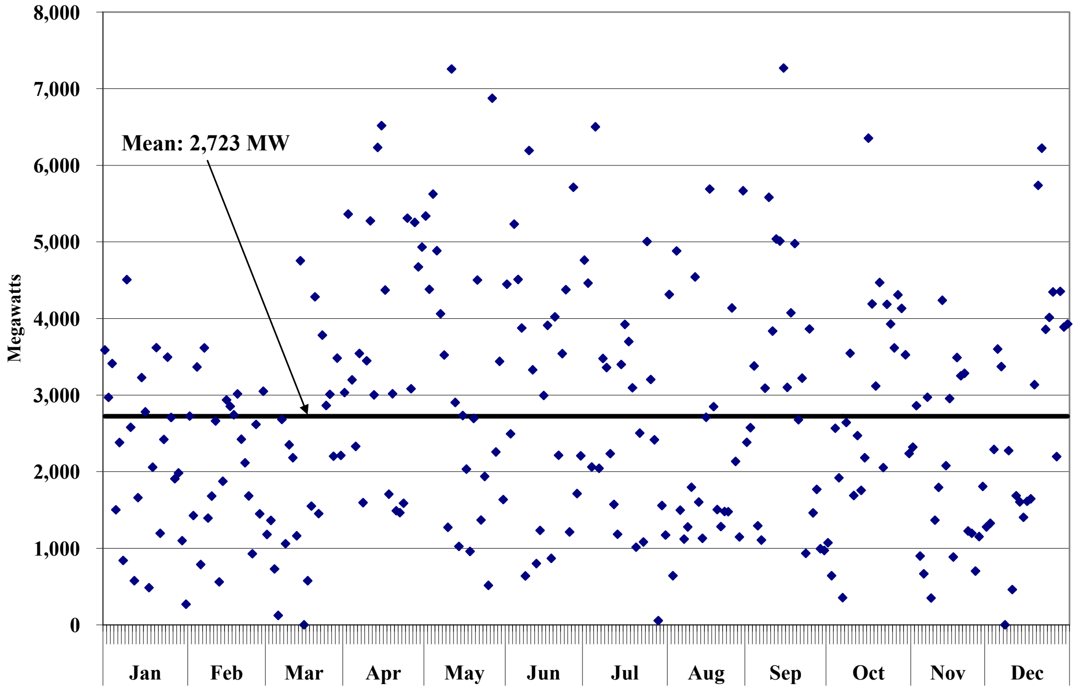


Figure 37 shows that the excess on-line capacity during daily peak hours on weekdays averaged 2,723 MW in 2008, which is approximately 7.6 percent of the average load in ERCOT. This is a reduction from prior years in which the same measure of on-online capacity average averaged 4,313, 2,927, and 3,020 MW in 2005, 2006 and 2007, respectively.

The overall trend in excess on-line capacity also 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

provide balancing energy.

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

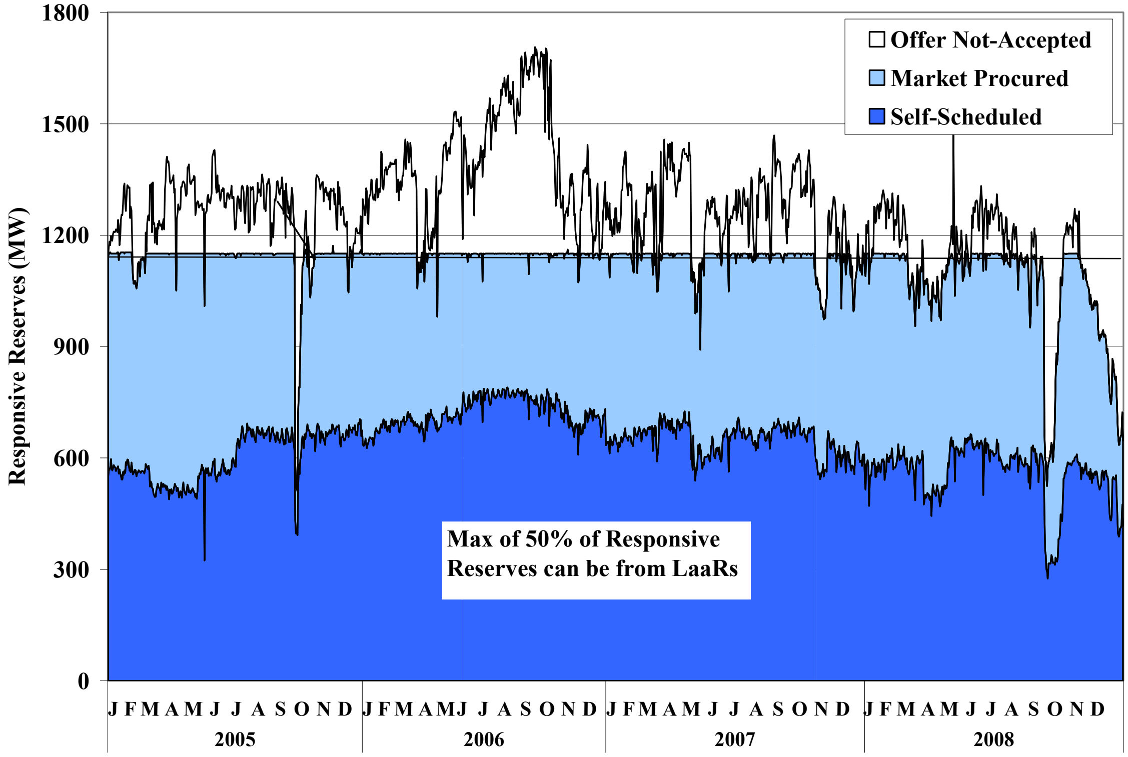
Demand response is a term that broadly refers to actions that can be taken by end users of electricity to reduce load in response to instructions from ERCOT or in response to certain market or system conditions. The ERCOT market allows participants with demand-response capability to provide energy and reserves in a manner similar to a generating resource. The ERCOT Protocols allow for loads to participate in the ERCOT administered markets as either Loads acting as Resources ("LaaRs") or Balancing Up Loads ("BULs").

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.

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 2008, 2,158 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 38 shows the amount of responsive reserves provided from LaaRs on a daily basis in 2008.

Figure 38: Provision of Responsive Reserves by LaaRs
Daily Average



The high level of participation by demand response participating in the ancillary service markets sets ERCOT apart from other operating electricity markets. Figure 38 shows that the amount of responsive reserves provided by LaaRs has held fairly constant at 1,150 MW since the beginning of 2005 (for reliability reasons, 1,150 MW is the limit of participation in the responsive reserve market by LaaRs). Notable exceptions were a period in September/October 2005 corresponding to Hurricane Rita, and a more prolonged decrease in September/October of 2008 corresponding to the Texas landfall of Hurricane Ike. Of interest in late 2008 is the post-hurricane recovery of the quantity of LaaRs providing Responsive Reserve followed by a steady reduction for the remainder of the year, which 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 or regulation services markets and their participation in the non-spinning reserves market was negligible in 2008. 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.

One change that may increase the participation in the non-spinning reserve market is the implementation of Protocol Revision Request No. 776 that was developed in late 2008 and was implemented in May 2009. This change will allow LaaRs that choose to provide this service to receive a daily capacity payment as in the past, but will also allow the LaaR to then offer the non-spinning reserve capacity into the balancing energy market that will determine the energy price at which the LaaR is willing to curtail its load. This change offers two benefits for LaaRs. First, assuming that the opportunity cost for LaaRs is typically much higher than the marginal cost of generating resources, the probability of deployment for LaaRs providing non-spinning reserves will be lower than in the past when non-spinning reserves were deployed independent of the marginal cost of the providers. Second, by allowing for the deployments to be based on energy price offers, LaaRs providing non-spinning reserves will be able to better predict and control the economics of providing the service in light of their particular business circumstances, as opposed to the prior practice in which non-spinning reserves were deployed as a price taker. These changes are expected to lead to the entry of some quantity of LaaRs into the non-spinning reserve market, which would obviously be an improvement compared to the history of no participation by LaaRs in the non-spinning reserve market.

E. Net Revenue Analysis

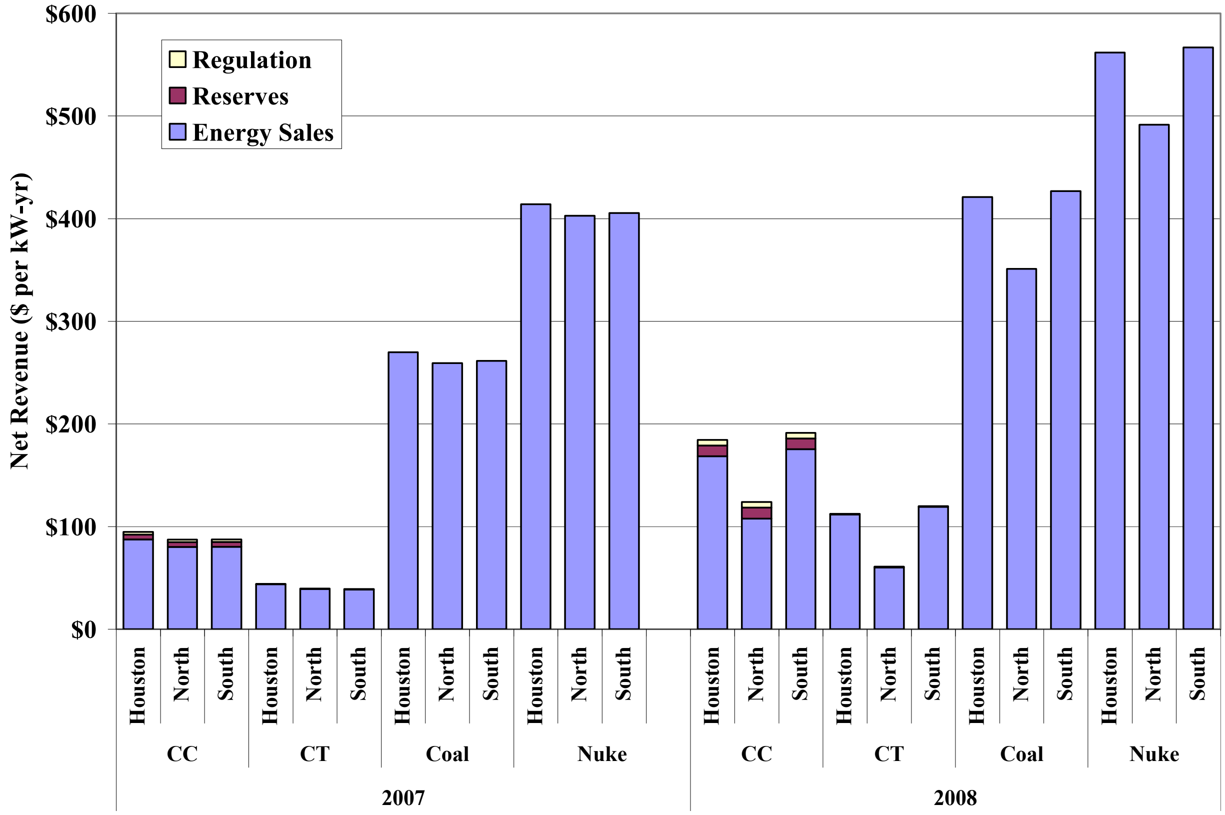
Net revenue is defined as the total revenue that can be earned by a generating unit less its variable production costs. Hence, it is the revenue in excess of short-run operating costs and is available to recover a unit's fixed and capital costs. Net revenues from the energy, operating reserves, and regulation 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 39 shows the results of the net revenue analysis for four types of units in 2007 and 2008. These are: (a) a gas combined-cycle, (b) a combustion turbine, (c) a new coal unit, and (d) a new 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 39: 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 MMbtu per MWh for a new coal unit. We assume variable operating and maintenance costs of $4 per MWh for the gas units and $1 per MWh for the coal unit. We assume variable costs of $5 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 39 shows that the net revenue increased substantially in 2008 in each zone compared to 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 2008 for a new gas turbine was approximately $120, $113 and $61 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 2008 for a new combined cycle unit was approximately $191, $185 and $124 per kW-year in the South, Houston and North Zones, respectively. These values indicate that the estimated net revenue in 2008 was sufficient to support new entry for a new gas turbine in the South and Houston zones and for a combined cycle unit in the South, Houston and North zones. However, as discussed later in this subsection, significant portions of the net revenue results for gas turbine and combined cycle units in 2008 can be attributed to anomalous market design related inefficiencies rather than fundamentals that would support an investment decision for new gas turbines and combined-cycle units.

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 have 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. For a new coal unit, the estimated net revenue requirement is approximately $190 to $245 per kW-year. The estimated net revenue in 2008 for a new coal unit was approximately $427, $421 and $351 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 2008 for a new nuclear unit was approximately $567, $562 and $492 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 sufficient to support new entry in 2008, as was the case in 2005, 2006 and 2007. Thus, it is not surprising that some market participants are building new baseload facilities and that several others have initiated activities that may lead to the construction of additional baseload facilities in the ERCOT region.

Although estimated net revenue grew considerably in 2008 compared to prior 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. Finally, and most importantly in 2008, net revenues can be inflated when prices clear at high levels due to inefficiencies in the market design. Similar to the case of market power, a market participant may be deterred from investing in new capacity if it believes that prevailing net revenues are largely due to market design inefficiencies that will be corrected.

Such market design inefficiencies were apparent in 2008. As discussed in Section III, the vast majority of price excursions in 2008 – particularly in the South and Houston Zones – were not a function of market fundamentals; rather, the price excursions were driven by inefficient congestion management techniques that have since been corrected and are not expected to materialize in the future, especially upon implementation of the nodal market in 2010. In addition to these transmission congestion issues, in 2008 the ERCOT Protocols provided for *ex post* re-pricing provisions in intervals in which non-spinning reserve prices were deployed that frequently resulted in scarcity level prices at times when ERCOT's operating reserve levels were not deficient. These rules were changed as a part of the aforementioned PRR 776, thereby reducing the probability of scarcity level prices during non-scarcity conditions going forward. Hence, a significant portion of the net revenue produced in 2008 is not reflective of fundamentals that would support an investment decision for new gas turbines and combined cycle units.

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 40 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.[14]

Figure 40: Comparison of Net Revenue of Gas-Fired Generation between Markets

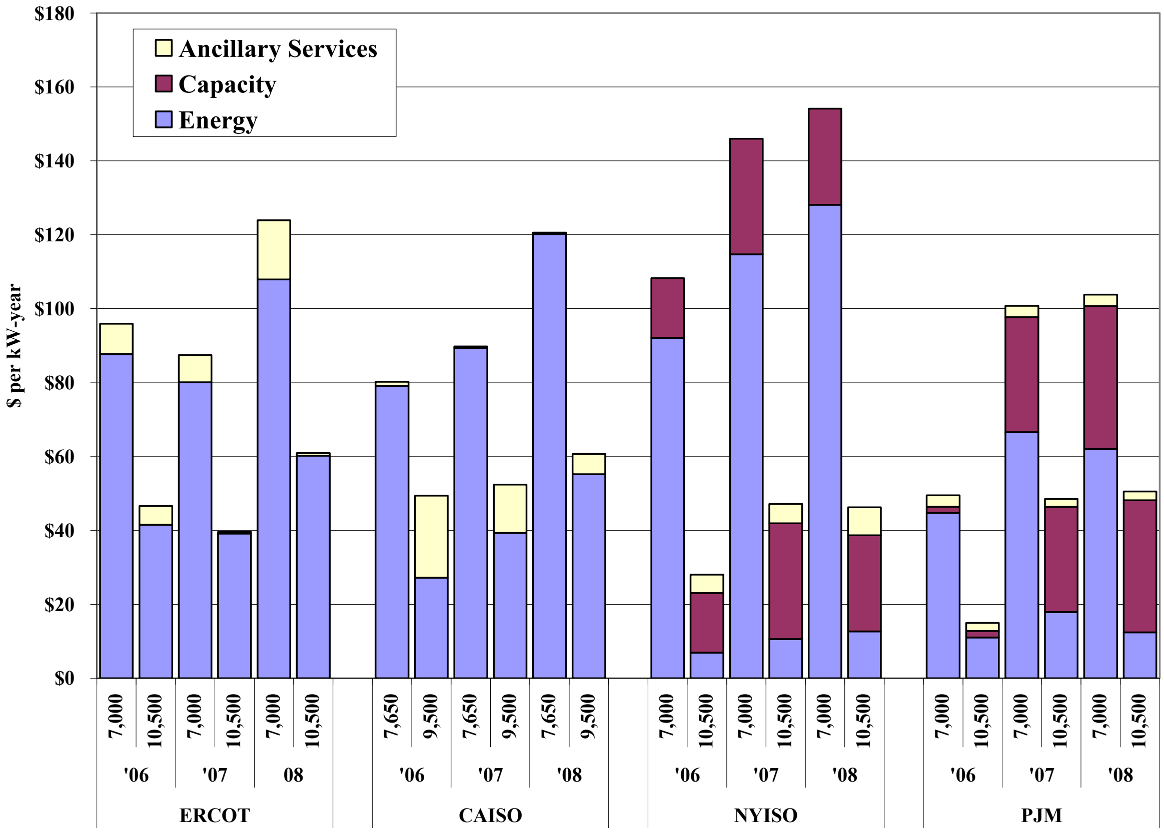


Figure 40 shows that net revenues increased in all markets from 2007 to 2008, with the exception of gas peaking units in New York that remained flat. In the figure above, net revenues are calculated for central locations in each of the five markets. However, there are load pockets within each market 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.

14 The California ISO does not report capacity and ancillary services net revenue separately, so it is shown as a combined block in Figure 40. 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.

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.

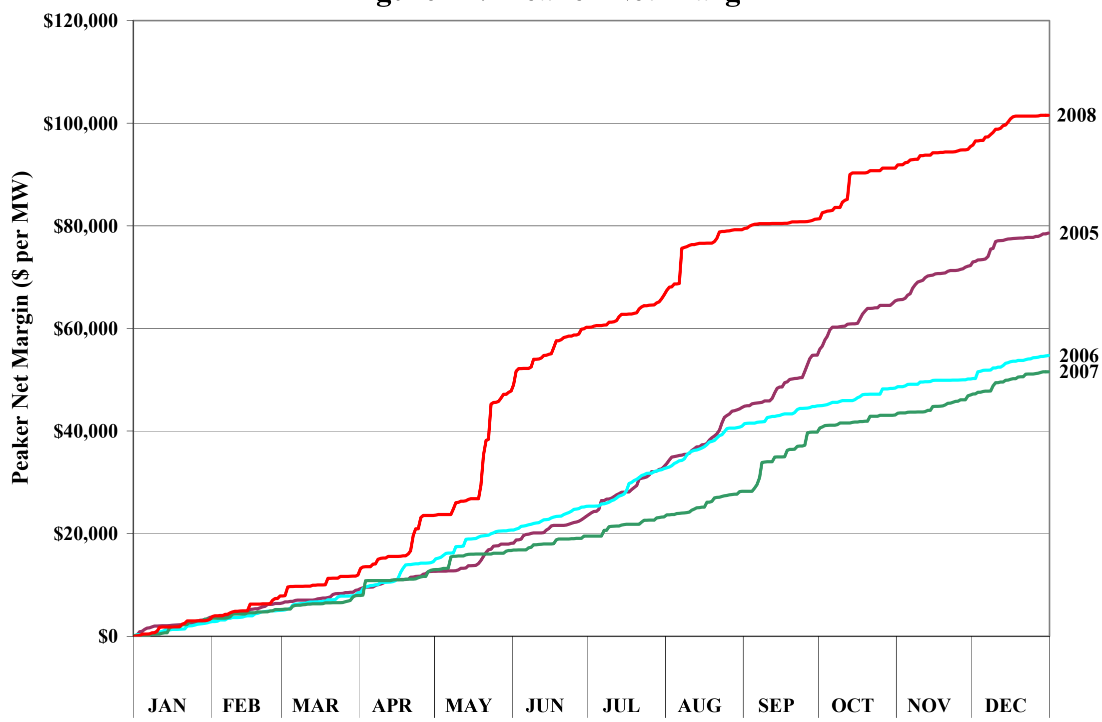
PUCT Subst. Rule 25.505 provides that the IMM may conduct an annual review of the effectiveness of the SPM. This subsection provides an assessment of the effectiveness of the SPM in 2008 under ERCOT's energy-only market structure.

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. Although the PNM was not in effect prior to 2007, Figure 41 shows the cumulative PNM that would have been produced for each year from 2002 through 2007.[15]

Figure 41: 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 41 and consistent with the previous findings in this section relating to net revenue, the PNM reached the level sufficient for new entry in only two of the last five years (2005 and 2008). In 2008, the peaker net margin and net revenue values rose substantially, surpassing the level required to support new peaker entry. However, as previously discussed, a significant portion of the net revenue increase in 2008 was associated with extremely inefficient transmission congestion management and inefficient

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

pricing mechanisms associated with the deployment of non-spinning reserves. Both of these issues have been corrected in the zonal market and will be further improved with the implementation of the nodal market in 2010. Absent these inefficiencies, net revenues would not have been sufficient to support new peaker entry in 2008. Beyond these anomalies, there were three other factors that significantly influenced the effectiveness of the SPM in 2008:

- A substantial decrease in out-of-merit deployments by ERCOT during declared short-supply conditions;
- A continued strong positive bias in ERCOT's day-ahead load forecast that tended to regularly commit online resources in excess of the quantity required to meet expected demand and operating reserve requirements; and
- The dependence on market participants to submit offers at or near the offer cap to produce scarcity level prices during legitimate shortage conditions.

1. Out-of-Merit Deployments during Shortage Conditions

As discussed in the 2007 SOM Report, ERCOT implemented new operating procedures in 2007 whereby it deployed Non-Spinning Reserve Service ("NSRS") when Adjusted Responsive Reserves ("ARR") were reduced to 2,500 MW. If NSRS was not procured, had already been deployed, or could not be timely deployed, ERCOT issued out-of-merit ("OOM") instructions to offline, quick-start units. ARR is a measure that is based upon available responsive reserves, but incorporates a discount factor that is applied to the capacity of online generating units. This discount factor was developed by ERCOT based on prior experience during emergency operating conditions, and is intended to account for the uncertainty in the actual maximum capacity that is deliverable when called upon during emergency conditions.

Although well-intended from a reliability perspective, from a market efficiency perspective, the use of the discount factor in 2007 created an "overlap" between market and reliability operations that often led to inefficient pricing outcomes during shortage and near-shortage conditions.

Efforts in 2007 to address these inefficiencies led to an interim measure that was implemented in January 2008 that increased the procurement of responsive reserves to offset the effect of the application of the discount factor, thereby significantly reducing the "overlap" between market and reliability operations that was frequently experienced in 2007. The responsive reserve procurement increase was linked directly to the magnitude of the discount factor. Additionally, Protocol Revision Request No. 750 was adopted in 2007 that provided for unannounced generator capacity testing with the objective of providing ERCOT enough confidence to eliminate the discount factor, which would also eliminate the increased procurement of responsive reserves.

In the 2007 SOM Report, we noted that "implementation of PRR 750 in 2008 will not only lead to the elimination of the discount factor, but will also eliminate the interim measure of increased procurement of responsive reserves. Ultimately, the successful implementation of PRR No. 750 should lead to more reliable and efficient operations in the ERCOT wholesale market." In fact, by August 2008, ERCOT had performed sufficient unit testing under PRR 750 that provided it with the confidence to substantially reduce the discount factor to a level that eliminated the increased procurements of responsive reserves. Together, the interim increase in responsive reserve procurements and implementation of PRR 750 worked very successfully to virtually eliminate the out-of-merit deployments by ERCOT during shortage conditions in 2008, thereby improving both the efficiency and reliability of market operations.

2. 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 being unable 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 42: Average Day Ahead Hourly Load Forecast Error by Season

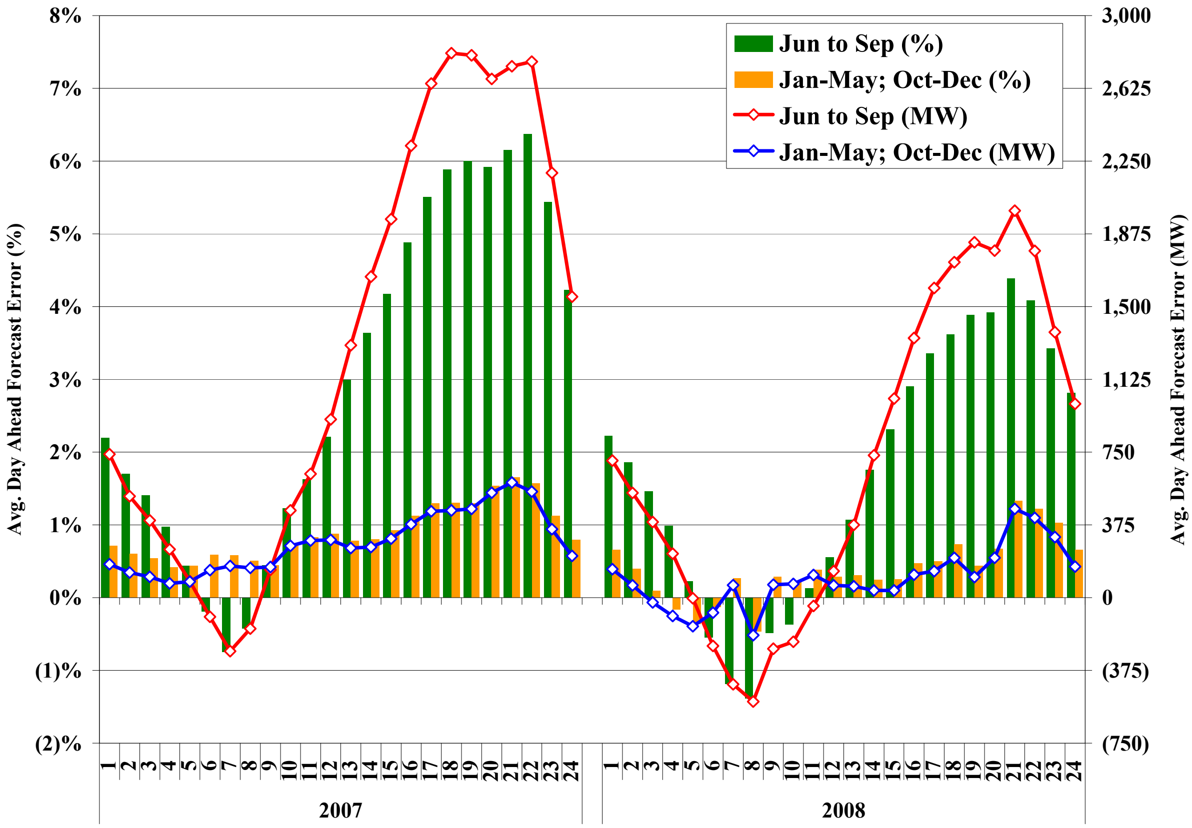
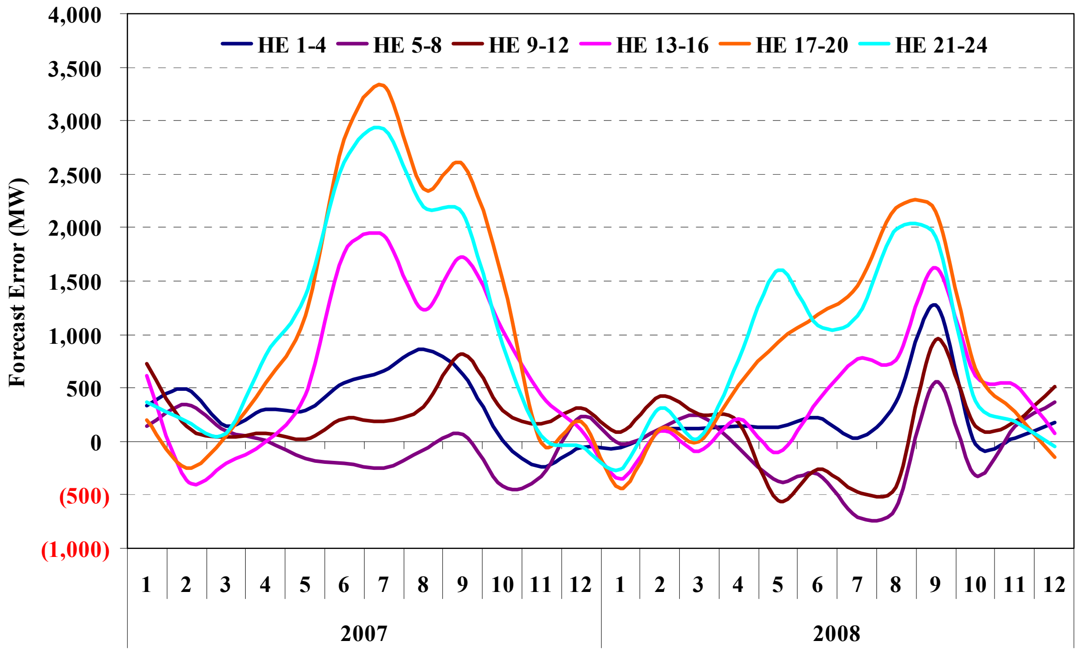


Figure 42 shows the average hourly day-ahead load forecast error for the summer months of June through September, and also for the months of January through May and October through December for 2007 and 2008. In this figure, positive values indicate a day-ahead load forecast that was greater than the 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 and 2008, 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 205 MW for the non-summer months, and an average over-forecast of 1,729 MW for the four summer months in 2008. Although the performance in 2008 was generally improved compared to 2007, Figure 42 clearly shows that the positive day-ahead load forecast bias observed in 2007 persists in 2008. Figure 43 shows another view of the same load forecast error data as in Figure 42 for 2007 and 2008 with the average megawatt error displayed for each month in four hour blocks (hours ending).

Figure 43: Average Day Ahead Load Forecast Error by Month and Hour Blocks



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.[16]

3. Dependence on High-Priced Offers by Market Participants

As a general principle, competitive and efficient market prices should be consistent with the cost of the marginal action taken to satisfy the market's demand. In the vast majority of hours, the marginal action is the dispatch of the most expensive online generator. 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:

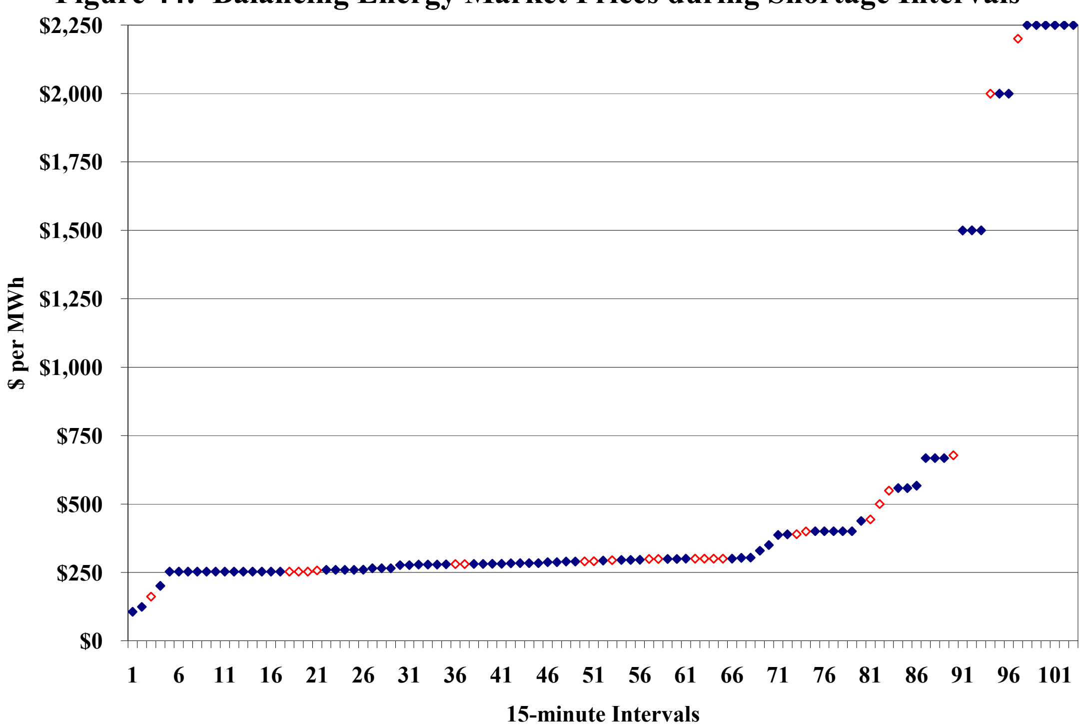
16 It is our understanding that ERCOT's current procedures allow to some extent for the deferral of the commitment of short-lead time resources.

- 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 44 shows the balancing market clearing prices during the 103 15-minute shortage intervals in 2008.

Figure 44: Balancing Energy Market Prices during Shortage Intervals



As shown in Figure 44, the prices during these 103 shortage intervals in 2008 ranged from $105 per MWh to the offer cap of $2,250 per MWh (prior to March 1, 2008, the offer cap was $1,500 per MWh), with an average price of $534 per MWh and a median price of $293 per MWh. The results in 2008 are similar to those in 2007 when there were 108 shortage intervals with an average price of $476 per MWh and a median price of $299 per MWh.

The data in Figure 44 are separated into solid blue and red outlined points. The blue points (79) represent true shortage conditions, whereas the red points (24) represent artificial shortage prices occurring as a result of large generation schedule reductions at the top of the hours from 10 PM to 1 AM. As discussed in more detail in Section I, the production of such artificial shortage prices under these conditions is the result of inefficiencies inherent to the current market design that will be significantly improved with the implementation of the nodal market.

Although each of the data points in Figure 44 represents system conditions in which the market was in shortage, the pricing outcomes are widely varied, with the majority of prices reflecting the marginal cost 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 and 2008.

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 recent years. In contrast, private investment in mid-merit and peaking resources in ERCOT has been relatively thin. In an energy-only market, net revenue expectations for mid-merit and peaking resources are much more sensitive the effectiveness of the shortage pricing mechanism than to factors such as 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 small market participants to effectively withhold lower cost resources by offering at prices dramatically higher than their marginal cost.

At least for the pendency of the zonal market, shortage pricing will continue to remain dependent upon the existence of high-priced offers by market participants, and results such as those experienced in 2007 and 2008 will continue to frustrate the objectives of the energy-only market design.[17] 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 shortage conditions. While important even in markets with a capacity market, 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.

17 Net revenue in 2008 was sufficient to support new entry for peaking resources in the Houston and South Zones. However, as discussed in Section III, the vast majority of price excursions in 2008 were not a function of market fundamentals; rather, the price excursions were driven by inefficient congestion management techniques and pricing mechanisms associated with the deployment of non-spinning reserves that have been corrected and are not expected to materialize in the future. Hence, the net revenue values produced in 2008 are generally not reflective of fundamentals that would support an investment decision for new gas turbine and combined cycle units.

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.

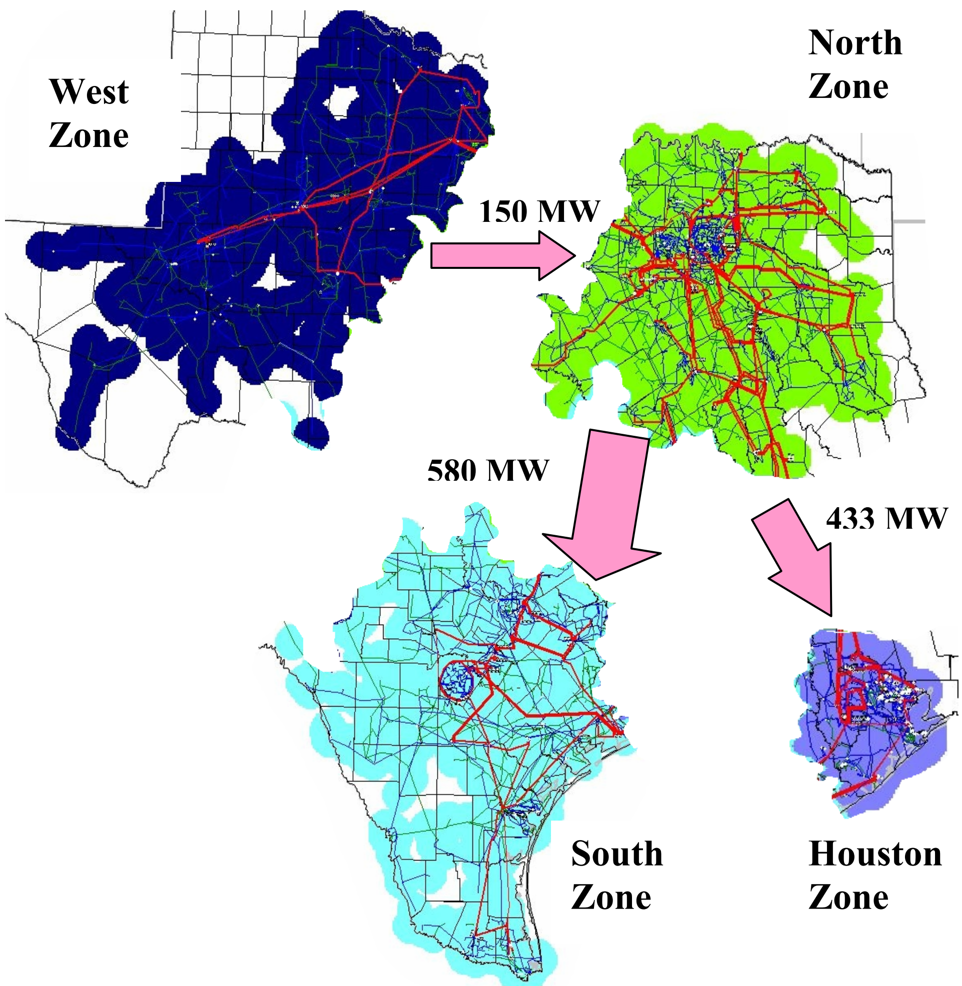
A. Electricity Flows between Zones

In 2008 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. From year-to-year, slight adjustments are sometimes made to the boundaries of the commercial pricing zones. However, the vast majority of customers remained in the same zone from 2007 to 2008. 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 2008.

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 45: Average SPD-Modeled Flows on Commercially Significant Constraints During All Intervals in 2008



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 150 MW, which is equivalent to saying that the average for the North-to-West CSC was *negative* 150 MW.

Figure 45 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.[18] In reality, the generators and loads within each zone can have widely varying 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 2008	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	150	139	*-11*	128	*-11*
North - South	580	570	*-10*	-44	*-614*
North-Houston	433	436	*3*	753	*317*

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 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.

[18] 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 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 614 MW and increased the calculated flows on the North to Houston CSC by 317 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 46 shows the actual average imports of power for each zone in 2008. In this figure, positive values represent imports, and negative values indicate exports.[19]

19 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 46.

Figure 46: Actual Zonal Net Imports

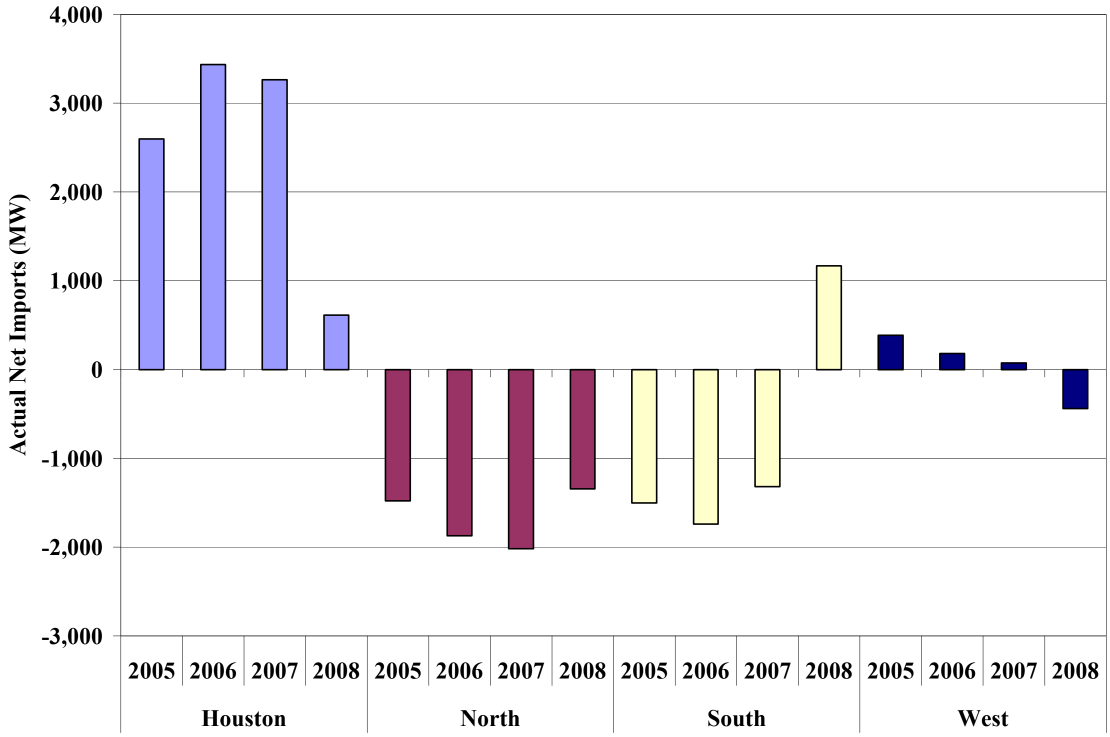


Figure 46 shows that the Houston Zone is a net importer of power, while the North Zone is a net exporter. The reduction in the Houston Zone imports in 2008 and corresponding change in the South Zone from a net exporter to a net importer can be attributed to the movement of the 2,700 MW South Texas Nuclear Project from the South Zone to the Houston Zone in 2008. The West Zone transitioned from a net importer in 2005 to a net exporter in 2008, which is reflective of 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 47 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 47: Average SPD-Modeled Flows on Commercially Significant Constraints During Transmission Constrained Intervals in 2008

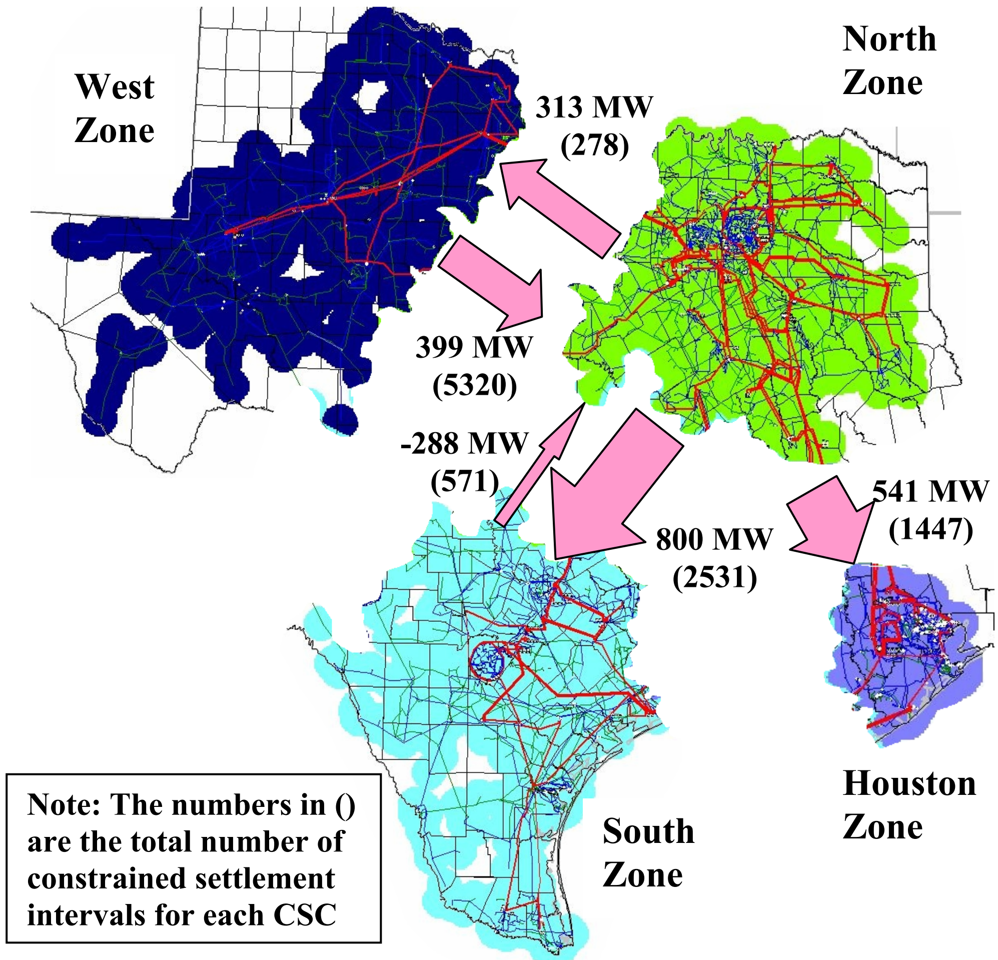


Figure 47 shows that inter-zonal congestion was most frequent in 2008 on the West to North and the North to South CSCs, followed by the North to Houston CSC. The West to North CSC exhibited SPD-calculated flows averaging 399 MW during 5,320 constrained intervals (15 percent of the totals intervals in the year). The North to South CSC exhibited SPD-calculated flows averaging 800 MW during 2,531 constrained intervals (7 percent of the total intervals), and the SPD-calculated average flow for the North to Houston CSC was 541 MW during 1,447 constrained intervals (4 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 2008	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	399	345	*-54*	335	*-10*
North - South	800	633	*-167*	139	*-494*
North-Houston	541	418	*-123*	827	*409*

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 CSC were relatively close to actual flows, whereas the average actual flows for the North to South 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 2008.

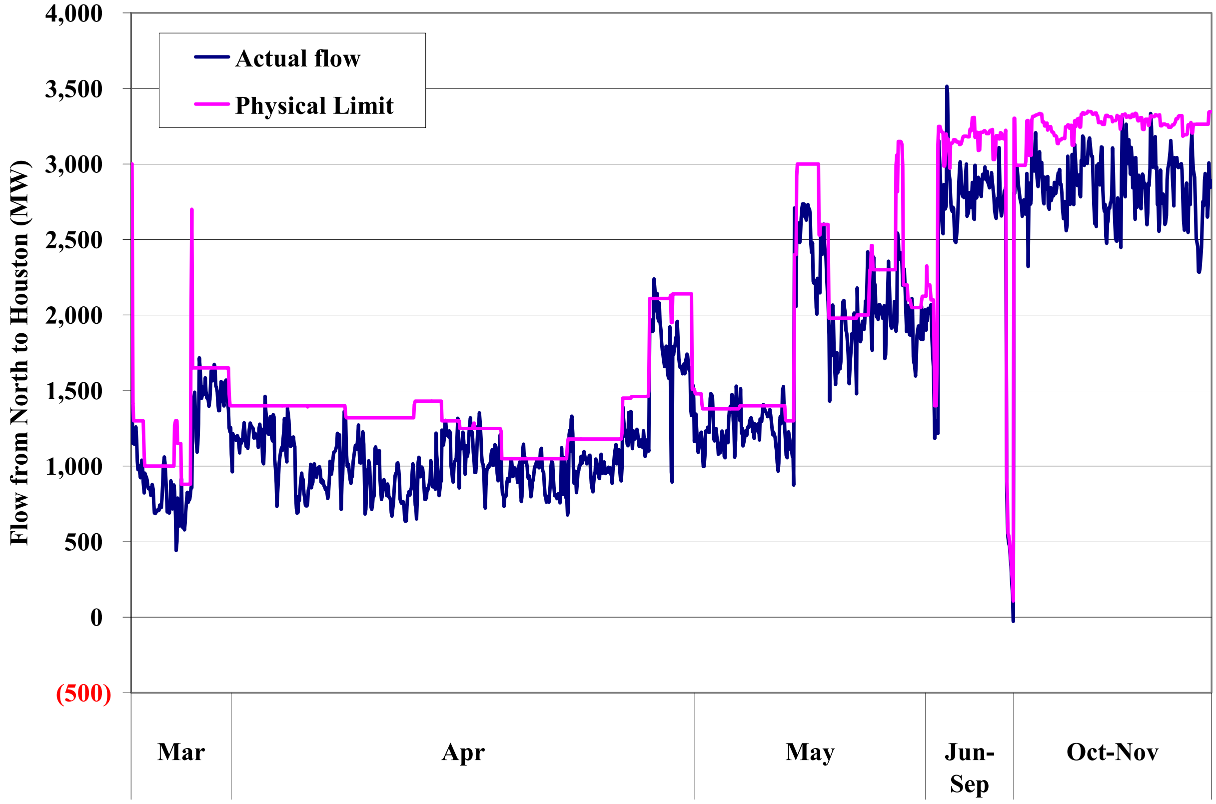
1. Congestion on North to South and North to Houston CSCs

The majority of congestion activity for the North to South and North to Houston CSCs was affected by common factors in 2008. Therefore, the analysis of both of these CSCs is provided in this single subsection.

The North to South CSC was new in 2008 and was added based on experience in 2007 of increased local congestion activity on this interface. In 2008 the North to South CSC was binding in 2,531 intervals (7 percent) with an annual average shadow price of $22 per MW. The North to Houston CSC was binding in 1,447 intervals (4 percent) with an annual average shadow price of $20.

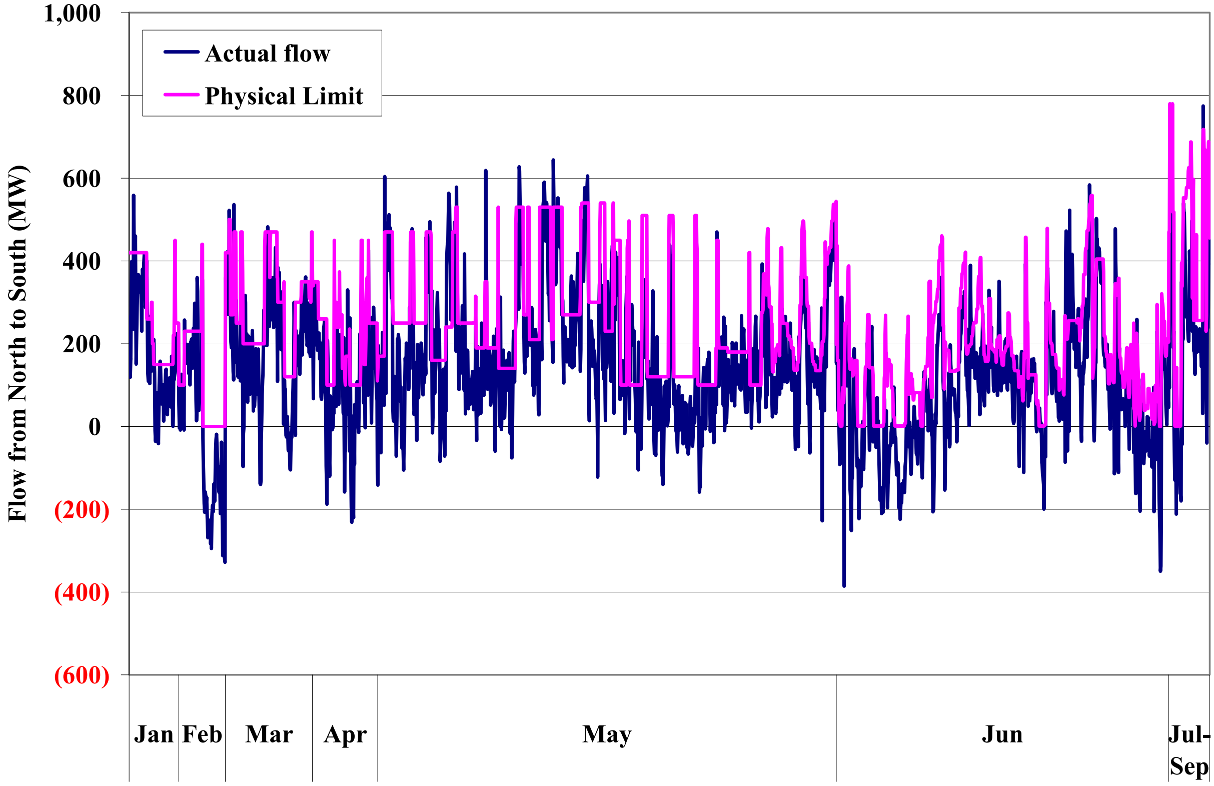
Beginning in April and continuing into May 2008, the frequency of congestion on the North to Houston CSC began to increase, at times becoming so significant that the constraint was unable to be resolved with available balancing energy in the zones where it was required. When congestion on a CSC cannot be resolved, maximum shadow prices are produced for the CSC that, in turn, produce balancing energy market prices in the deficient zones that can approach or even exceed the system-wide offer caps (the system-wide offer cap was $2,250 per MWh beginning March 1, 2008). Figure 48 shows the actual flows versus the physical limit for the North to Houston CSC in 2008 during intervals when the CSC was binding.

Figure 48: Actual Flows versus Physical Limits during Congestion Intervals North to Houston



Similar to the North to Houston CSC, the frequency of congestion on the North to South CSC began to increase beginning in May and continuing into early June 2008. Like the North to Houston CSC, the North to South CSC also experienced an increasing number of intervals in which the congestion on the interface could not be resolved, thereby producing maximum shadow prices on the constraint and associated high balancing energy market prices in the deficient zones (Houston and South). Figure 49 shows the actual flows versus the physical limit for the North to South CSC in 2008 during intervals when the CSC was binding.

Figure 49: Actual Flows versus Physical Limits during Congestion Intervals North to South



Historically, the inability to resolve a zonal constraint has been a relatively rare occurrence. In fact, excluding the North to Houston and North to South CSCs, the other CSCs together averaged only 15 intervals in 2008 with shadow prices that were greater than or equal to the current maximum CSC shadow price of $5,000 per MW. In contrast, the North to South and North to Houston CSCs experienced shadow prices greater than or equal to $5,000 per MW in 92 and 87 intervals, respectively.

The sharp increase in the frequency of occurrence of unresolved congestion on the North to Houston and North to South CSCs prompted the IMM, in consultation with ERCOT and the PUCT, to initiate in early May 2008 a detailed examination of ERCOT's congestion management procedures. This investigation quickly revealed that ERCOT rules permitted certain transmission elements to be managed with zonal balancing energy deployments when, in actuality, the congestion on these elements was neither effectively nor efficiently resolvable with zonal balancing energy deployments (the transmission elements that can be designated to be managed with zonal balancing energy deployments in the same manner as the CSC are referred to as "Closely Related Elements (CREs)" in the ERCOT Protocols).

Under the current zonal market model, transmission congestion is resolved through a bifurcated process that consists of either (1) zonal balancing energy deployments, or (2) local, unit-specific deployments. Because the pricing and incentives for both load and resources are better aligned with zonal congestion management techniques under the zonal market model, it is preferable to manage transmission congestion using zonal balancing energy to the extent it is effective and efficient.

However, for the CREs in question related to the North to Houston and North to South CSCs, zonal balancing energy deployments were neither effective nor efficient in resolving the transmission congestion. The result was an increasing frequency of the deployment of substantial quantities of energy in both the Houston and South Zones up to the point of exhaustion, thereby triggering the maximum shadow prices for these CSCs and associated high balancing energy prices in the South and Houston Zones that approached or even exceeded the system-wide offer cap of $2,250 per MWh.

To address these market and reliability issues, in late May 2008, again in consultation with ERCOT and the PUCT, the IMM submitted Protocol Revision Request ("PRR") 764, which improved the definition of those transmission elements eligible to be designated as CREs. PRR 764 was processed through the ERCOT committees on an expedited basis and was implemented on June 9, 2008.

While PRR 764 effectively resolved the issues encountered in April through June 2008 within the context of the zonal market model framework, the implementation of the nodal market will eliminate the current bifurcated congestion management process by providing simultaneous, unit-specific solutions that will always present the most effective and efficient congestion management alternatives to the system operator. As a point of comparison, Figure 50 shows the pricing implications of unresolved North to South congestion under the zonal model, and Figure 51 shows the pricing implications of the same unresolved constraint under a nodal market model.

Figure 50: Pricing Contours of Unresolved Congestion in the Zonal Market

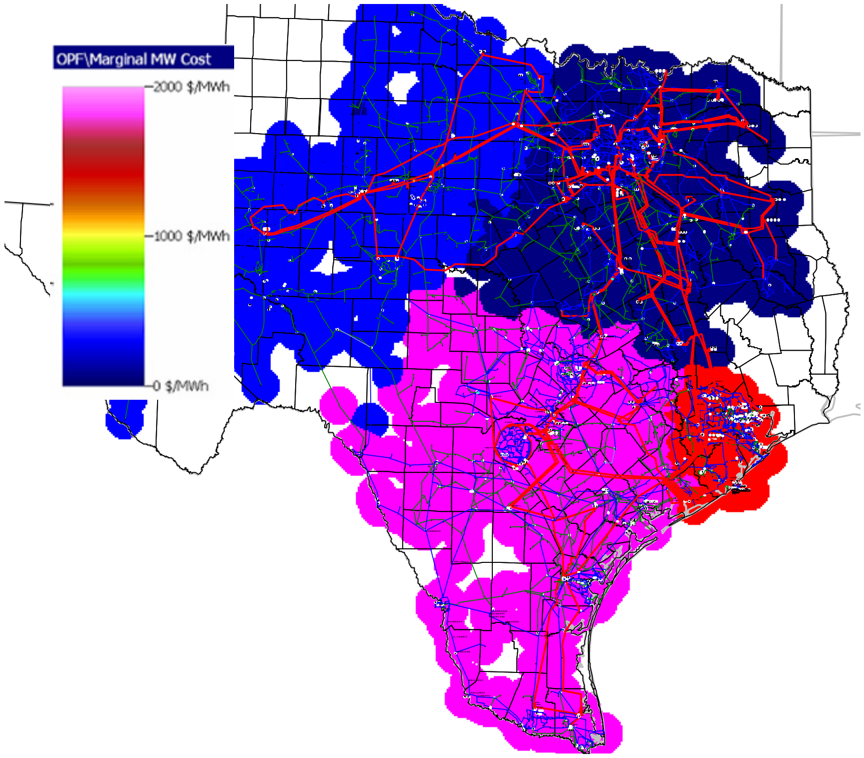
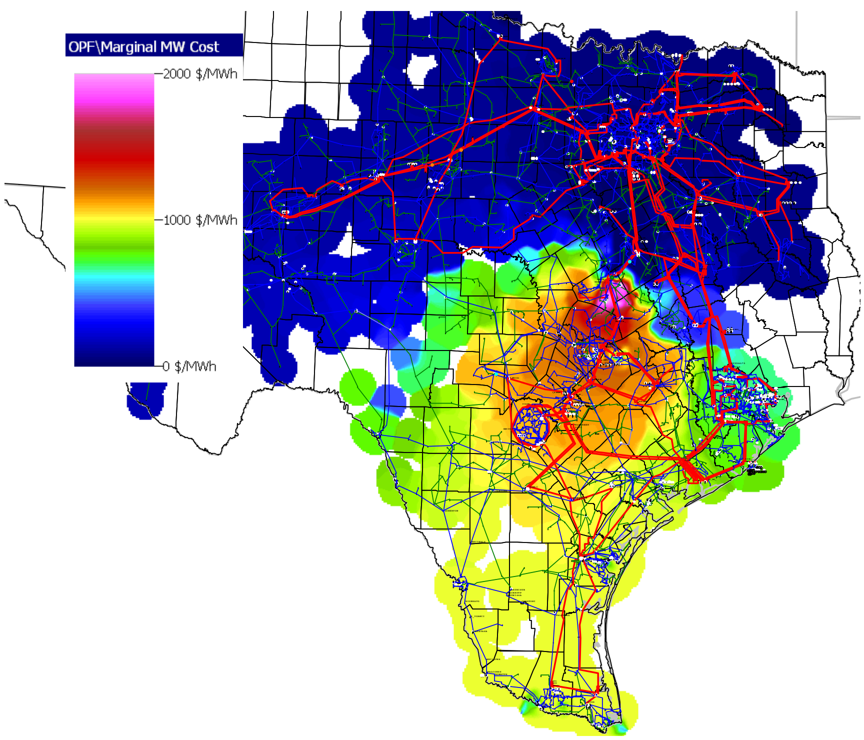


Figure 51: Pricing Contours of Unresolved Congestion in the Nodal Market



These graphics show that an unresolved constraint produces extremely high market clearing prices that are very widespread under the zonal model. In contrast, for the same unresolved constraint, the resulting high prices remain much more localized under the nodal model. These differences in pricing outcomes are due to the use of zonal average shift factors under the zonal model, as previously discussed, compared to the use of location-specific shift factors under the nodal model. Further, because the nodal market employs unit-specific offers and dispatch, the control of power flows on the system is much more flexible and precise than under the zonal model. Hence, it is much less likely under nodal dispatch to even encounter unresolvable constraints such as those experienced on the North to South and North to Houston interfaces in 2008.

In consideration of these differences, we have estimated the benefits that the nodal market would have produced by allowing more efficient resolution of the congestion on the North to South and North to Houston interfaces in 2008. To produce this estimate we assume that a nodal market model would not have eliminated the existence of congestion, but would have been able to resolve the congestion more efficiently without the frequent exhaustion of all available resources. Additionally, consistent with the results shown in Figure 50 and Figure 51, we recognize that the distribution of extremely high market-clearing prices due to congestion is much more localized under the nodal model than the widespread distribution under the zonal model, thereby affecting the average load zone price to a lesser degree under the nodal model. To account for these factors, in each interval that either the North to Houston or the North to South CSC was binding, we limited the price in the import-constrained zones (Houston and South) to a maximum of 20 times the price of natural gas and recalculated the annual average balancing energy market price for the Houston and South Zones in 2008.

This analysis indicates that the annual average balancing energy market price in the Houston and South Zones would have been reduced by approximately $10.42 per MWh under these assumptions. With approximately 168 million MWh consumed in the Houston and South Zones in 2008, the implied value of this reduction is approximately $1.75 billion in 2008. However, because of existing contracts and the fact that the 2008 congestion excursions were contained to a relatively short period of time and not expected to recur in the future, not all customers were directly affected by these wholesale price increases. Assuming that only 5 to 10 percent of customers in the South and Houston Zones were directly affected by the significant price increases in the balancing energy market and short-term bilateral markets associated with the North to Houston and North to South congestion, this analysis indicates that the efficiencies of the nodal market, had it been in place, could have reduced the annual costs for customers by $87 to $175 million in 2008. This analysis estimates only the savings that could have occurred through more efficient congestion management during the periods of acute North to Houston and North to South congestion, and does not include the benefits that the nodal market will provide more generally with respect to congestion management and other dispatch efficiency improvements.

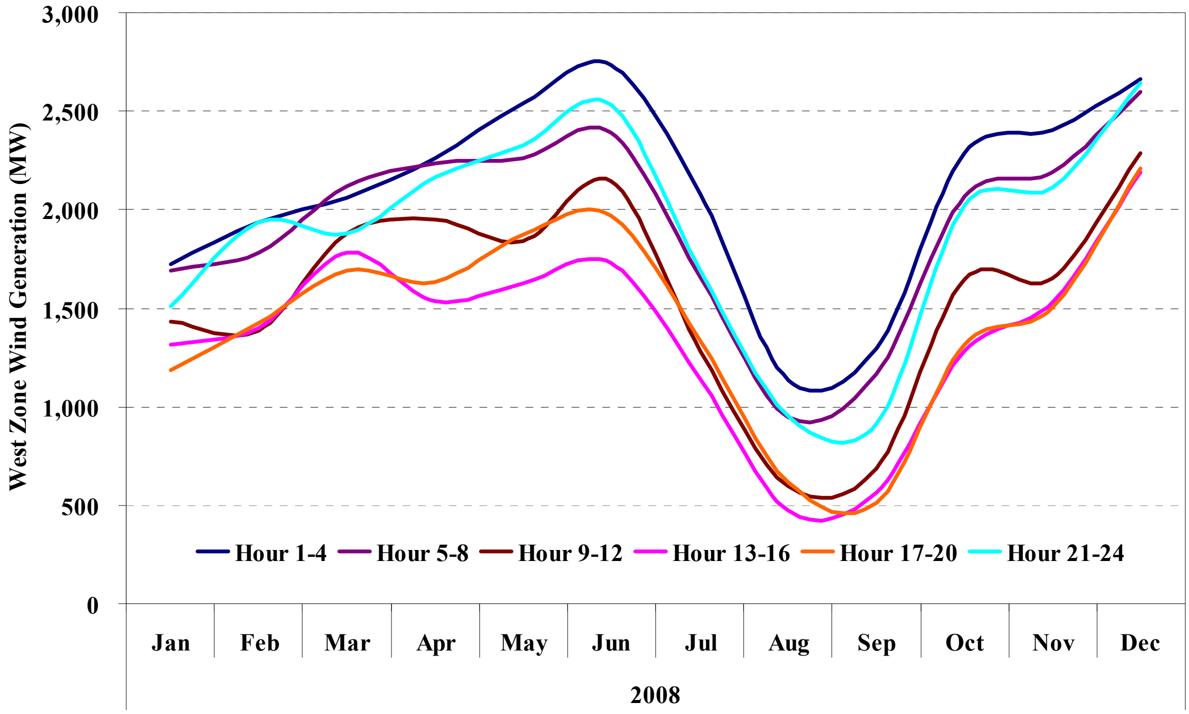
2. Congestion on West to North CSC

In 2008 the West to North CSC was binding in 5,320 intervals (15 percent). This was more frequent than any other CSC in 2008 and 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 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. The installed wind capacity in ERCOT grew from approximately 4.5 GW at the beginning of the year to approximately 8.1 GW by December 2008, with more than 90 percent of wind capacity located in the West Zone.

Average load in the West Zone was 2,547 MW in 2008, with a minimum of 1,828 MW and a maximum of 3,910 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 52 shows the average West Zone wind production for each month in 2008, with the average production in each month shown separately in four hour blocks.[20]

[20] Figure 52 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 52 are lower than the production levels that would have materialized absent transmission constraints.

Figure 52: Average West Zone Wind Production



Depending on load levels, transmission system topology and other system operating conditions, up to 4 GW of wind generation in the West Zone can be reliably produced before encountering a transmission export constraint on the West to North CSC. Figure 53 shows the actual flows and the physical limit for the West to North CSC in 2008 for intervals in which the CSC was binding.

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

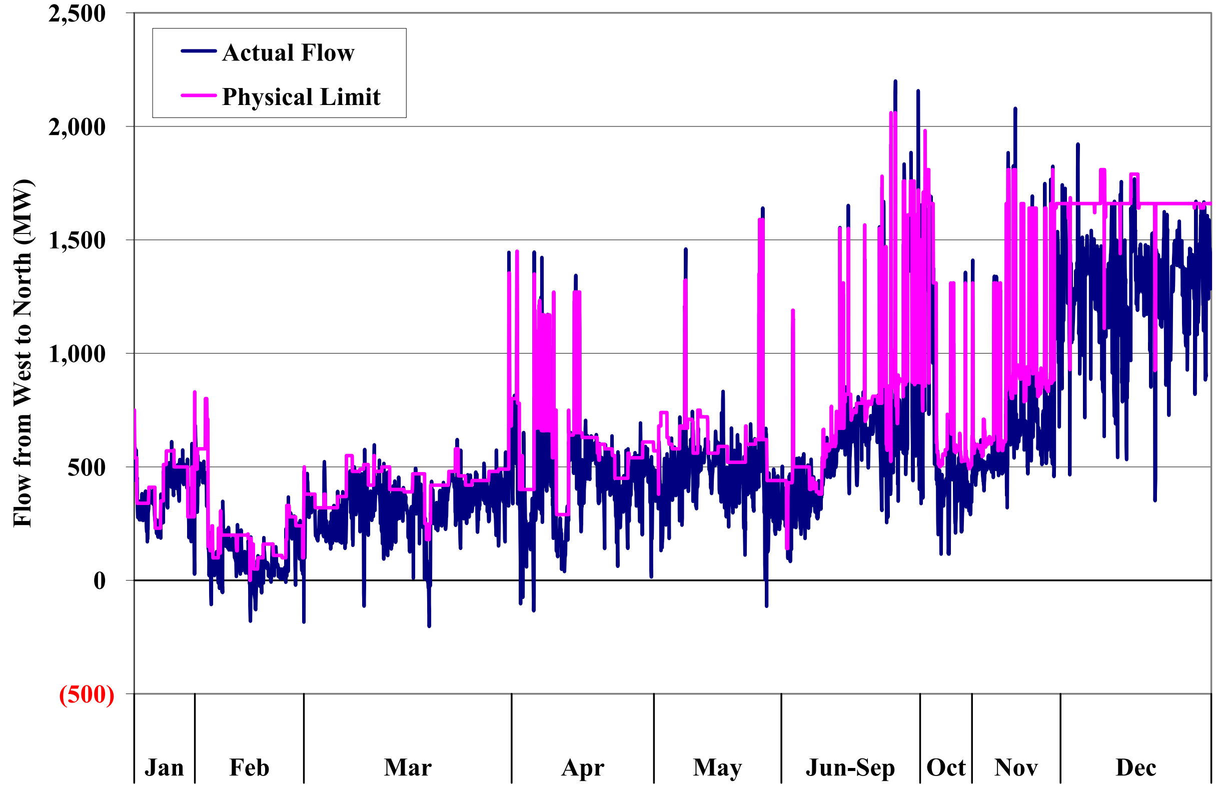


Figure 53 shows that over one-half of the binding intervals occurred in January through May 2008, with relatively fewer binding intervals in the summer months, and congestion activity increasing again in October through December.[21] Hence, although Figure 52 shows actual wind production in January through June that was comparable or lower than October through December, West to North congestion activity was more frequent in the first half of the year.

This higher frequency of congestion on the West to North CSC in the early part of the year was affected by two factors. First, the West to North CSC was subject to similar CRE issues previously described for the North to South and North to Houston CSC during this timeframe, although to a much lower degree. While the majority of the congestion on the North to South and North to Houston CSCs was related to constraints that were not included as CREs with the implementation of PRR 764, the majority of the congestion on the West to North CSC in the first

21 The West to North CSC consists of thermal transfer limits and a dynamic stability limit. Both are represented in Figure 53. Generally, in intervals with a physical limit less than 1,000 MW, the thermal limit was binding, and in intervals with a physical limit greater than 1,000 MW, the dynamic stability limit was binding.

part of the year was related to constraints that remained as CREs following the implementation of PRR 764. The second and more significant explanation for the increased congestion on the West to North CSC in the first part of the year is that a major double-circuit 345 kV transmission line was out of service for maintenance for most of February and March, thereby significantly reducing the West to North transfer capability. However, even with the changes implemented with PRR 764 and most major transmission lines in service, Figure 53 shows that congestion activity increased again in October through December 2008, and that this increase in congestion activity corresponds to the increase in wind production shown in Figure 52 for these months.

Although plans exist through the Competitive Renewable Energy Zone ("CREZ") project to significantly increase the transmission export capability from the West Zone, it is likely given the current transmission infrastructure and the level of existing wind facilities in the West Zone that the quantity of wind production that can be reliably accommodated in the West Zone will continue to be significantly limited for several years until such transmission improvements can be completed.

3. Congestion on North to West CSC

The North to West CSC was binding in 278 intervals (0.8 percent) in 2008. The North to West CSC was primarily binding under system conditions where wind generation in the West Zone is very low and few of the thermal generating units in the West Zone are online. Figure 54 shows the actual flows and the physical limit for the North to West CSC in 2008 for intervals in which the CSC was binding.

Figure 54: Actual Flows versus Physical Limits during Congestion Intervals North to West

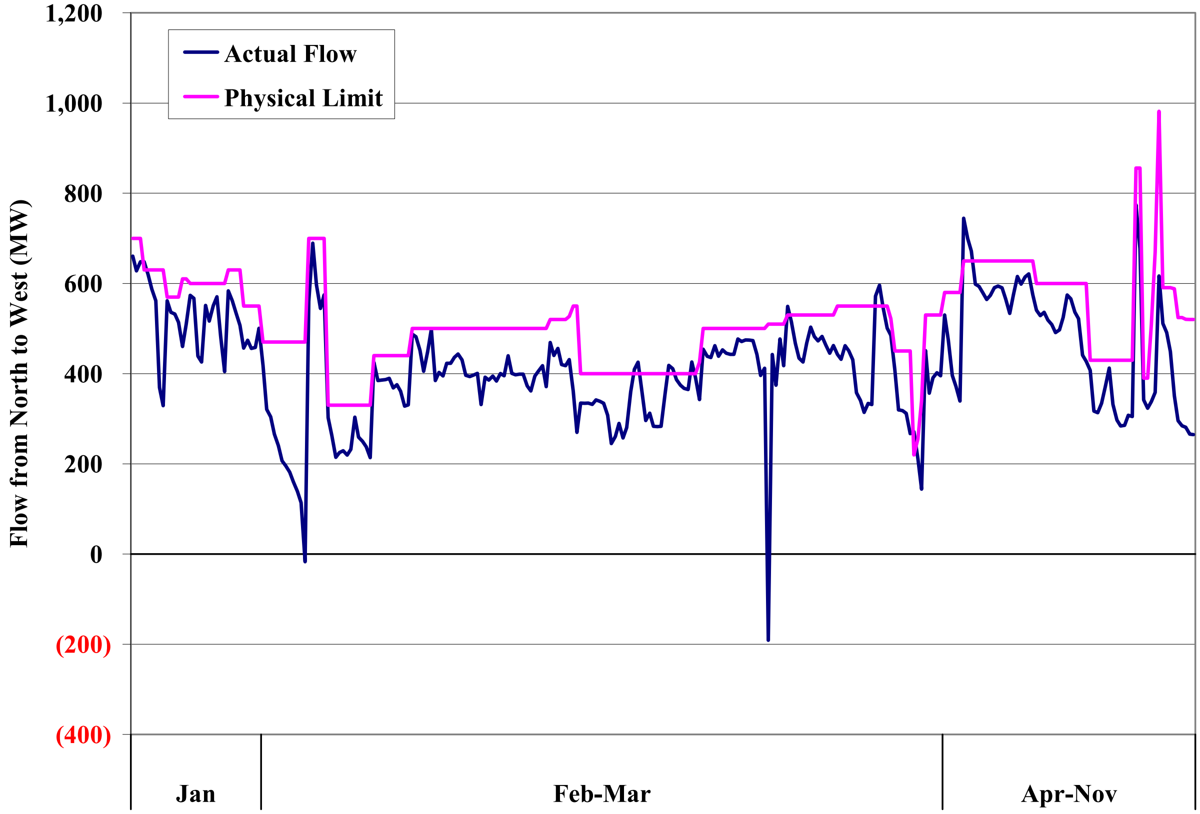


Figure 54 shows that the majority of the binding intervals on the North to West CSC occurred during the first quarter of 2008. The transmission outages discussed in relation to the West to North CSC also affected the import capability for North to West in the February and March timeframe. Additionally, as with the North to Houston and North to South CSCs, some of congestion on the North to West CSC was related to constraints that were removed as eligible zonal constraints with the implementation of PRR 764 in June 2008, although the magnitude of congestion on the North to West CSC associated with these issues was much less than for the North to Houston and North to South CSCs.

4. Congestion on South to North CSC

The South to North CSC was binding in 571 intervals (1.6 percent) in 2008. Figure 55 shows the actual flows and the physical limit for the South to North CSC in 2008 for intervals in which the CSC was binding

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

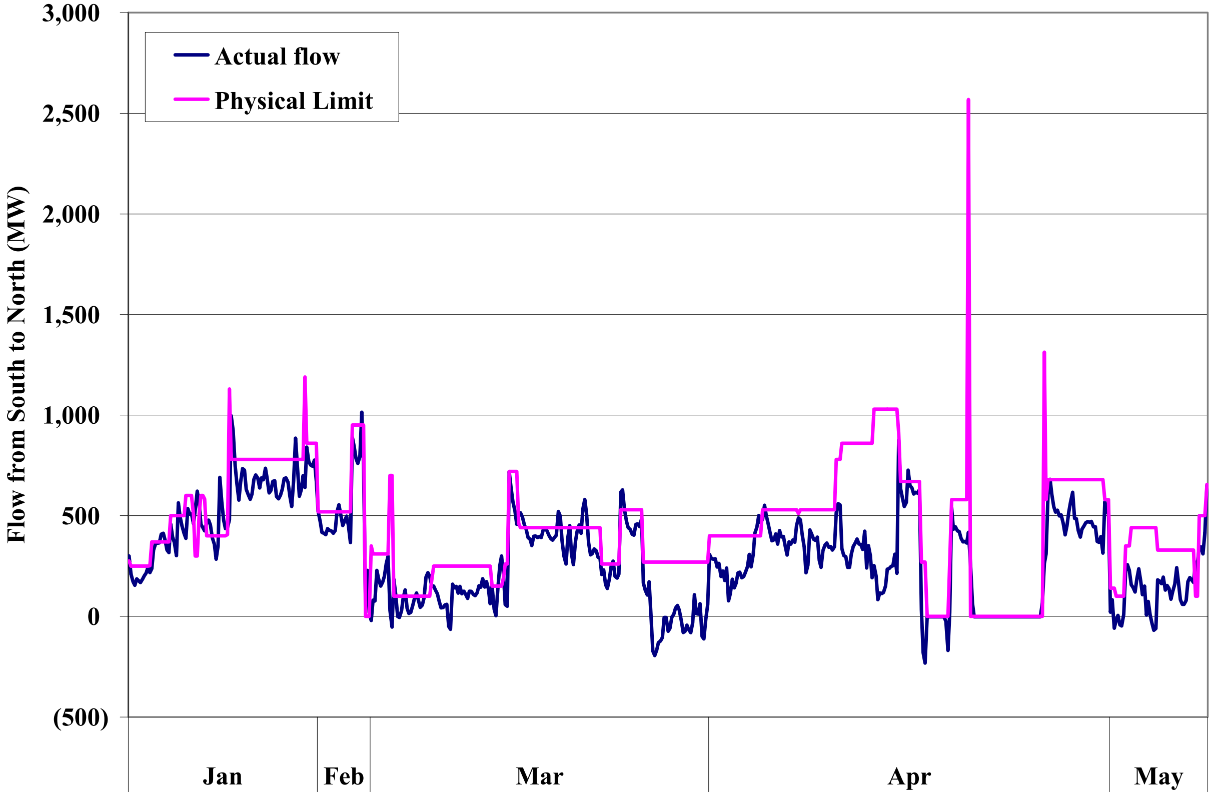


Figure 55 shows that all of the binding intervals on the South to North CSC occurred during the months of January through May 2008. For the South to North CSC, practically all of the congestion during this timeframe was related to constraints that were removed as eligible zonal constraints with the implementation of PRR 764 in June 2008.

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 2008.

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

CSC 2008	Average Physical Limit (MW)	Average Actual Flow (MW)	Avg. Physical Limit - Avg. Actual Flow (MW)
North to South	234	139	94
North to Houston	1,948	1,655	293
South to North	444	297	148
West to North	733	568	165
North to West	516	419	97

Table 4 shows that, for all CSCs in 2008, the average actual flow was considerably less than the average physical limit. For all CSCs combined, the average actual flow was 21 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.

The nodal market will provide many improvements, including unit-specific offers and shift factors, 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 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 2008. Figure 56 shows the average number of TCRs and PCRs awarded for each of the CSCs in 2008 compared to the average SPD-modeled flows during the constrained intervals.

Figure 56: Transmission Rights vs. Real-Time SPD-Calculated Flows Constrained Intervals - 2008

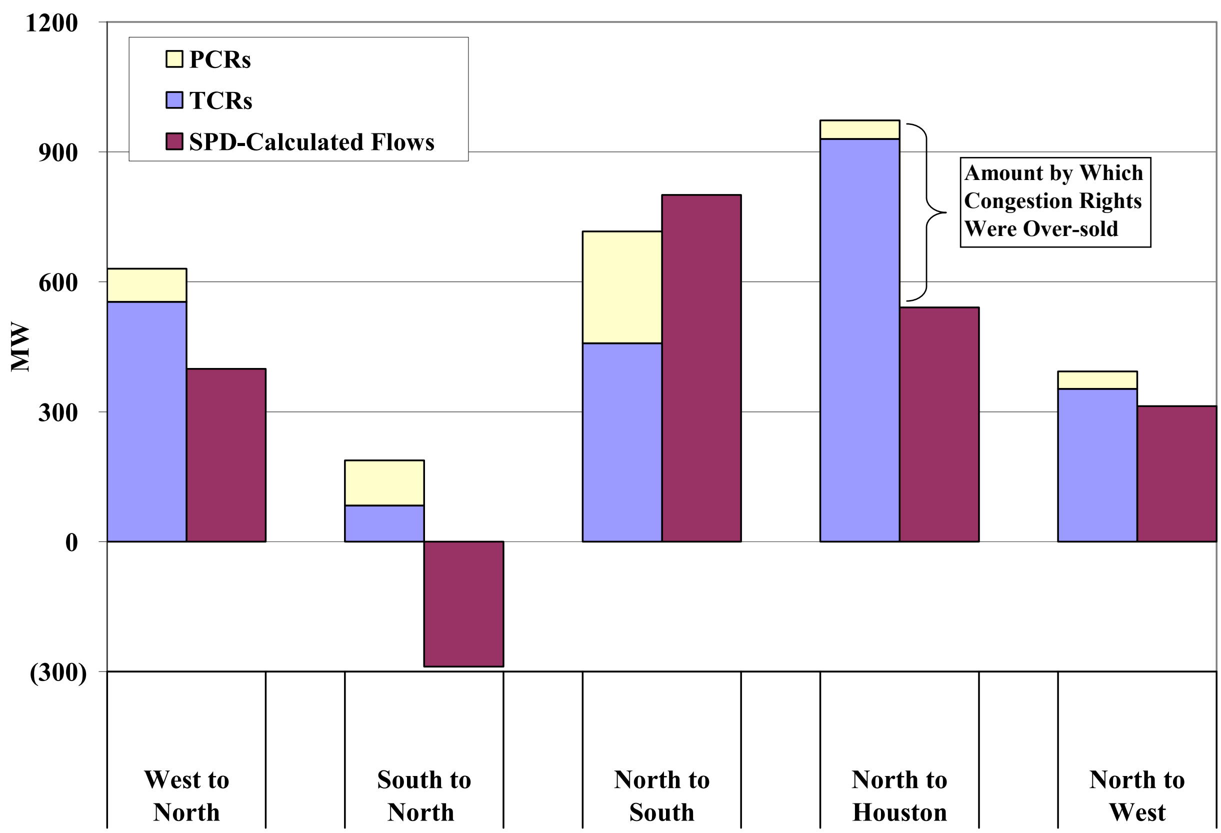


Figure 56 shows that total congestion rights (the sum of PCRs and TCRs) on all the interfaces exceeded the average real-time SPD-calculated flows during constrained intervals except for the North to South CSC. These results indicate that the congestion rights were oversold in relation to the SPD-calculated limits for most CSCs, even though fewer TCRs were awarded in 2008 than in 2007. For example, congestion rights for the North to Houston CSC were oversold by an average of 432 MW.

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.

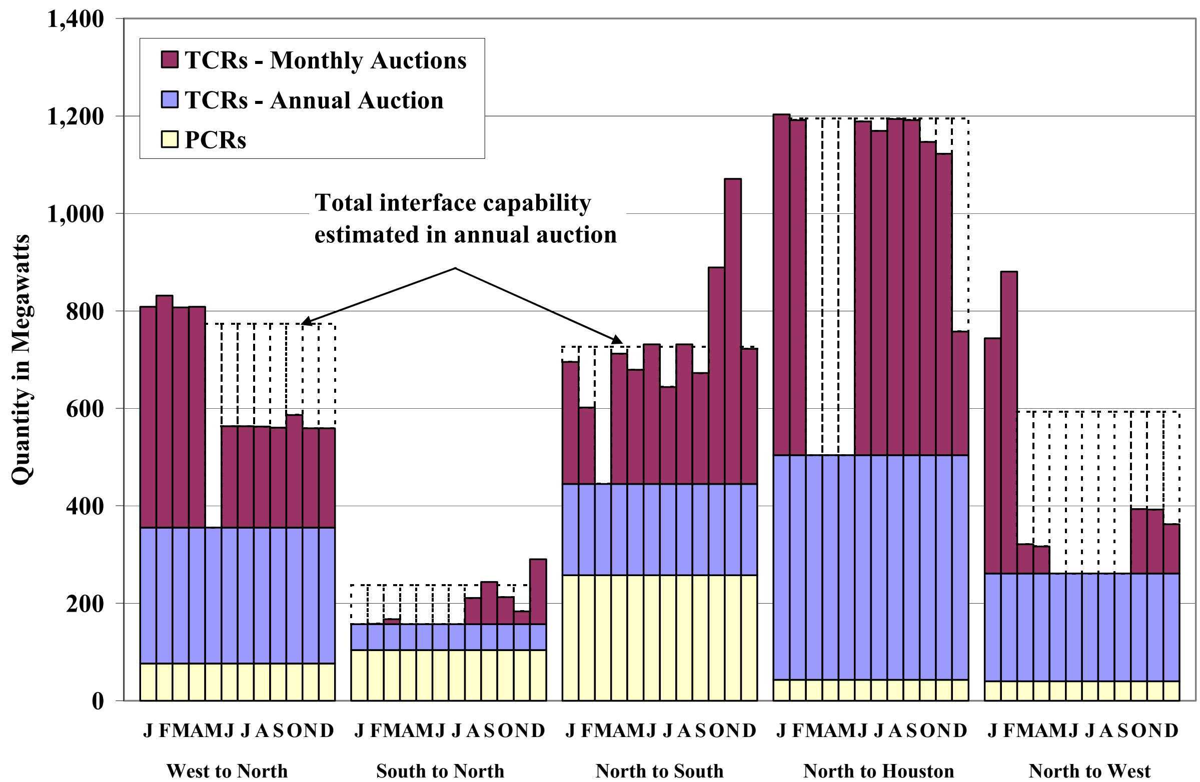
As an example, suppose the SPD-calculated flow limit is 300 MW for a particular CSC during a constrained interval and that holders of congestion rights own a total of 800 MW over the CSC. ERCOT will receive congestion rents from the balancing energy market to cover precisely 300 MW of the 800 MW worth 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 that cause 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 57 shows the quantity of each category of congestion rights for each month during 2008. The quantities of PCRs and annual TCRs are constant across all months and were determined before the beginning of 2008, while monthly TCR quantities can be adjusted monthly.

Figure 57: Quantity of Congestion Rights Sold by Type

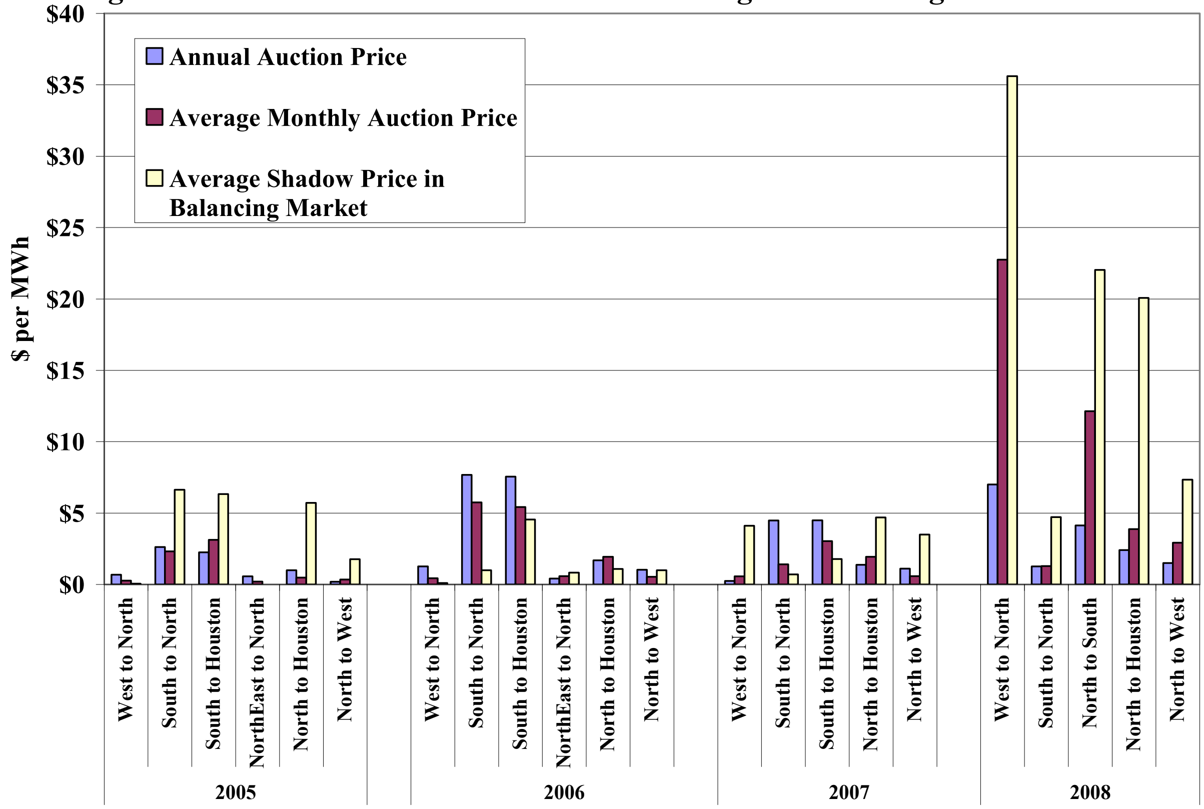


When the monthly planning studies indicate changes from the summer study, revisions are often made to the estimated transmission capability. Therefore, the auctioned congestion rights may increase or decrease relative to the amount estimated in the summer study. The shadow boxes in the figure represent the capability estimated in the summer study that is not ultimately sold in the monthly auction. When there is no shadow box in Figure 57, 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, North to West 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 and North to West 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 58 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.

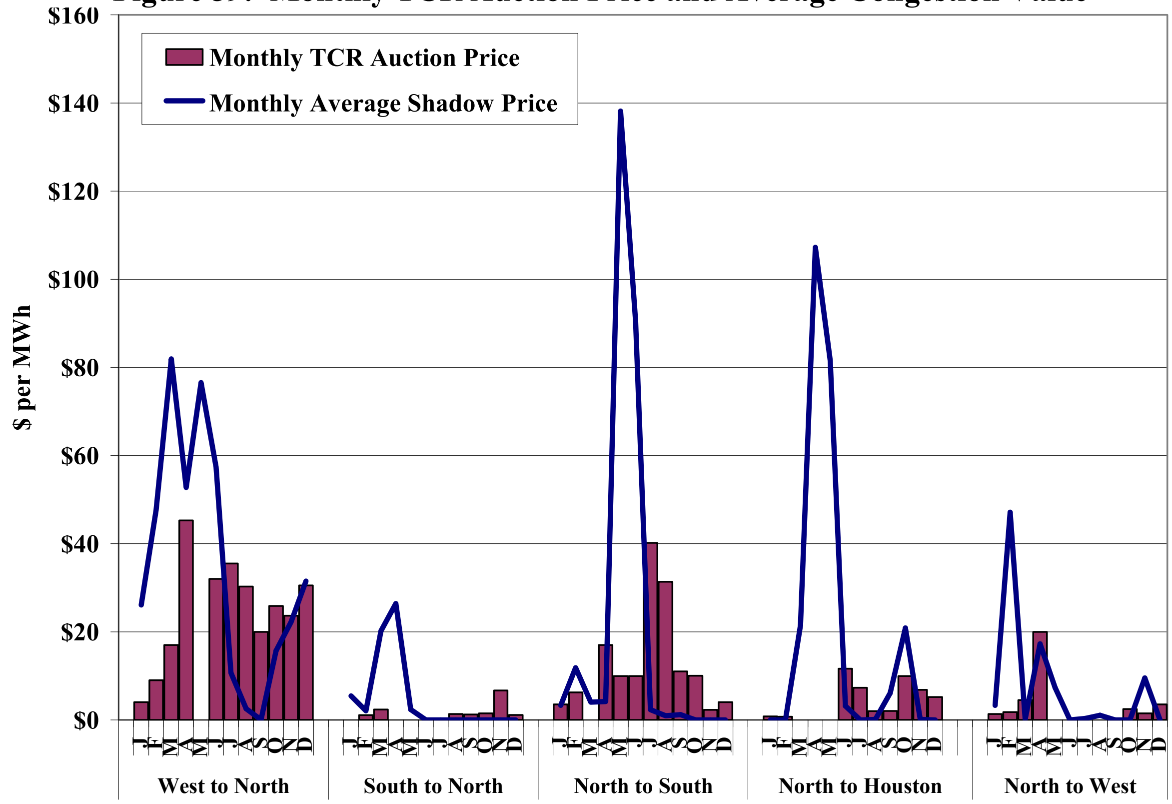
Figure 58: TCR Auction Prices versus Balancing Market Congestion Prices



This figure shows that the TCR annual auction prices were significantly lower than the value of congestion in real-time in 2008. This suggests that participants are not able to forecast annual interzonal congestion costs and accurately value the TCRs in the annual auction, and instead rely more upon historical market outcomes. Monthly TCR auction prices were more consistent with real-time congestion prices; however, real-time shadow price exceeded the average monthly auction price for all the five CSCs in 2008.

Figure 59 compares monthly TCR auction prices with monthly average real-time CSC shadow prices from SPD for 2008. The TCR auction prices are expressed in dollars per MWh.

Figure 59: Monthly TCR Auction Price and Average Congestion Value



The significant divergence in the monthly TCR auction prices and the real-time shadow prices indicates that market participants did a poor job predicting and valuating the real-time cost of zonal congestion in 2008. These outcomes appear to reflect the existence of flaws in the congestion management procedures that were not recognized until ERCOT had difficulty managing interzonal congestion in the spring of 2008. Just as ERCOT was unaware of these flaws, market participants could not anticipate that problems with the congestion management system would lead to the high shadow prices that occurred in April through June 2008.

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.[22] 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 59 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.

22 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 60: TCR Auction Revenues, Credit Payments, and Congestion Rent

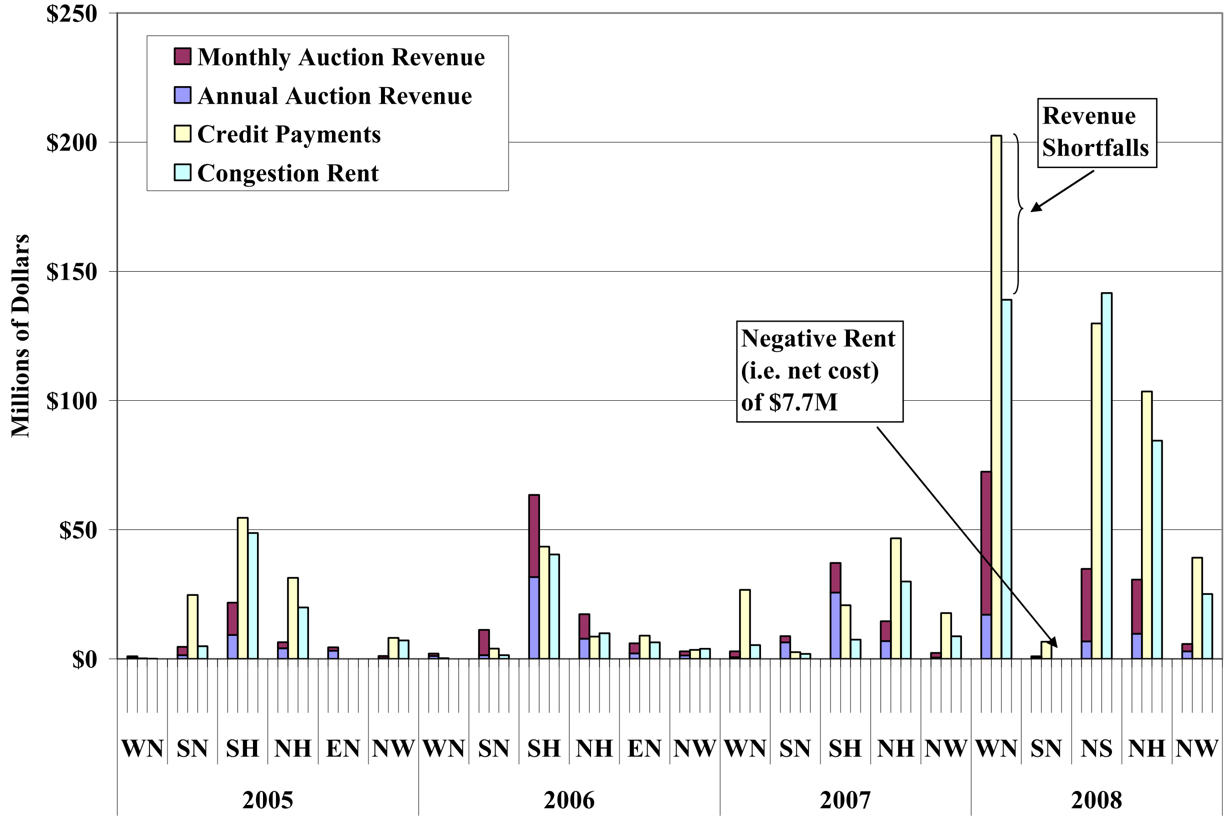


Figure 60 shows that the West to North, North to Houston and North to West CSCs each had significant revenue shortfalls in 2008. When congestion rents fall significantly below payments to TCR holders, it implies that the SPD-calculated flows across constrained interfaces have been systematically lower than the amount of TCRs sold for the interfaces. In contrast, the North to South CSC produced a slight surplus in 2008 (*i.e.*, congestion rents exceeded TCR credit payments), which indicates that the North to South CSC had SPD-calculated flows that were generally greater than the amount of TCRs sold. The South to North CSC actually experienced a negative $7.7 million congestion rent in 2008, which not only produces a revenue shortfall, but also shows that the average SPD-calculated flow across this constraint when it was binding was negative when the actual flows were positive, which is a product of the issues associated with the use of zonal average shift factors as described in the first part of this Section.

Figure 60 also shows that payments to TCR holders have consistently exceeded the congestion rents that have been collected from the balancing market since the creation of the TCR market. Congestion rents covered 68, 90 and 47 percent of payments to TCR holders in 2005, 2006 and

2007, respectively. Congestion rents covered 79 percent of the payments to TCR holders in 2008, with an annual net revenue shortfall of $99 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 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.

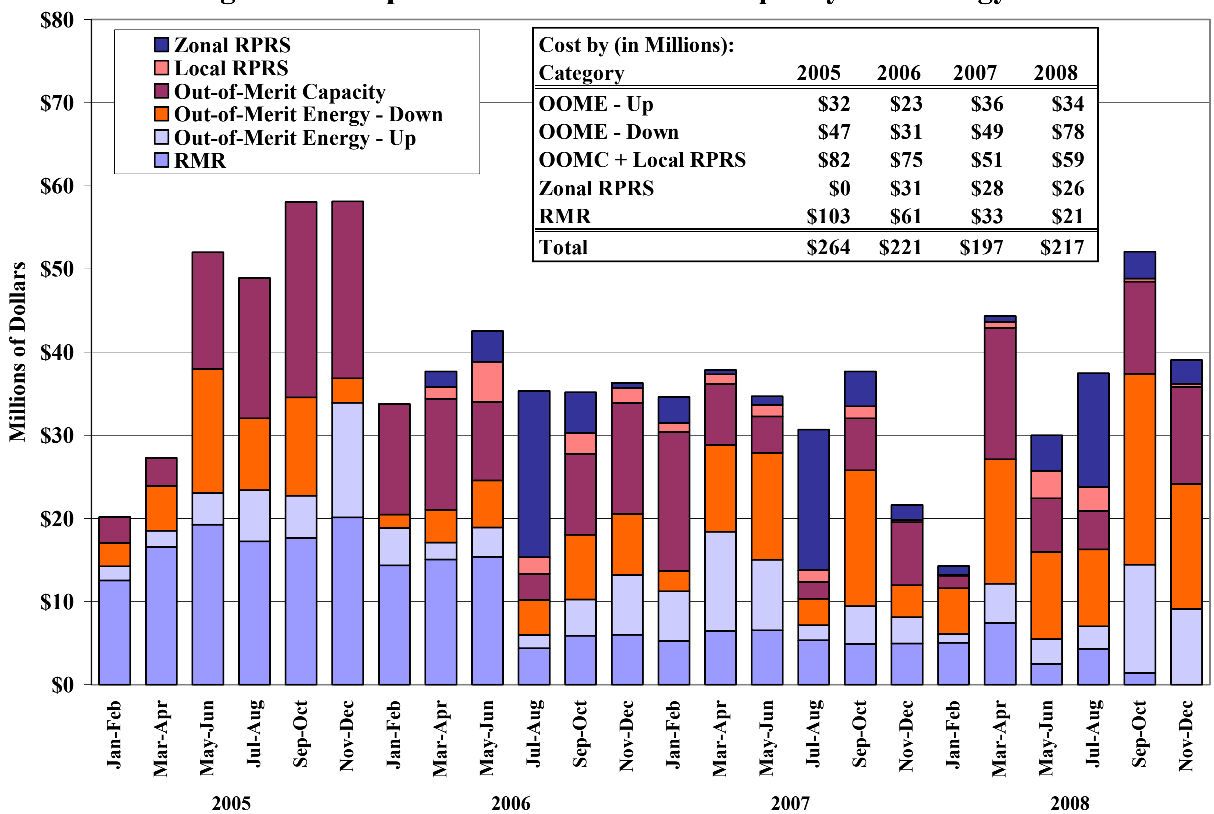
When a unit is dispatched out of merit (OOME up or OOME down), the unit is paid for a quantity equal to the difference between the scheduled output based on the unit's resource plan and the actual output resulting from the OOME instruction from ERCOT. The payment per MWh for OOME is a pre-determined 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. However, as a part of the RMR exit strategy process, all units were removed from RMR status by October 2008. Units contracted to provide RMR service to ERCOT are compensated for start-up costs, energy costs, and are also paid a standby fee. Figure 61 shows each of the four categories of uplift costs from 2005 to 2008.

Figure 61: Expenses for Out-of-Merit Capacity and Energy



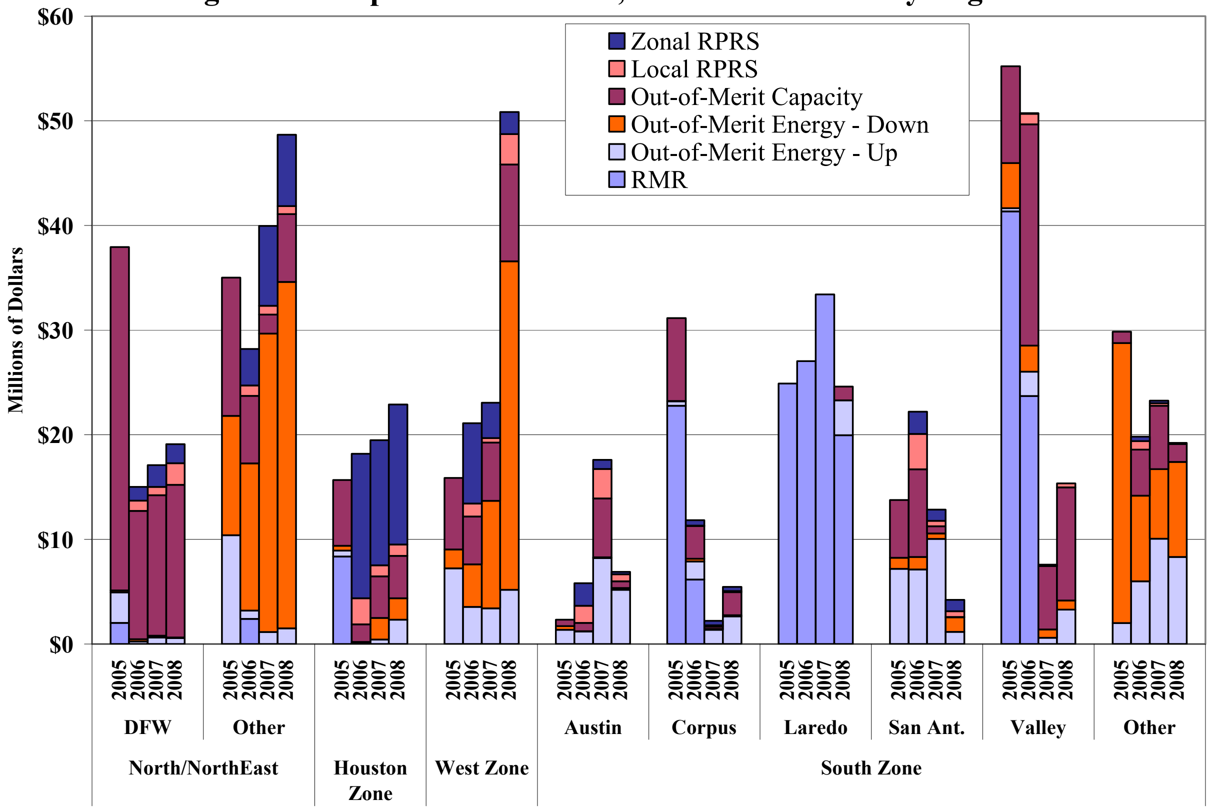
The results in Figure 61 show that overall uplift costs for RMR units, OOME units, OOMC/Local RPRS and Zonal RPRS[23] units were $217 million in 2008, which is a $20 million increase over the $197 million in 2007. OOME Down costs accounted for the most significant portion of the change in 2008, increasing from $49 million in 2007 to $78 million in 2008.

23 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. The RPRS procurement tool was not in production in 2005, thus all capacity procurements were conducted via OOMC in 2005.

OOMC/Local RPRS costs increased from $50 million in 2007 to $60 million in 2008, and RMR costs decreased from $33 million in 2007 to $20 million in 2008. Figure 61 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 reliability. The rest of the analyses in this section evaluate in more detail where these costs were caused and how they have changed between 2005 and 2008. Figure 62 shows these payments by location.

Figure 62: Expenses for OOME, OOMC and RMR by Region



The most significant changes in 2008 compared to 2007 (*i.e.*, an increase or decrease of more than $5 million in a category by location) shown in Figure 62 are as follows:

- OOME Down costs in the West Zone increased by $21 million in 2008. This increase was associated with the significant addition of wind capacity in the West Zone.
- OOMC and OOME Down costs in the North Zone each increased by $5 million in 2008. These increases were associated with local reliability requirements near the load centers

in the North Zone, and with transmission outages resulting in the OOME Down of coal/lignite units.

- OOMC costs in the Valley area of the South Zone increased by $5 million in 2008. This increase was associated with the more frequent need for local capacity to be online to maintain Valley import limits.
- RMR costs in the Laredo area of the South Zone decreased by $13 million in 2008. This decrease was associated with the termination of the Laredo RMR contract in October 2008.
- OOME Up costs in the San Antonio area of the South Zone decreased by $9 million in 2008, and OOMC costs in the Austin area of the South Zone decreased by $5 in 2008. These decreases were associated with the completion of transmission improvements between Austin and San Antonio.

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 2008. We examine market structure using a pivotal supplier analysis, which indicates that suppliers were pivotal in the balancing energy market at a frequency in 2008 that was similar to 2007, but much less than 2005 and 2006. This analysis also shows that the frequency with which a supplier was pivotal increased at higher levels of demand, which is consistent with 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 2008.

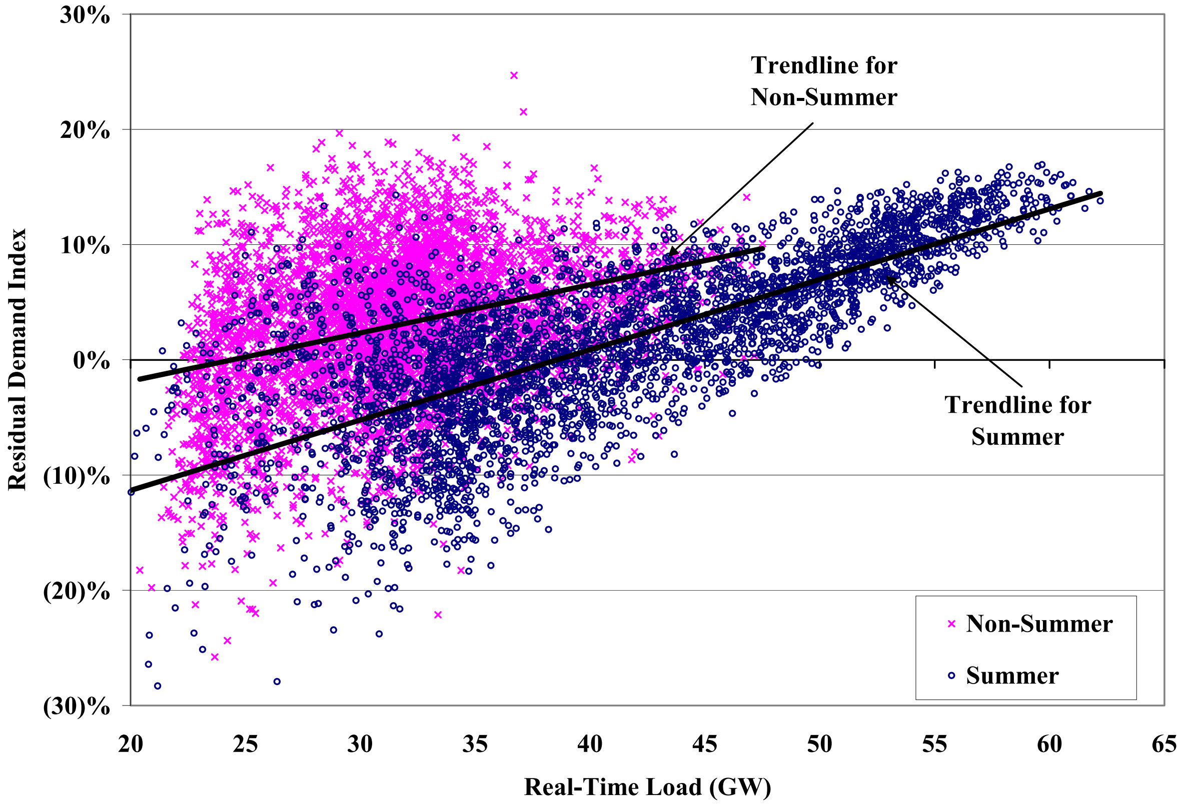
A. Structural Market Power Indicators

We analyze market structure 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 63 shows the RDI relative to load for every hour in 2008. 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.

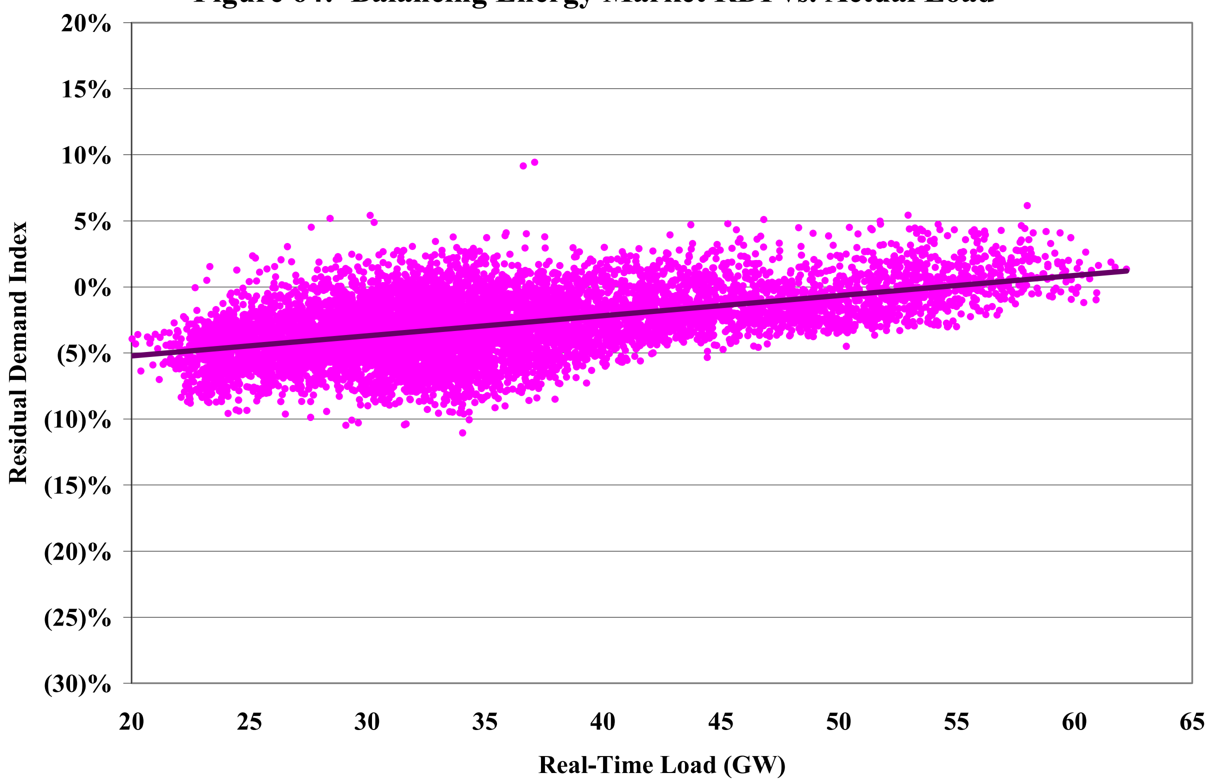
Figure 63: Residual Demand Index



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 60 percent of hours. The RDI was typically positive 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 70 percent of hours during the non-summer period. 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 64 shows the RDI in the balancing energy market relative to the actual load level.

Figure 64: Balancing Energy Market RDI vs. Actual Load

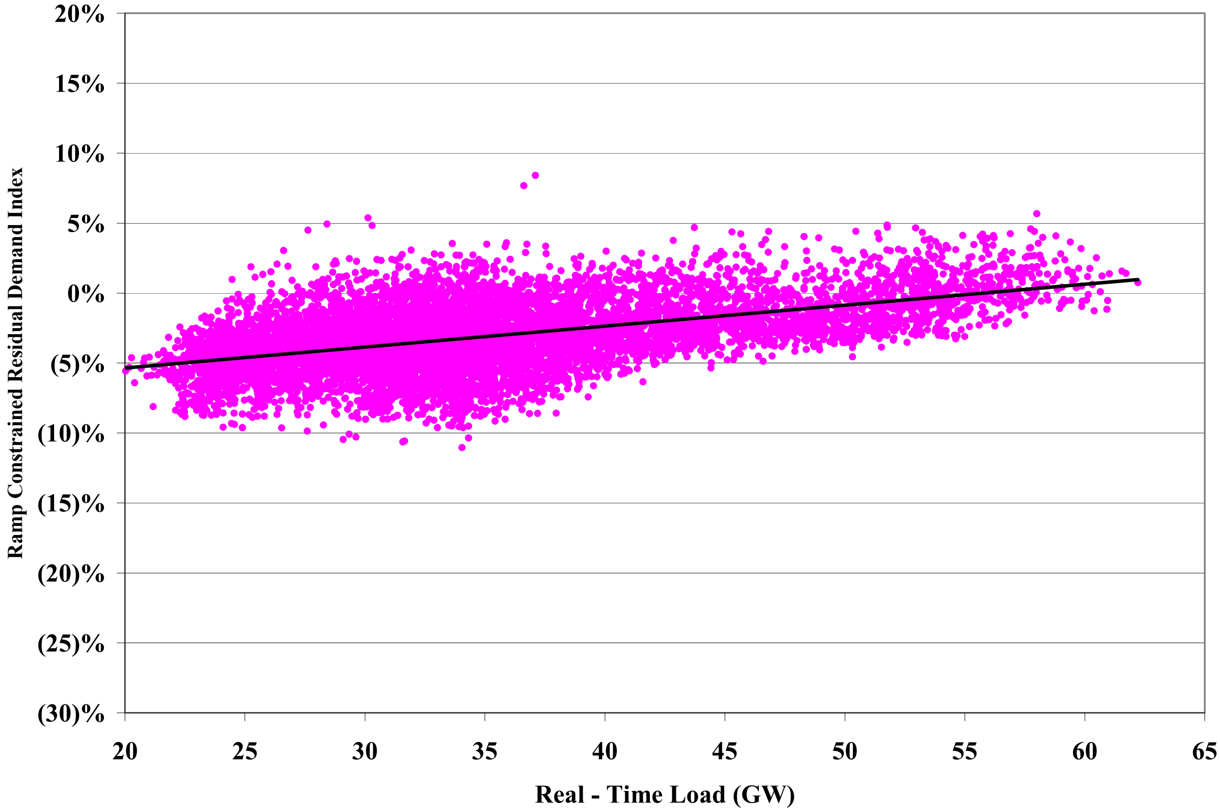


Ordinarily, the RDI is used to measure the percentage of load that cannot be served without the resources of the largest supplier, assuming that the market could call upon all committed and quick-start capacity[24] owned by other suppliers. Figure 64 limits the other supplier's capacity to

24 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

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 65 shows the same data as in Figure 64 except that the balancing energy offers are further limited by portfolio ramp constraints in each interval.

Figure 65: Ramp-Constrained Balancing Energy Market RDI vs. Actual Load

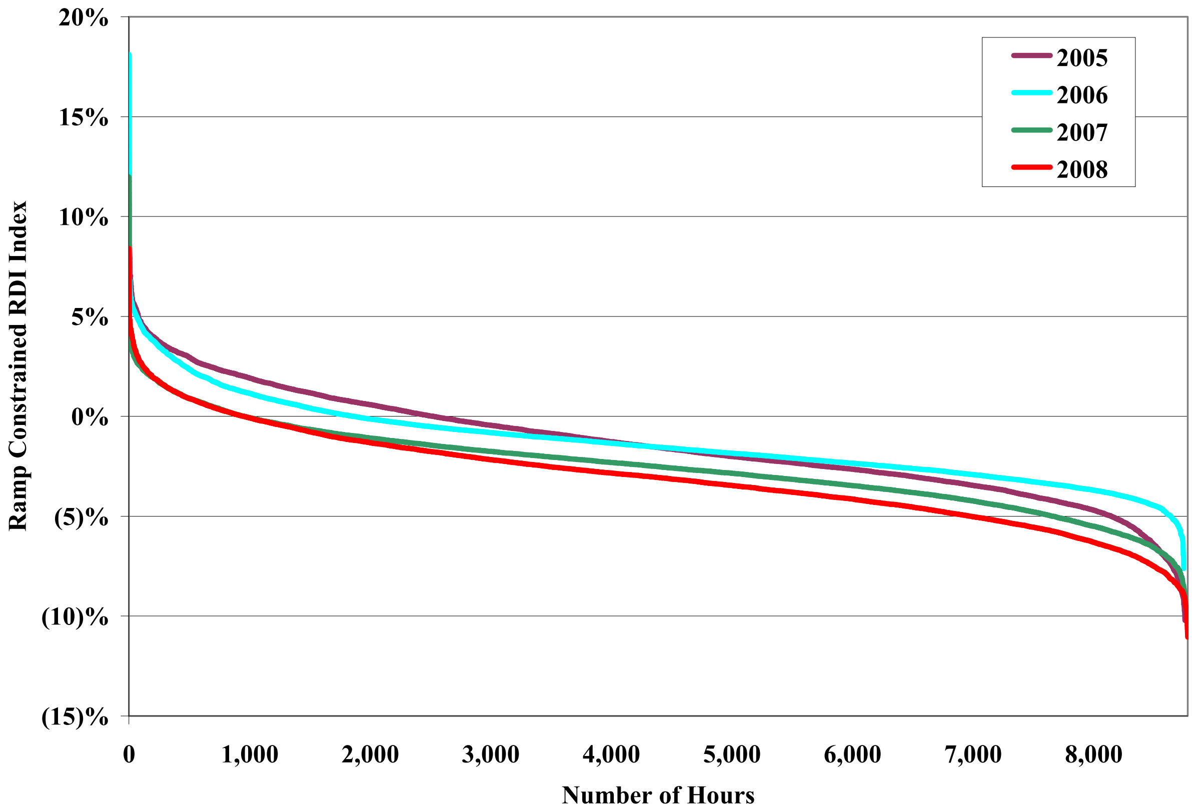


In 2008, the instances when the RDI was positive occurred over a wide range of load levels, from 26 GW to 63 GW. The RDI results for the balancing energy market shown in the preceding two figures help explain how transient price spikes can occur under mild demand while large amounts of capacity are available in ERCOT. The balancing energy market RDI data and trend line for 2008 are similar in shape to 2007. The frequency of data points that are positive in 2008 is similar to the frequency in 2007 as well, although the frequency of data points that are positive is significantly lower in 2007 and 2008 than in 2005 and 2006. This difference

of starting-up and reaching full output after receiving a deployment instruction from the balancing energy market.

is highlighted in Figure 66 which compares the balancing energy market RDI duration curves for 2005 through 2008.

Figure 66: 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 from 29 percent of the hours in 2005, to 21 percent of the hours in 2006, and to less than 11 percent of the hours in 2007 and 2008. These results indicate that the structural competitiveness of the balancing energy market in 2008 maintained the improvement exhibited in 2007 compared to prior years.

B. Evaluation of Supplier Conduct

The previous sub-section 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.

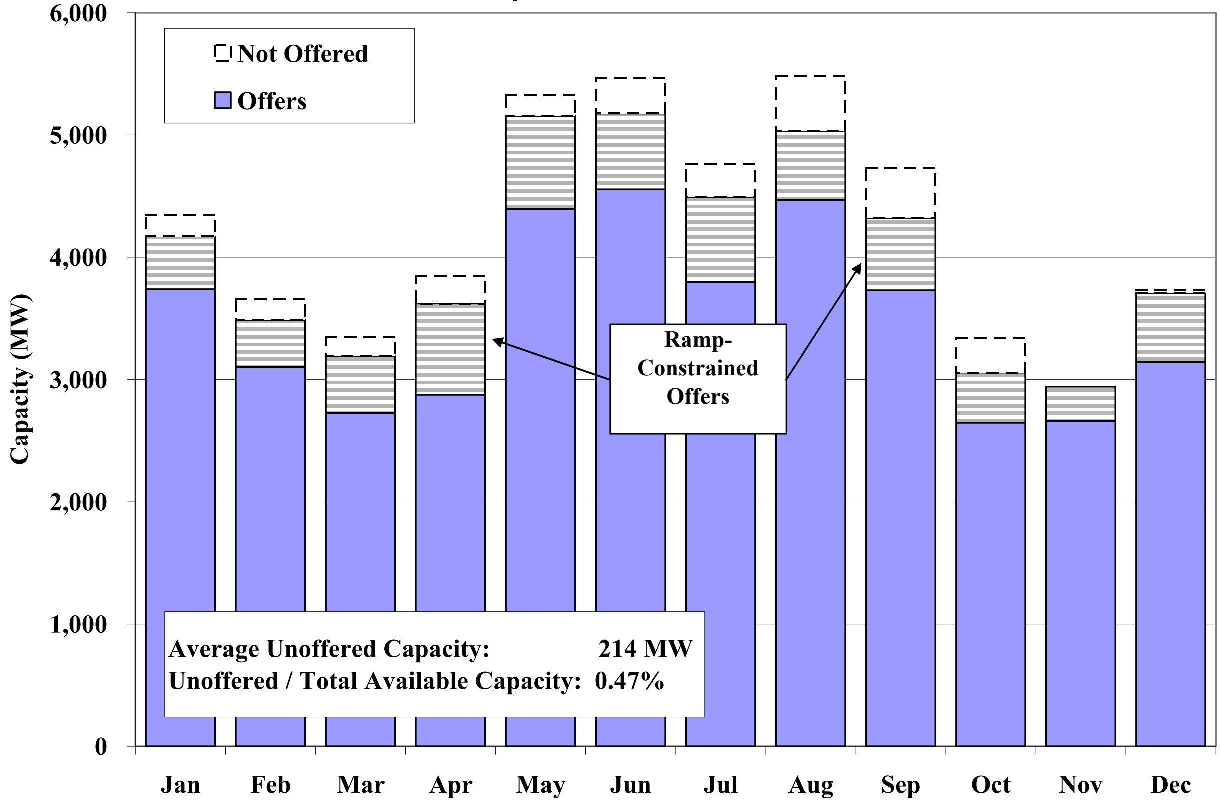
1. Balancing Energy Market Offer Patterns

In this section, we evaluate balancing energy offer patterns by analyzing the rate at which capacity is offered.[25] Figure 67 shows the average amount of capacity offered to supply up balancing service relative to all available capacity.

Figure 67 shows only slight variation in 2008 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").

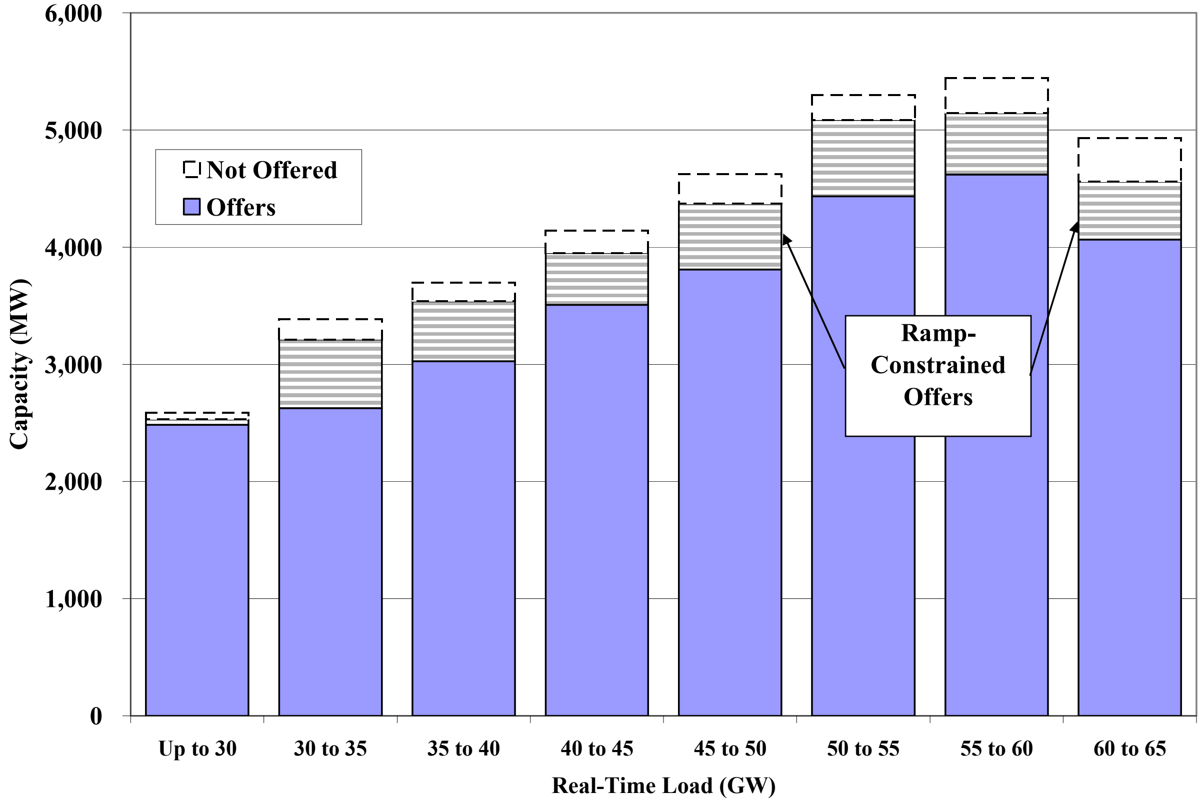
[25] The methodology for determining the quantities of un-offered capacity is detailed in the 2006 SOM Report (2006 SOM Report at 63-65).

Figure 67: Balancing Energy Offers Compared to Total Available Capacity Daily Peak Load Hours



Un-offered energy 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 68 shows the same data as the previous figure, but arranged by load level for daily peak hours in 2008. 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 68: Balancing Energy Offers Compared to Total Available Capacity Daily Peak Load Hours



The figure indicates that in 2008 the average amount of capacity available to the balancing market increased gradually up to 60 GW of load and then declined at higher levels. The decline in balancing energy available at higher load levels is associated with the fact that scheduled generation increases at higher load levels, thereby leaving less residual capacity available to be offered as balancing energy. As indicated in the figure, the quantity of un-offered capacity increases slightly as load levels increase, although the quantity of un-offered capacity remains relatively flat as a percentage of system demand.

The pattern of un-offered capacity shown in Figure 68 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 do not change significantly 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 63 through Figure 65 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 69 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 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 69: Short-Term Deratings by Load Level and Participant Size
June to August, 2008

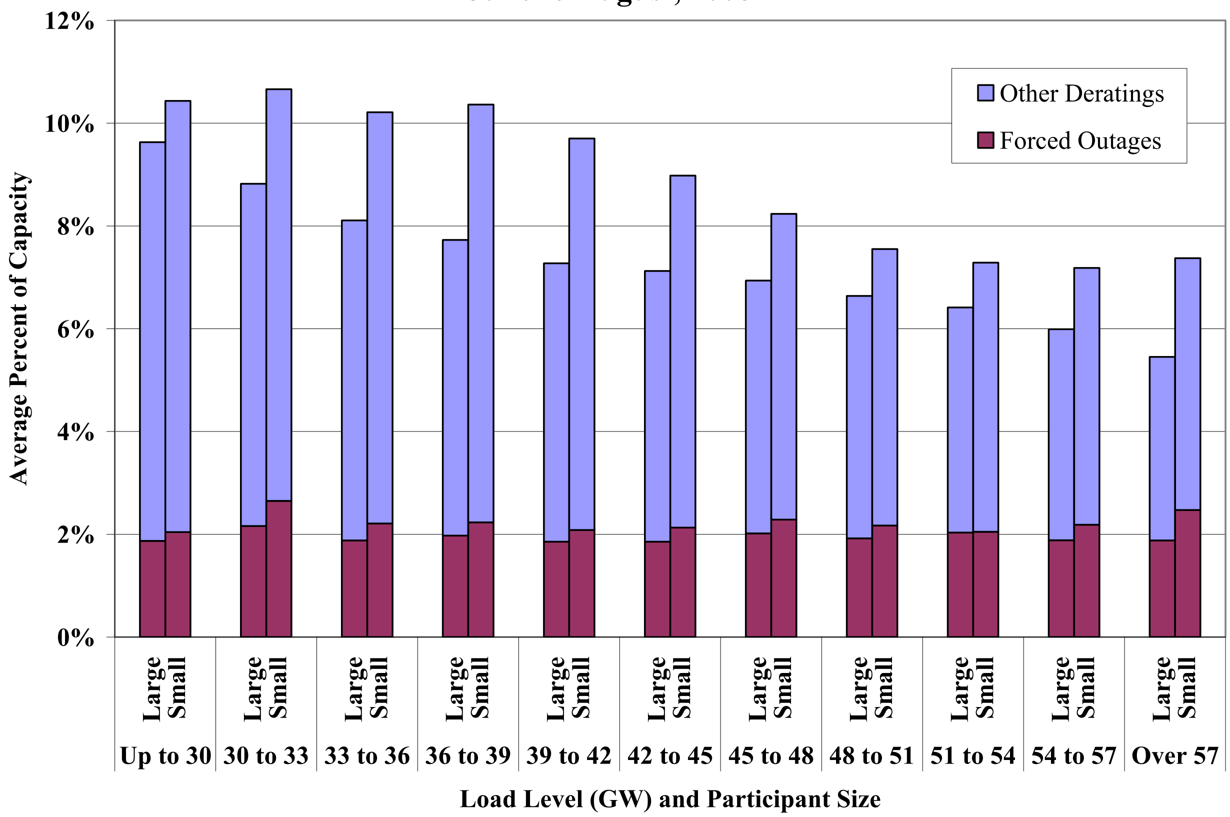


Figure 69 suggests that as electricity demand increases, both large and small market participants tend to make more capacity available to the market. For both large and small suppliers, the combined short-term derating and forced outage rates decreased from approximately 10 percent at low demand levels to about 5 and 7 percent respectively 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. Furthermore, large suppliers' deratings and outages generally decline as load levels increase. 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.[26]

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 70 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 2008.

[26] 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 70: Incremental Output Gap by Load Level

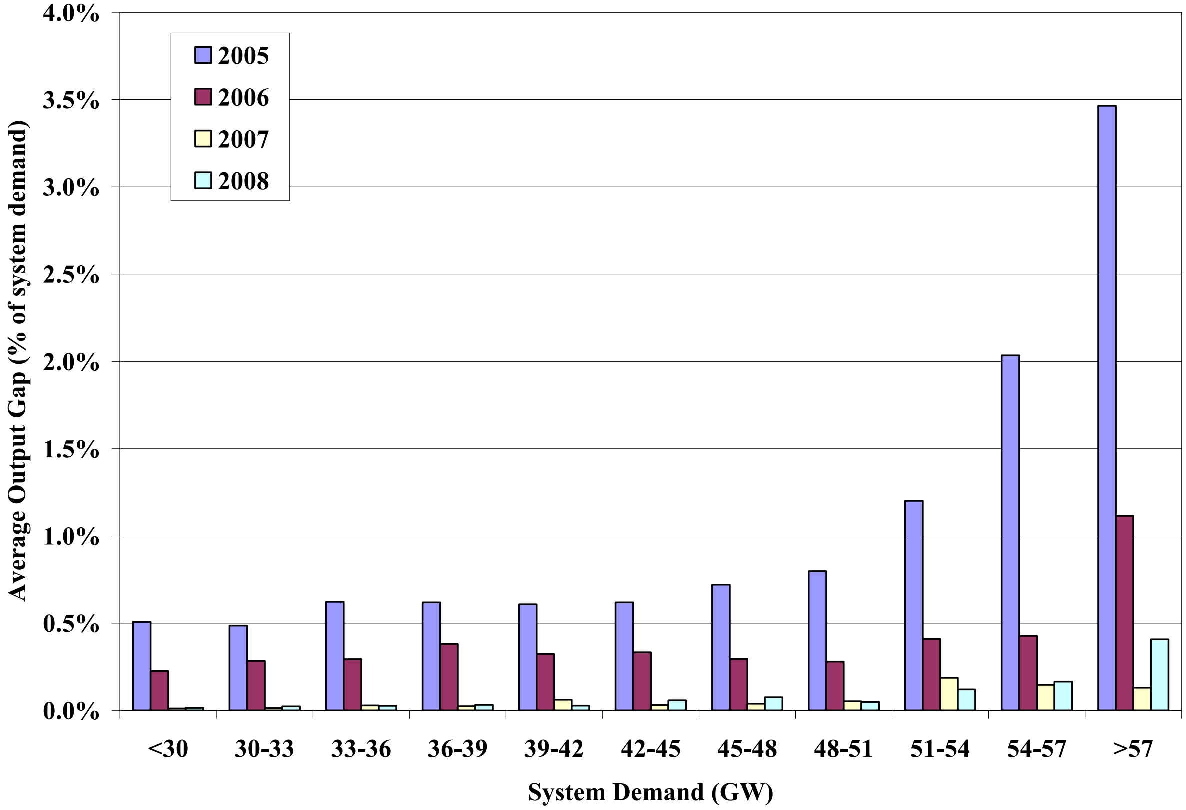


Figure 70 shows that the competitiveness of supplier offers improved considerably in 2006 compared to 2005, followed by even more substantial improvement in 2007 and 2008. Although 2008 exhibited a higher average incremental output gap at the highest load levels, the overall magnitude remains small and does not raise significant economic withholding concerns.

Figure 71 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 71: Output Gap by Load Level and Participant Size

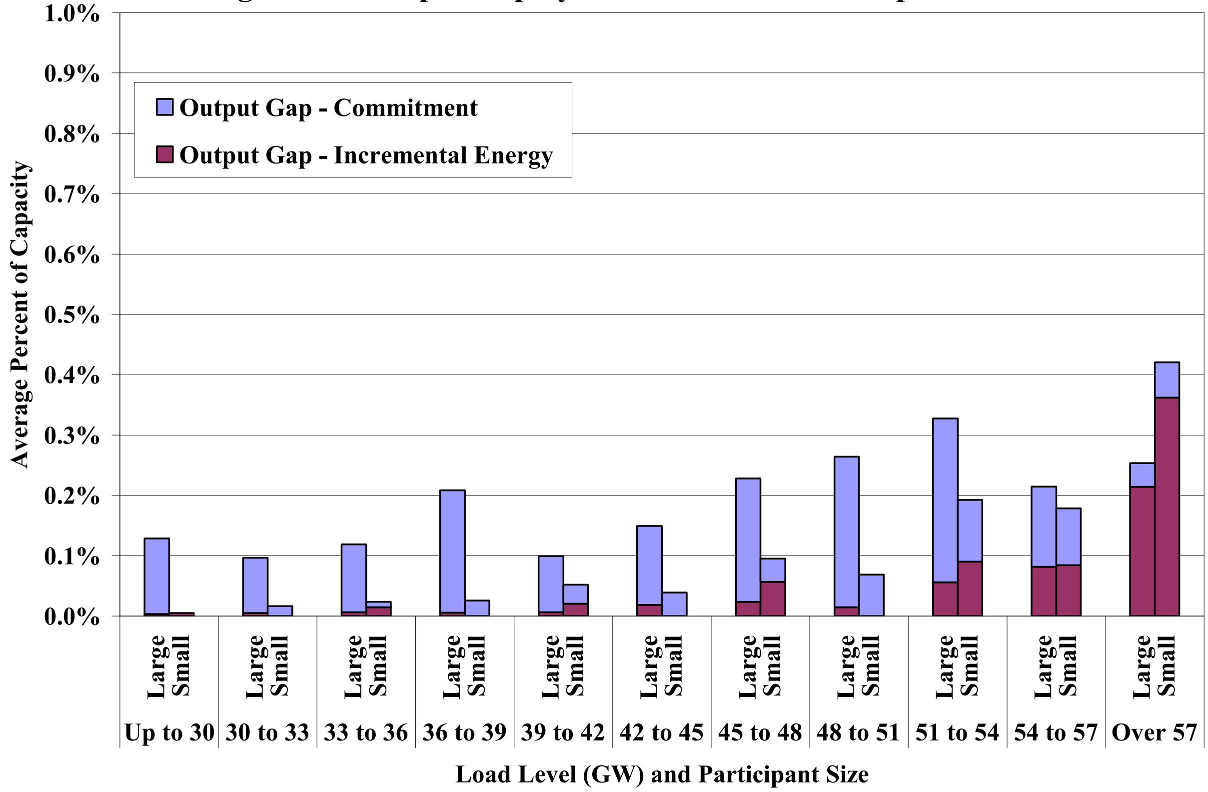


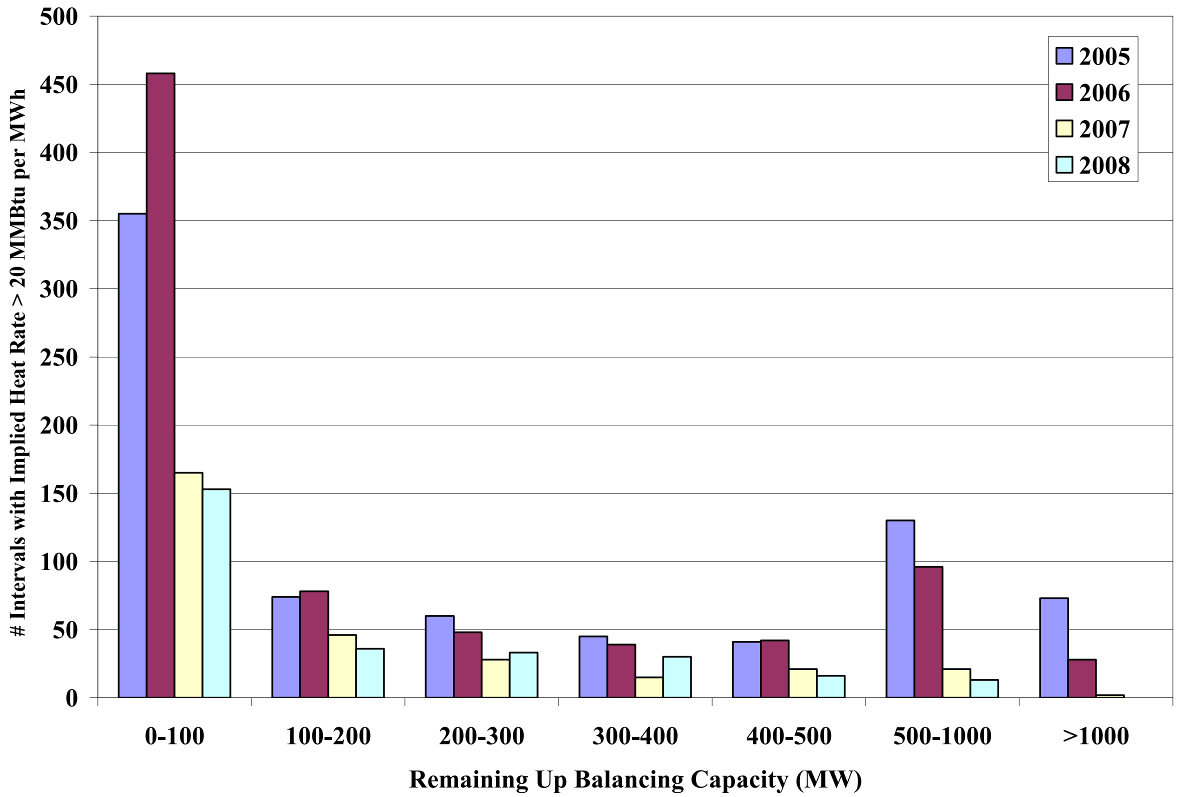
Figure 71 shows that the output gap quantities for incremental energy of large and small suppliers were very low across all load levels. Overall, the output gap measures in 2008 were comparable with the levels in 2007, with both years showing significant improvement over 2005 and 2006.[27] Figure 71 also shows that the increase in the incremental output gap for all market participants in 2008 at the highest load levels shown in Figure 70 is not only small in overall magnitude, but is higher for small participants than for large participants, and therefore does not raise competitive concerns.

A final measure used to evaluate the competitiveness of the market outcomes in 2008 analyzes the number of balancing energy market price spikes compared to the quantity of remaining Up Balancing capacity. If the market is operating competitively, price spikes should occur during shortage and near shortage conditions, and the number of price spikes should reduce significantly as the amount of available surplus capacity increases.

27 *See* 2005, 2006 and 2007 SOM Reports.

For the purpose of this analysis, a price spike is measured as an interval in which the balancing energy market price exceeded an implied heat rate of 20 MMBtu per MWh, which is greater than the marginal costs of most online generating units. However, the marginal cost of offline quick start units is often greater than this threshold. Thus, some of the price spikes in this figure are indicative of the deployment of quick start gas turbines, particularly in 2007 and 2008 when several market participants had well over 1,000 MW of quick start capability qualified to provide balancing energy. In contrast, in 2005 only one market participant had quick start unit qualified to provide balancing energy (Austin Energy; 7 units and approximately 330 MW), and in 2006 one additional market participant had qualified quick start gas turbines (CPS Energy; 4 units and approximately 200 MW).

Figure 72: Price Spikes vs. Remaining Up Balancing Capacity



The results in Figure 72 indicate very competitive market outcomes in 2008, with over 95 percent of the price spikes occurring during intervals with less than 500 MW of Up Balancing capacity remaining.[28] These results show significant improvement over 2005 and 2006 when

28 The data in Figure 72 exclude intervals where there was zonal congestion or when non-spinning reserves were deployed.

only 74 and 84 percent, respectively, of the price spikes occurred during intervals with less than 500 MW of available Up Balancing capacity remaining.

The changes in the market outcomes from 2005 through 2008 shown in Figure 72 are consistent with expectations given the improvements in structural and supplier conduct competitiveness over this timeframe that are highlighted in Figure 66 and Figure 70.

Overall, based upon the analyses in this section, we find that the ERCOT wholesale market performed competitively in 2008.