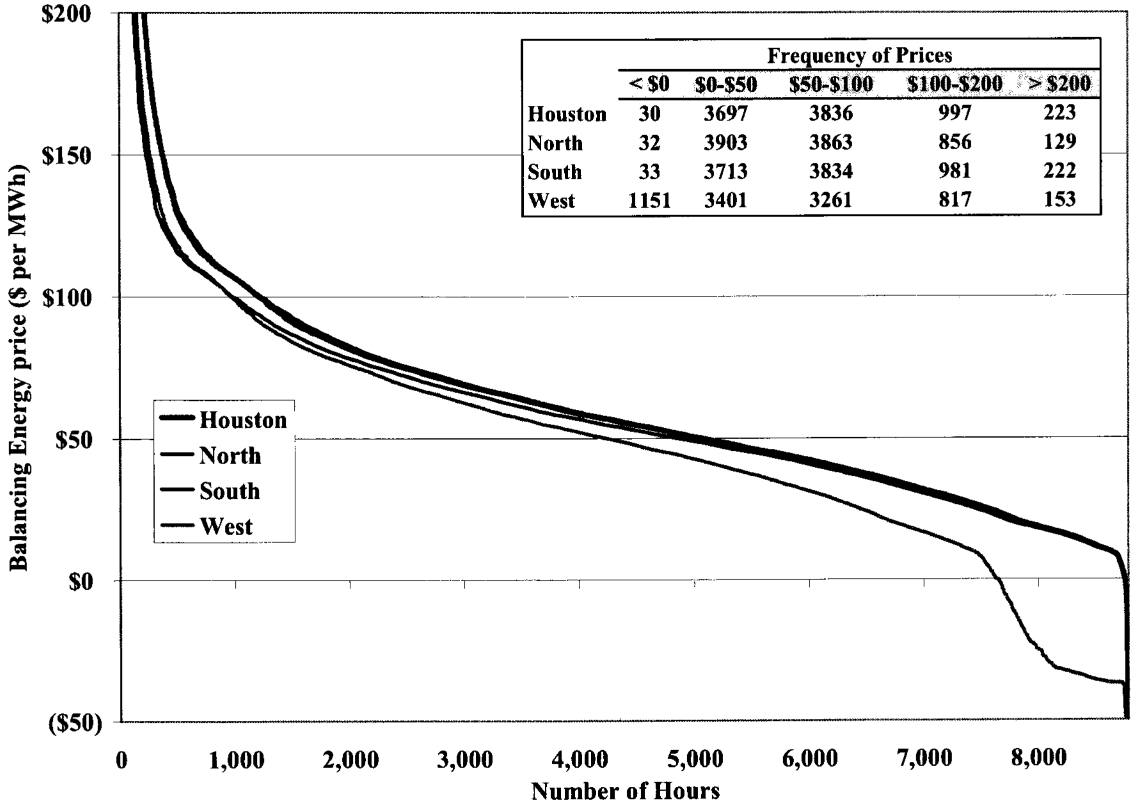
Figure 5: Zonal Price Duration Curves

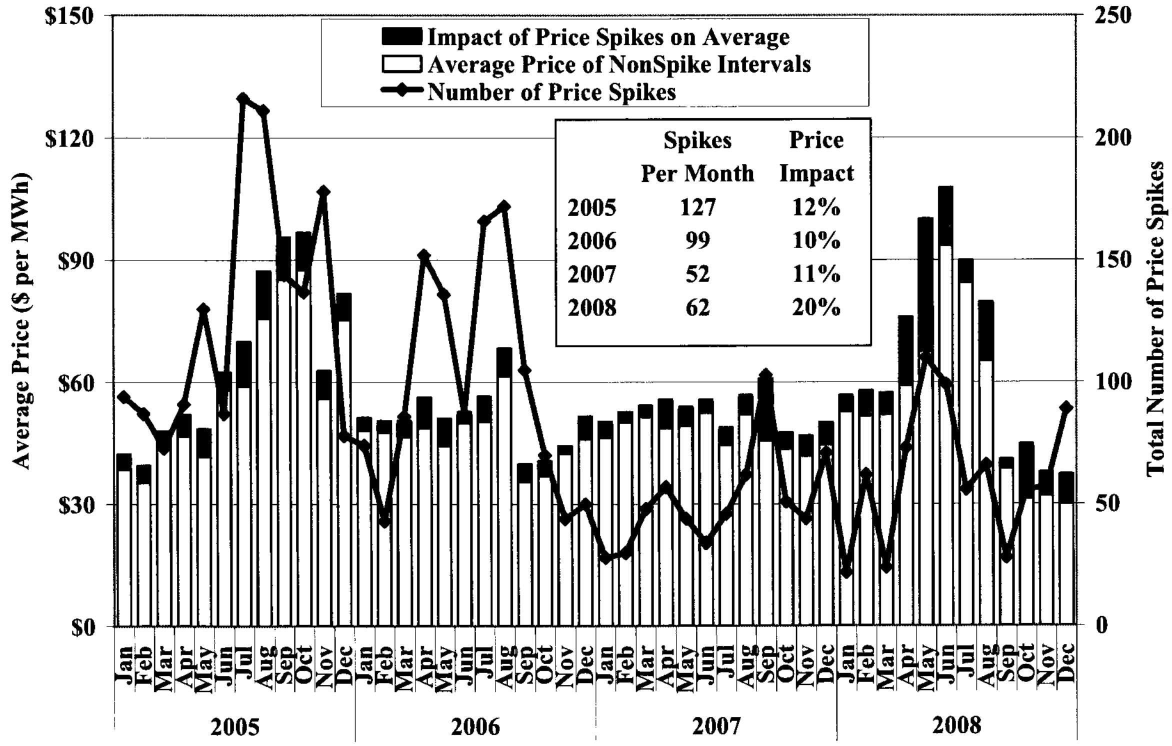


	Frequency of Prices				
	< $0	$0-$50	$50-$100	$100-$200	> $200
Houston	30	3697	3836	997	223
North	32	3903	3863	856	129
South	33	3713	3834	981	222
West	1151	3401	3261	817	153

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



	Spikes Per Month	Price Impact
2005	127	12%
2006	99	10%
2007	52	11%
2008	62	20%

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

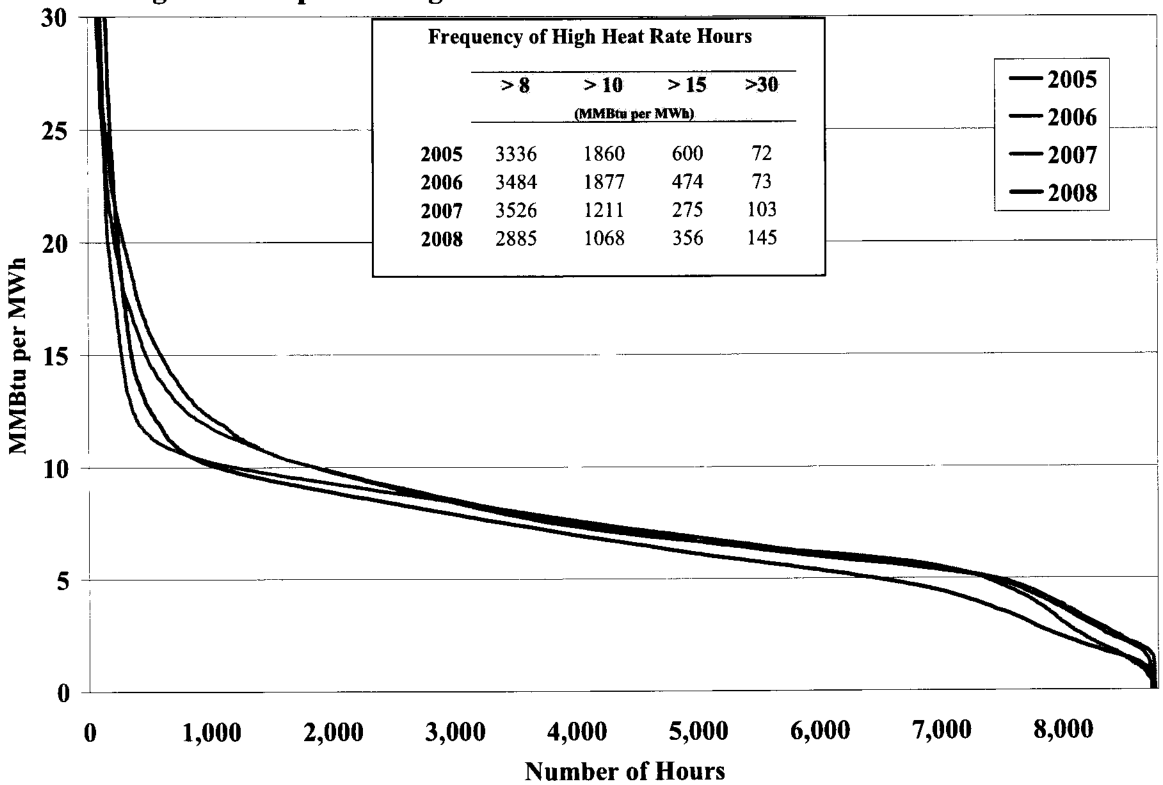
The pricing patterns shown in the prior sub-section are driven to a large extent by changes in fuel prices, natural gas prices in particular. However, prices are influenced by a number of other factors as well. To clearly identify changes in electricity prices that are not driven by changes in natural 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.

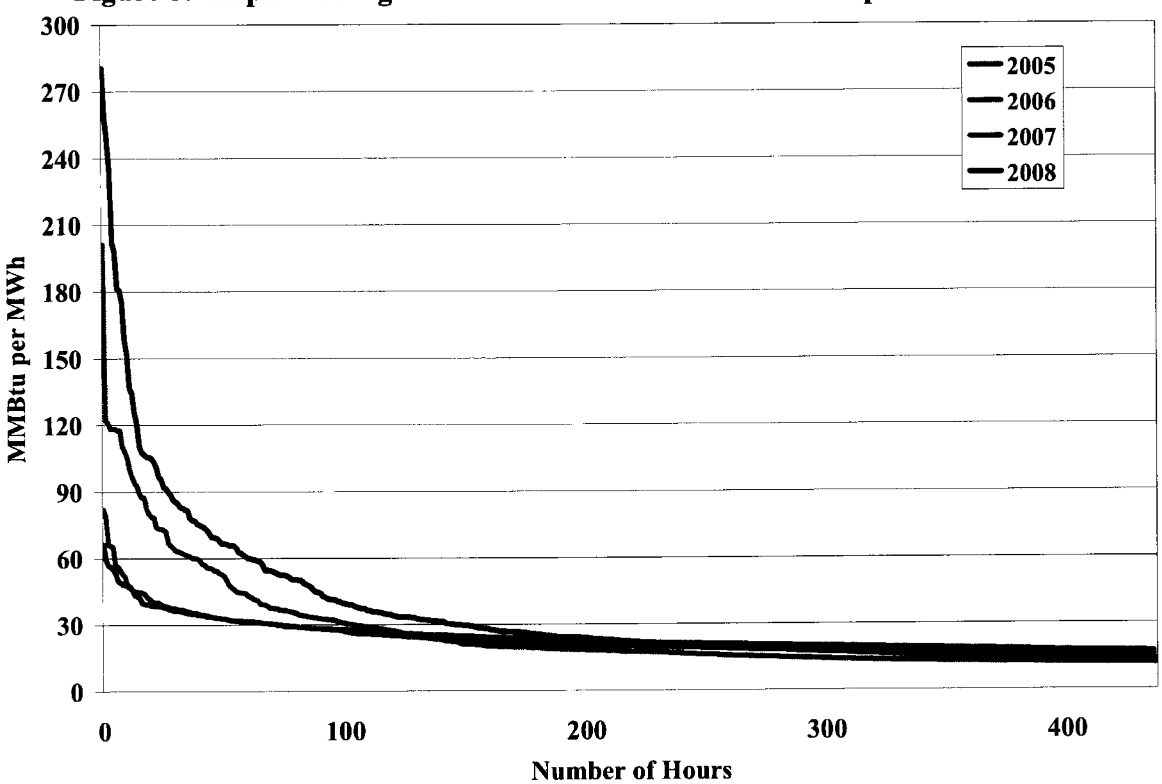
8 This methodology implicitly assumes that electricity prices move in direct proportion to changes in natural gas prices.

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



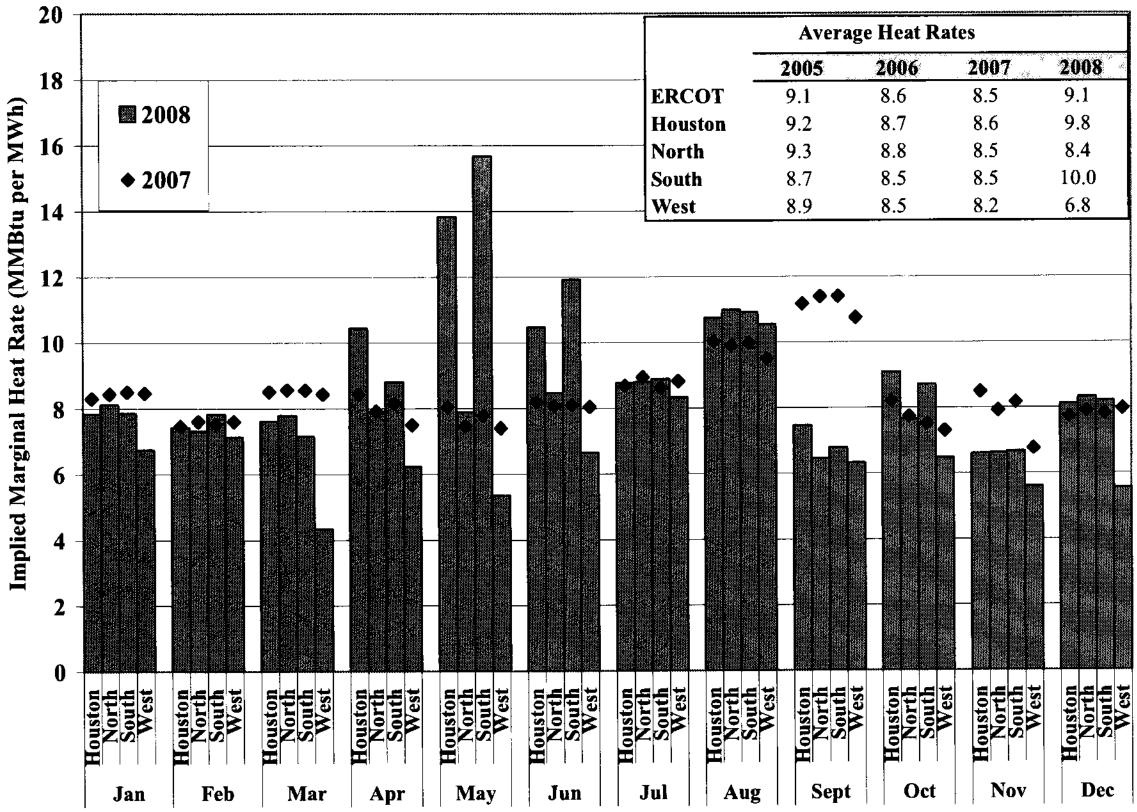
	> 8	> 10	> 15	>30
	(MMBtu per MWh)			
2005	3336	1860	600	72
2006	3484	1877	474	73
2007	3526	1211	275	103
2008	2885	1068	356	145

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



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

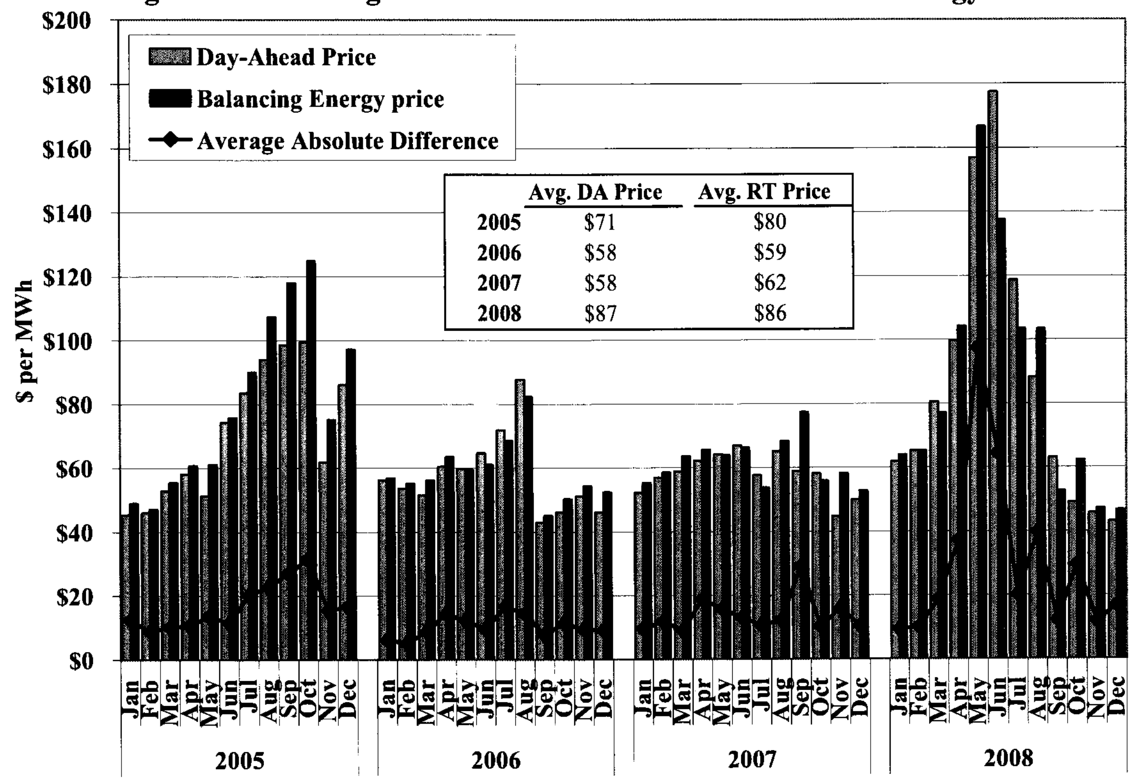


Figure 10 shows price convergence during peak periods (*i.e.*, weekdays between 6 AM and 10 PM). Day-ahead prices averaged $87 per MWh in 2008 compared to an average of $86 per MWh for real-time prices. Although the day-ahead and real-time prices exhibited good average convergence in 2008, Figure 10 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

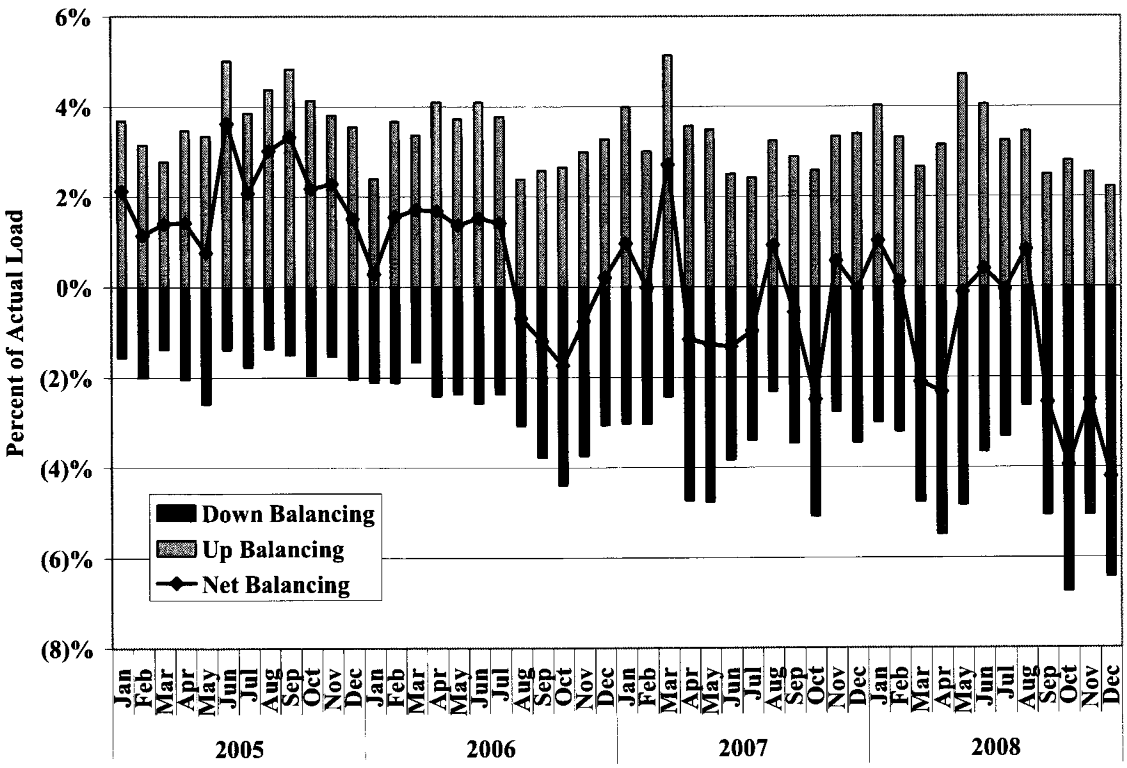
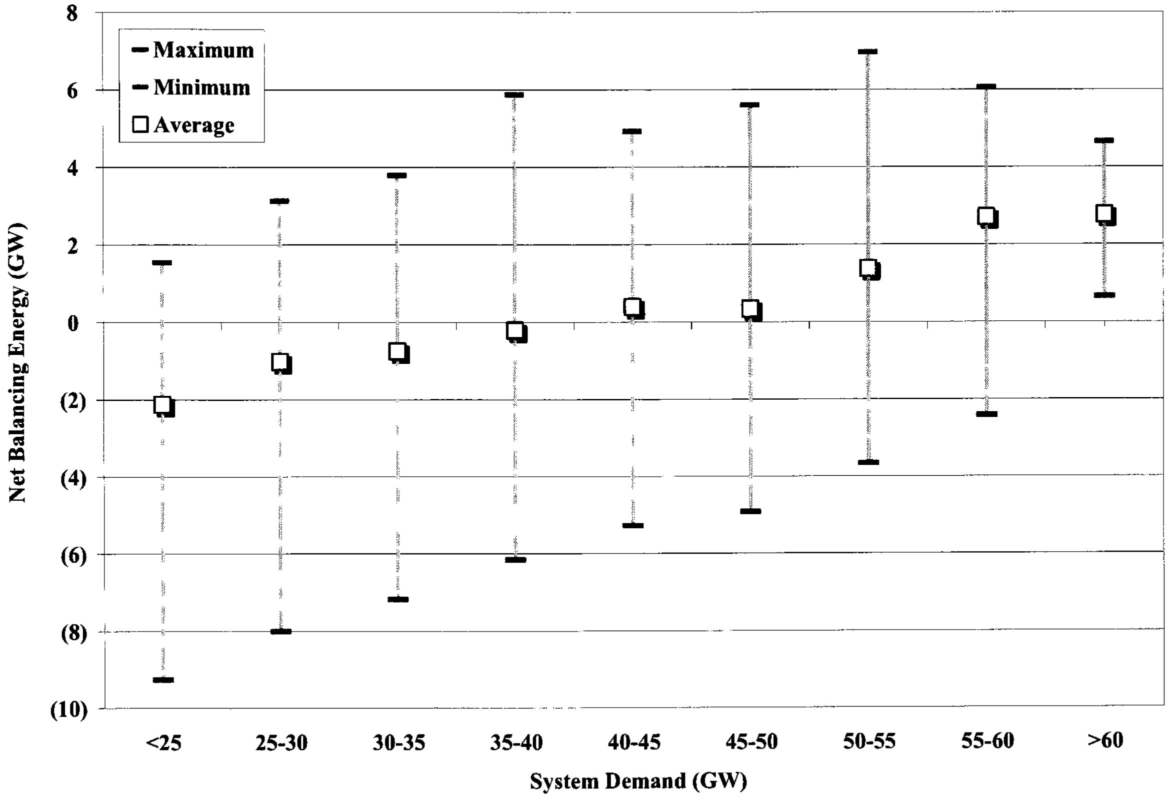


Figure 11 shows that the total volume of up balancing and down balancing energy as a share of actual load increased from an average of 6.8 percent in 2007 to 7.7 percent in 2008. Starting in August 2006, the average volume of down balancing energy began to increase. In 2008, for the first time the average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

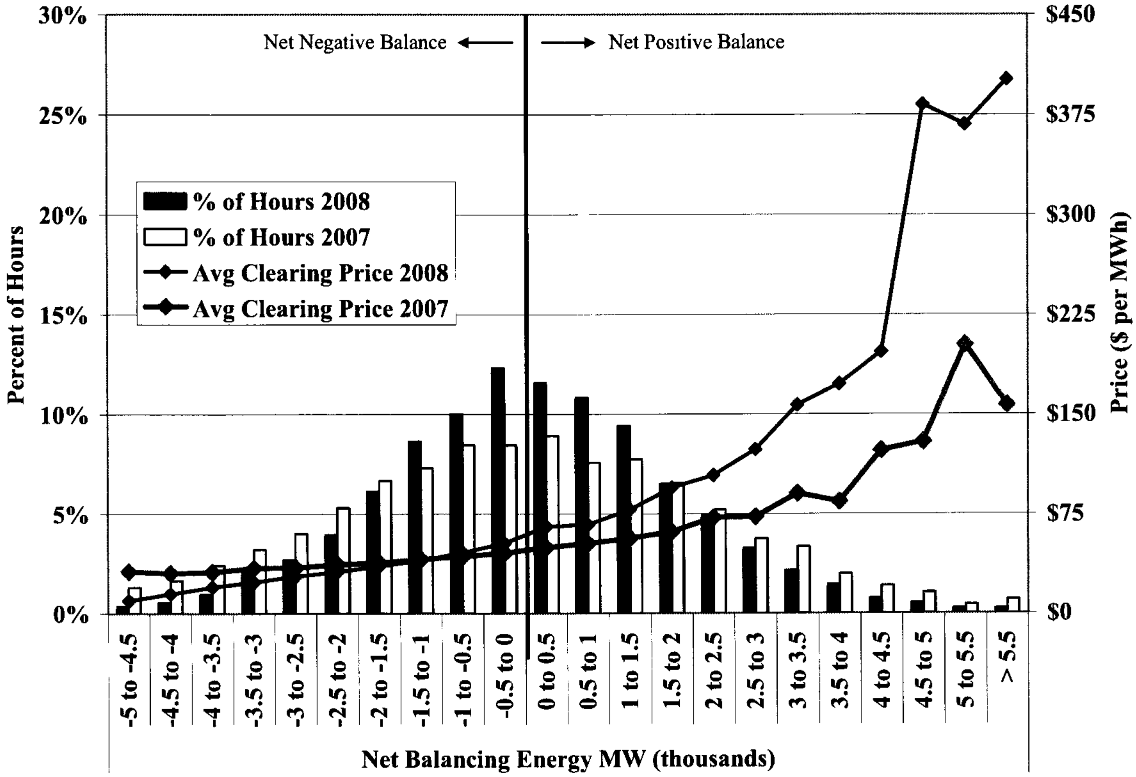


Figure 13 shows a relatively symmetrical distribution of net balancing energy purchases in 2007 centered around zero gigawatts, but 2008 is skewed more toward the left, meaning more down than up balancing energy was deployed. This is consistent with Figure 11 which showed that on average the quantity 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.

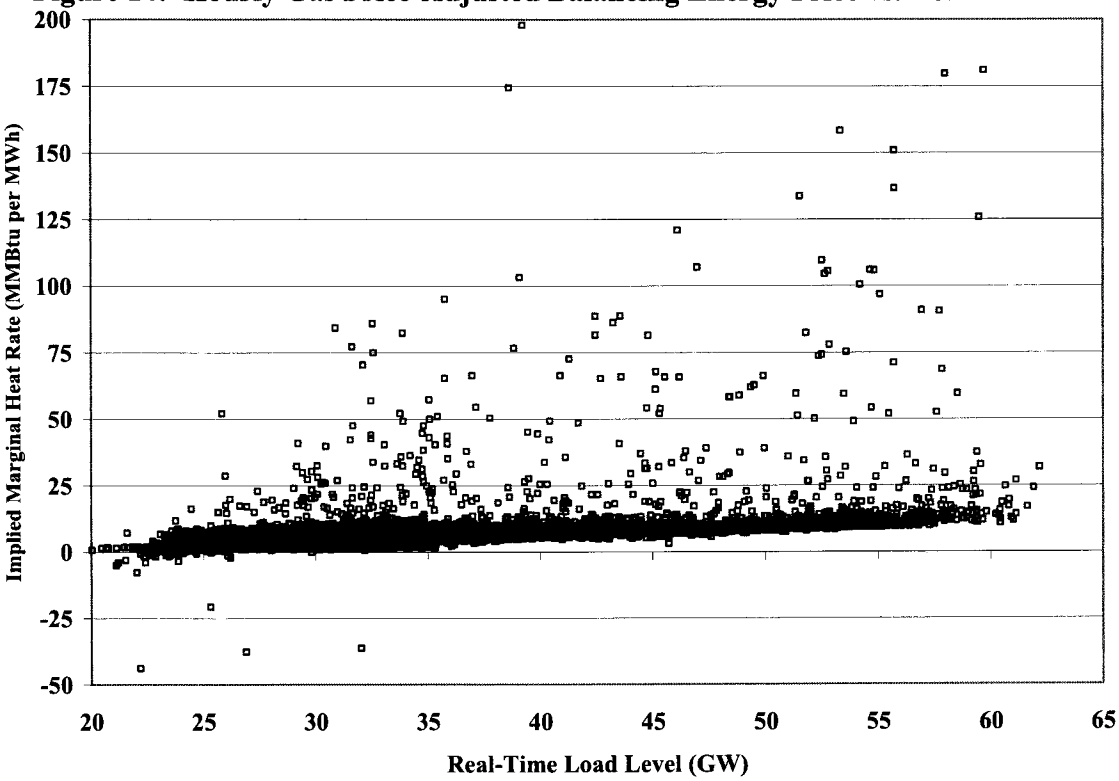
5. Determinants of Balancing Energy Prices

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



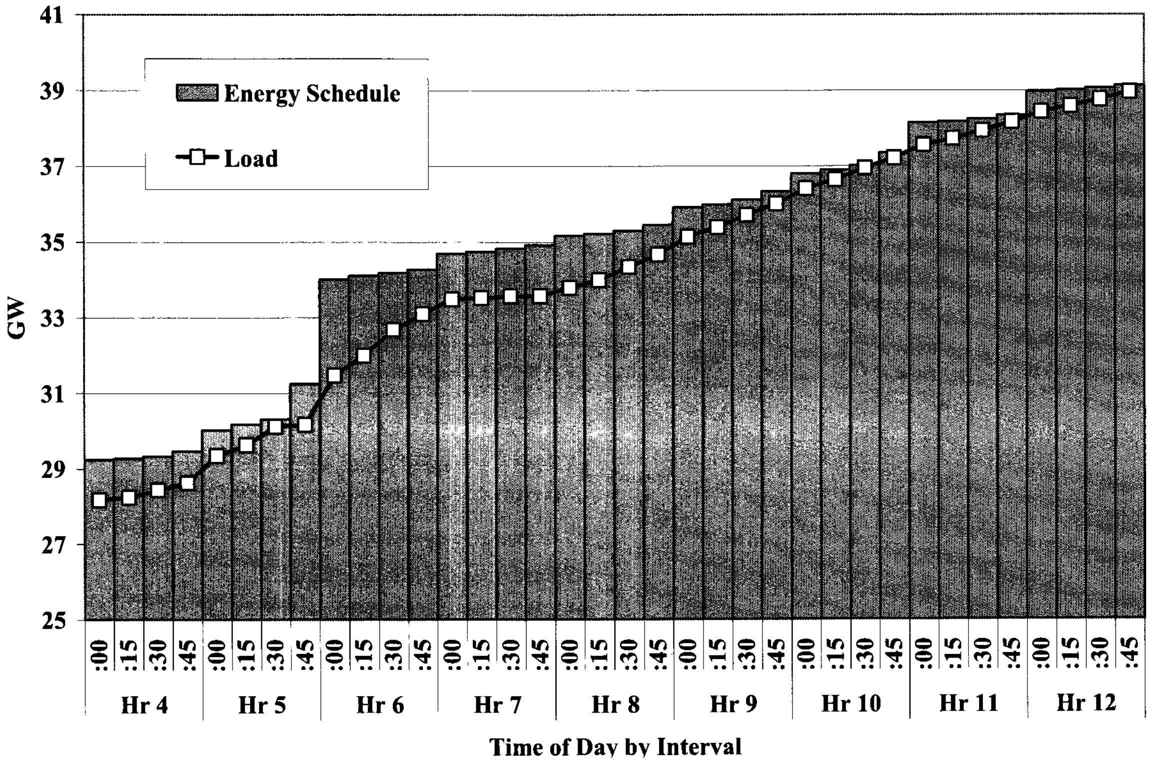
The figure indicates a positive correlation between real-time load and the clearing price in the balancing market. Although prices were generally higher at higher load levels, the analysis shown in Figure 13 indicates that the net volume of energy purchased in the balancing energy market is often a stronger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

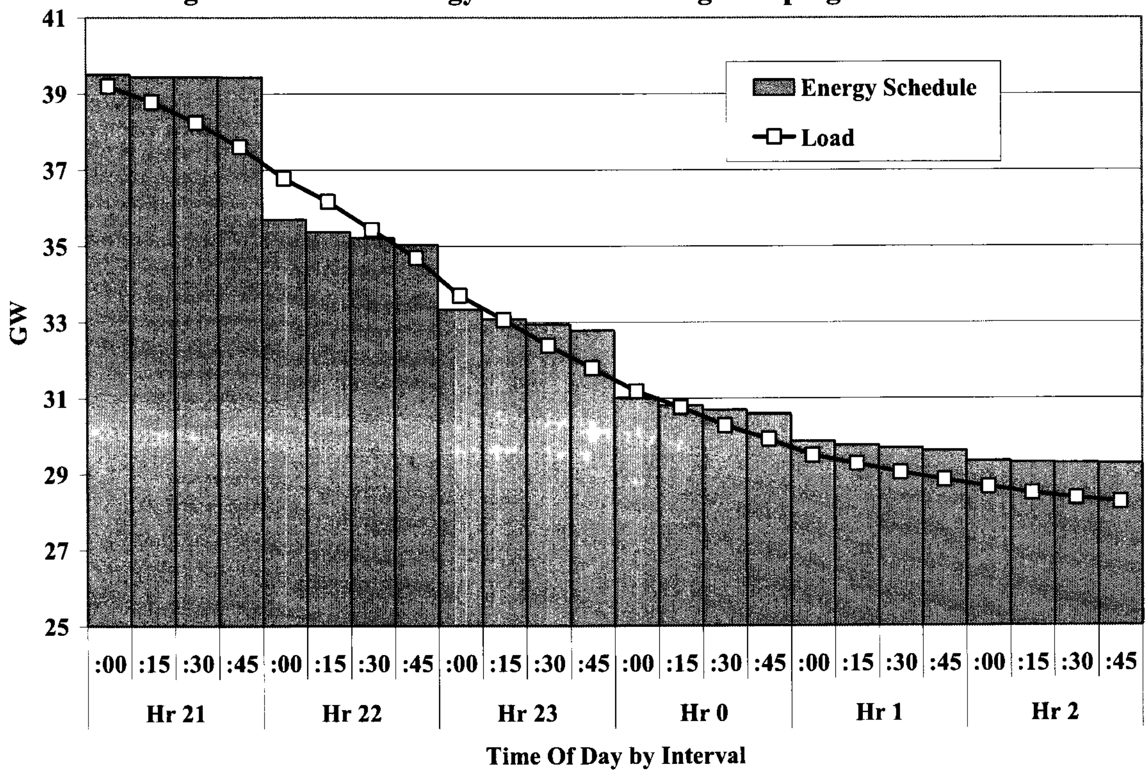
Figure 15: Final Energy Schedules during Ramping Up Hours



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

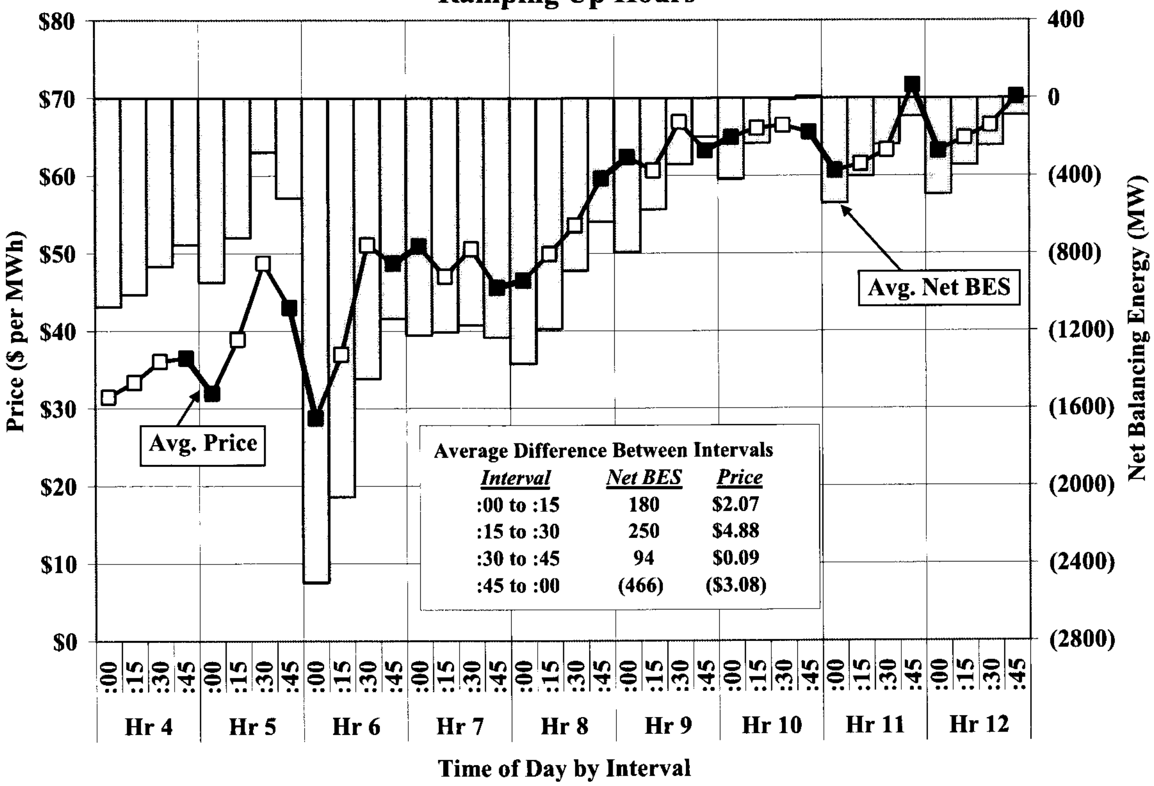


Figure 17 reveals two key aspects of the balancing energy market. First, as discussed above, balancing energy prices are highly correlated with balancing energy deployments. Second, with the exception of hour 7, there is a distinct pattern of increasing net balancing energy deployments during the hour. 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

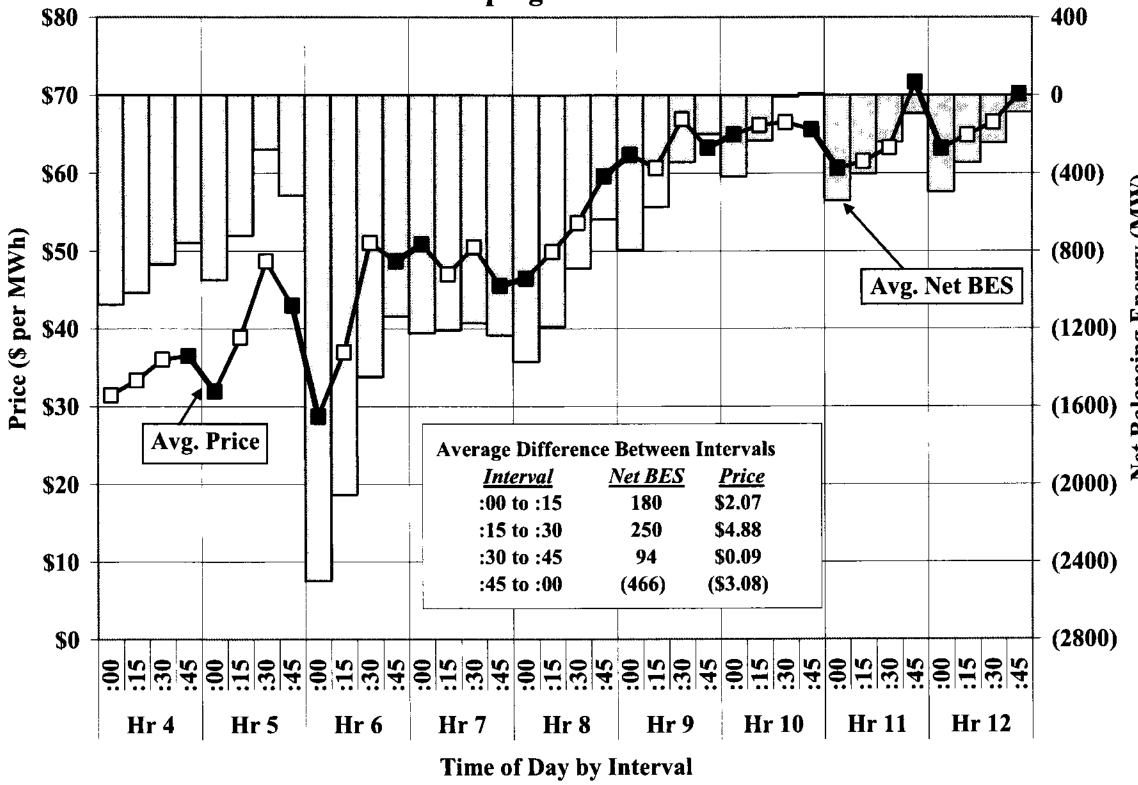


Figure 18 shows the same analysis for the ramping down hours. During ramping down hours, at the beginning of the hour, actual load tends to be higher than energy schedules, resulting in substantial balancing energy purchases, particularly in hour 22. At the end of the hour actual load tends to be lower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

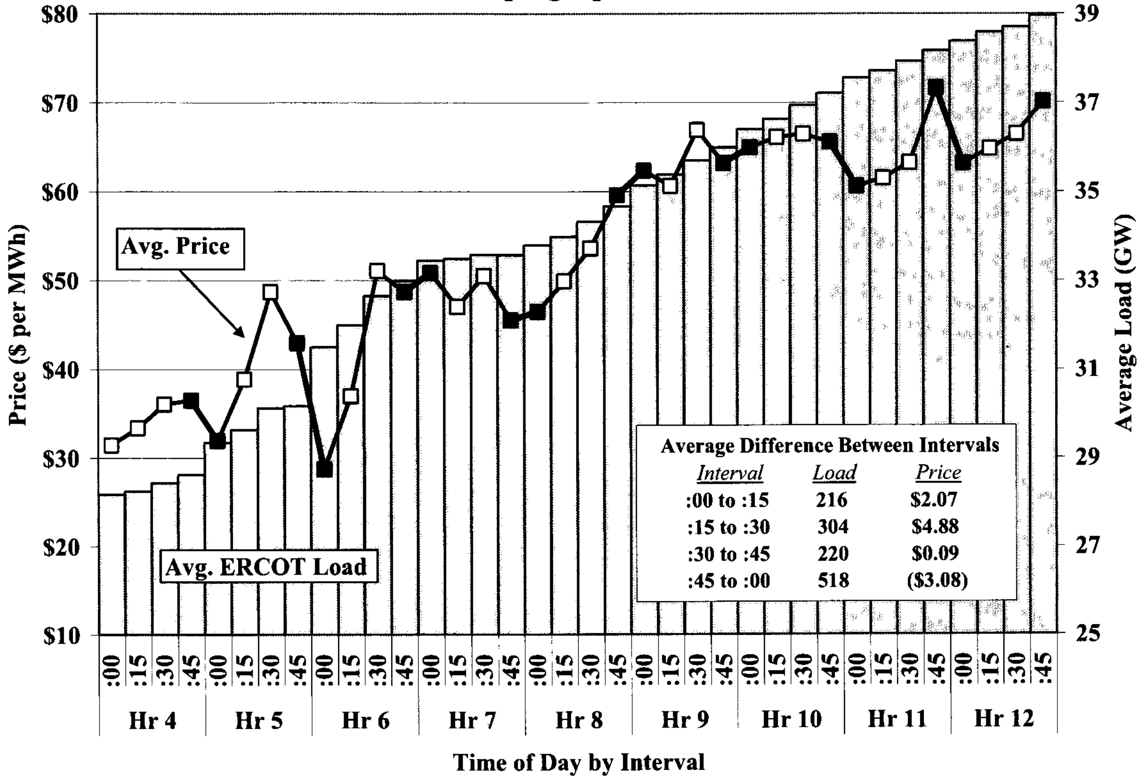


Figure 19 shows that, with the exception of hour 7, load steadily increases in every interval and prices generally move upward from an average of $31 per MWh at 4:00 AM to $70 per MWh at 12:45 PM. If actual load were the primary determinant of energy prices, the balancing energy prices would rise gradually 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

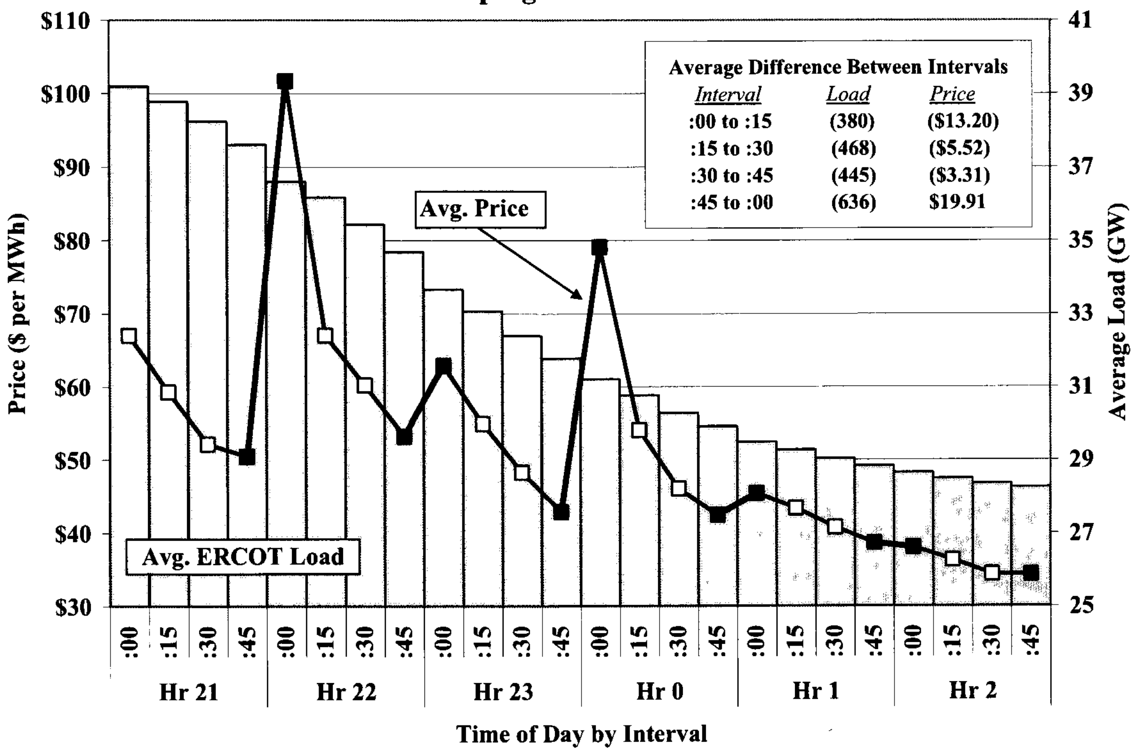


Figure 20 shows that while balancing energy prices decrease over these intervals, the pattern is similar to that exhibited in the ramping up hours. The balancing energy price decreases in each interval of the hour before rising substantially in the first interval of the following hour. The balancing 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.

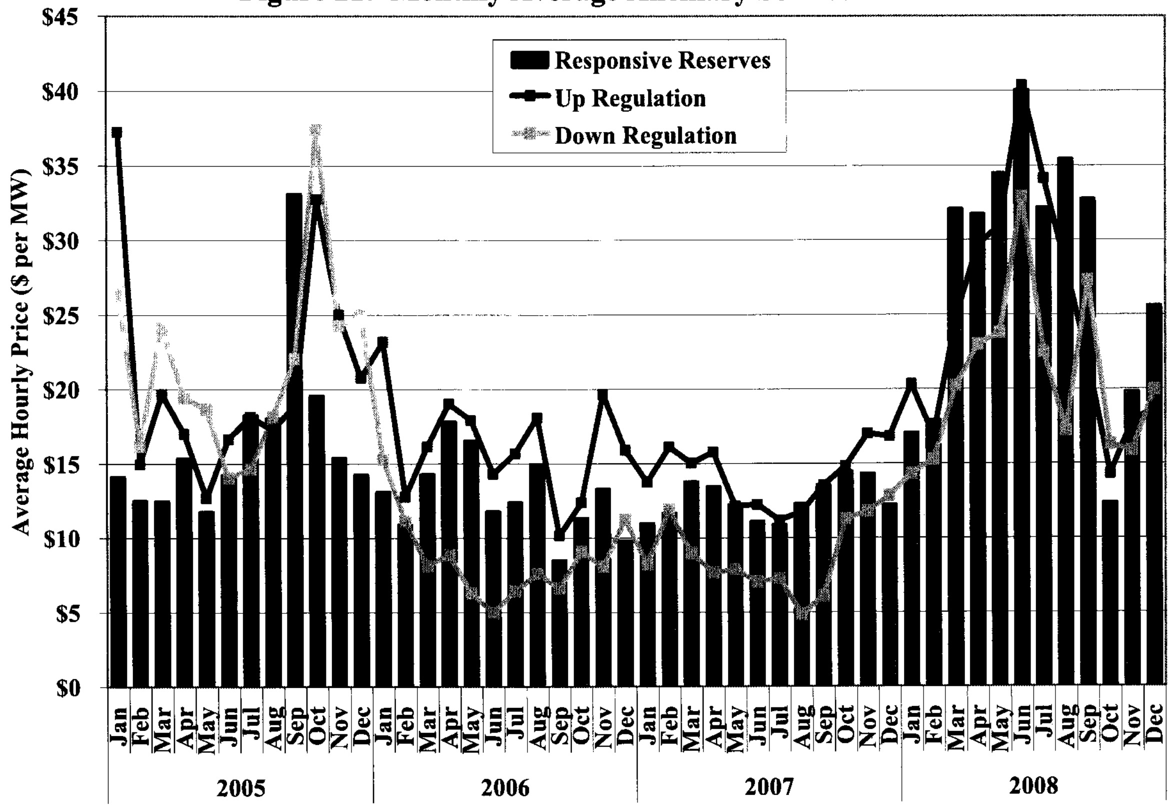
1. Reserves and Regulation Prices

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

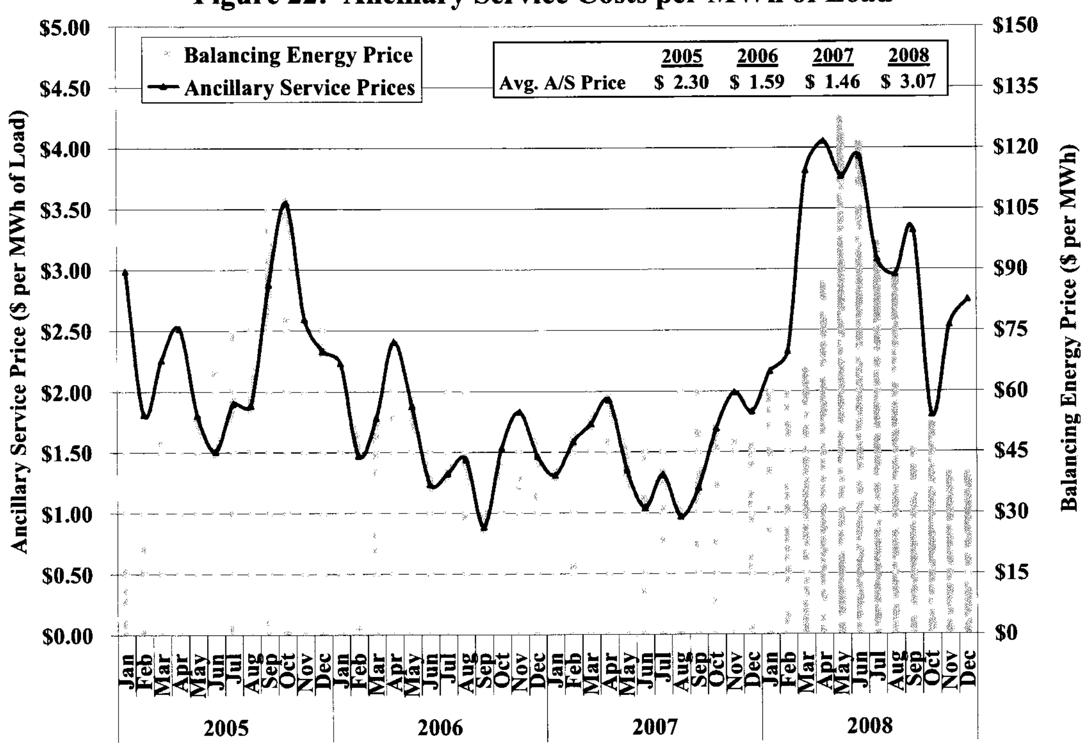


Figure 22 shows that total ancillary service costs are generally correlated with balancing energy price movements which, as previously discussed, are highly correlated with natural gas price movements. The average ancillary service cost per MWh of load increased to $3.07 per MWh in 2008 compared to $1.46 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

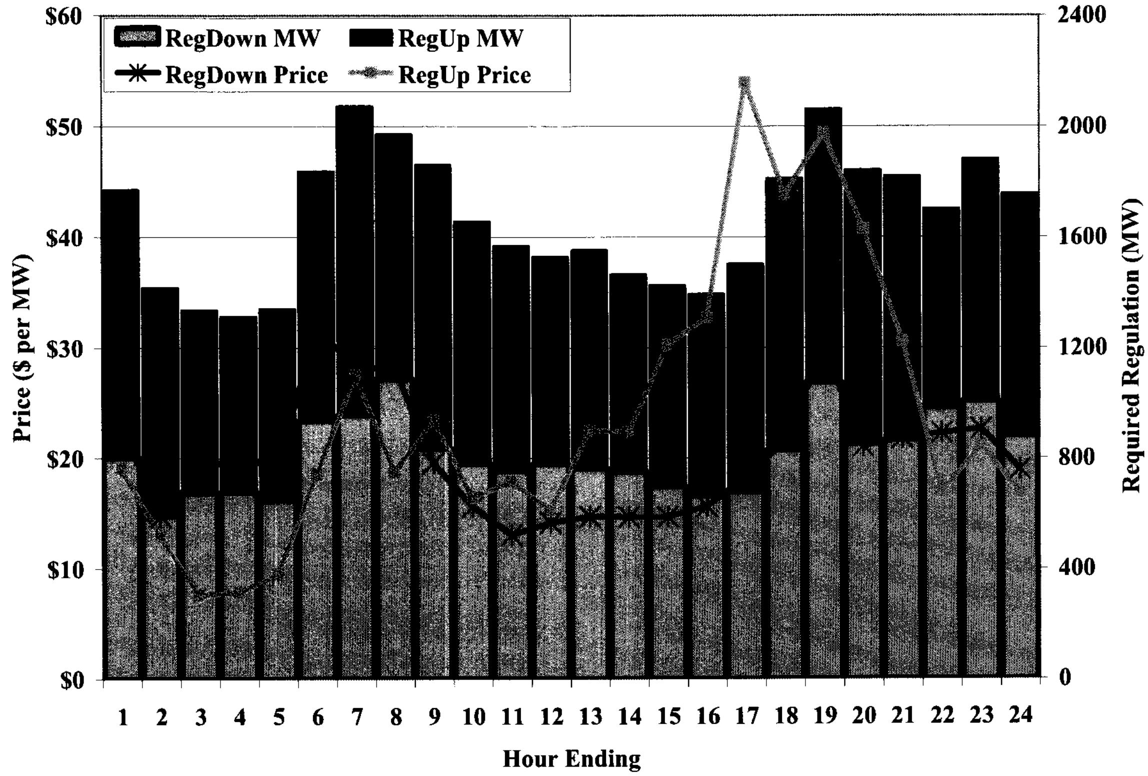


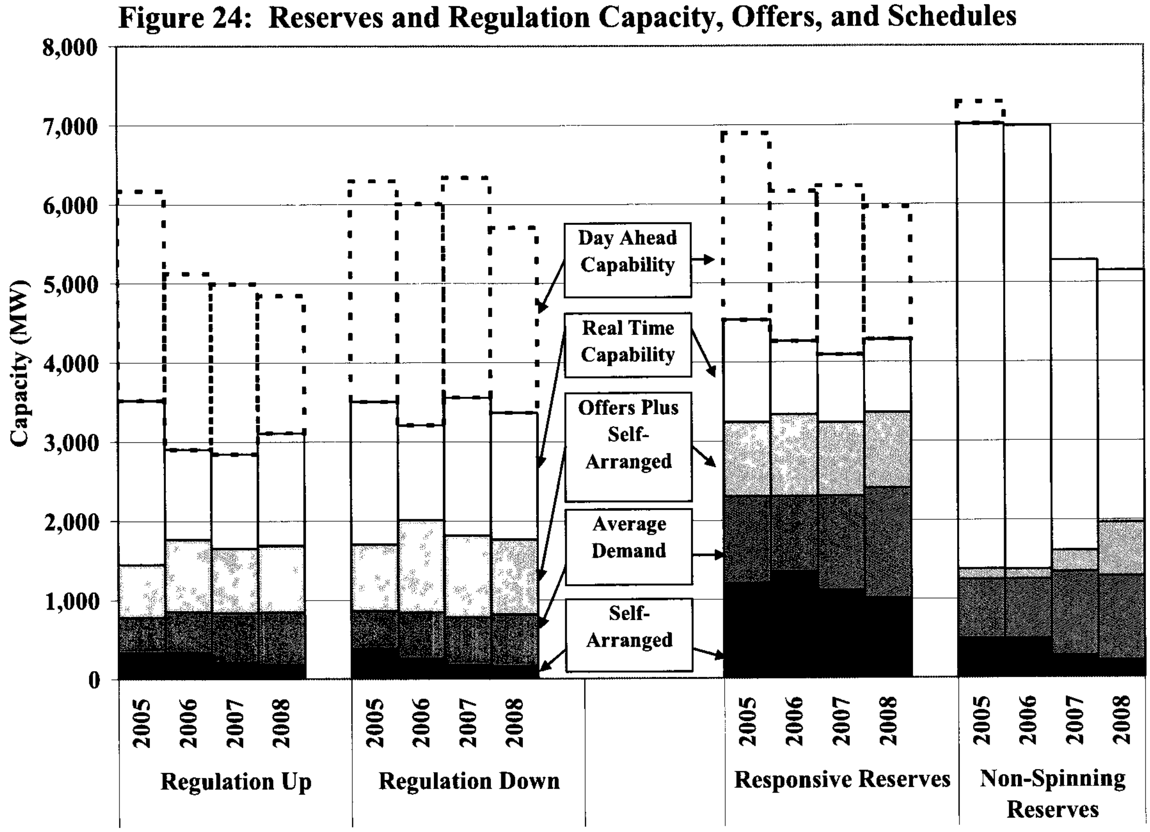
Figure 23 indicates that average regulation prices are generally correlated with the regulation quantity purchased and the typical load pattern in ERCOT. During non-ramping hours, such as overnight and late morning, up and down regulation prices are at their lowest levels. During the ramping hours in 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.

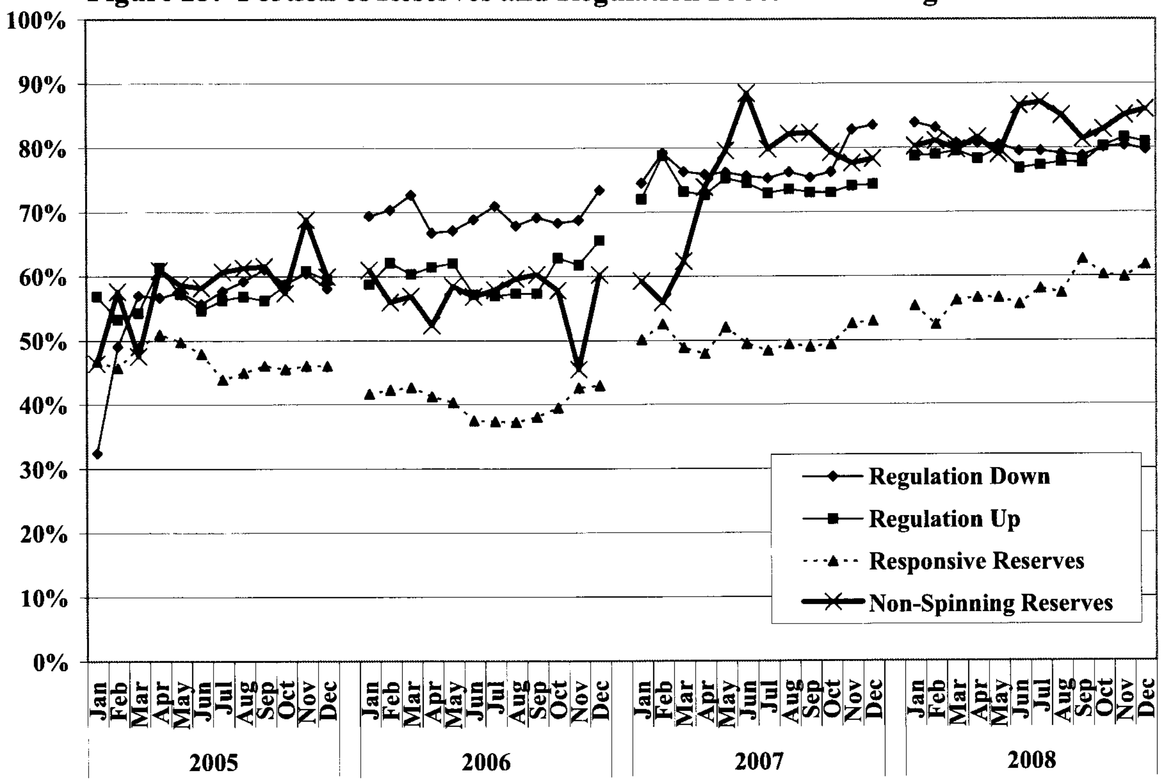
For non-spinning reserves, Figure 24 includes the capability of units that QSEs indicate are able to ramp-up in thirty minutes and able to start-up on short notice. The total capability shown in this figure does not account for capacity of online resources. However, it should be noted that any on-line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

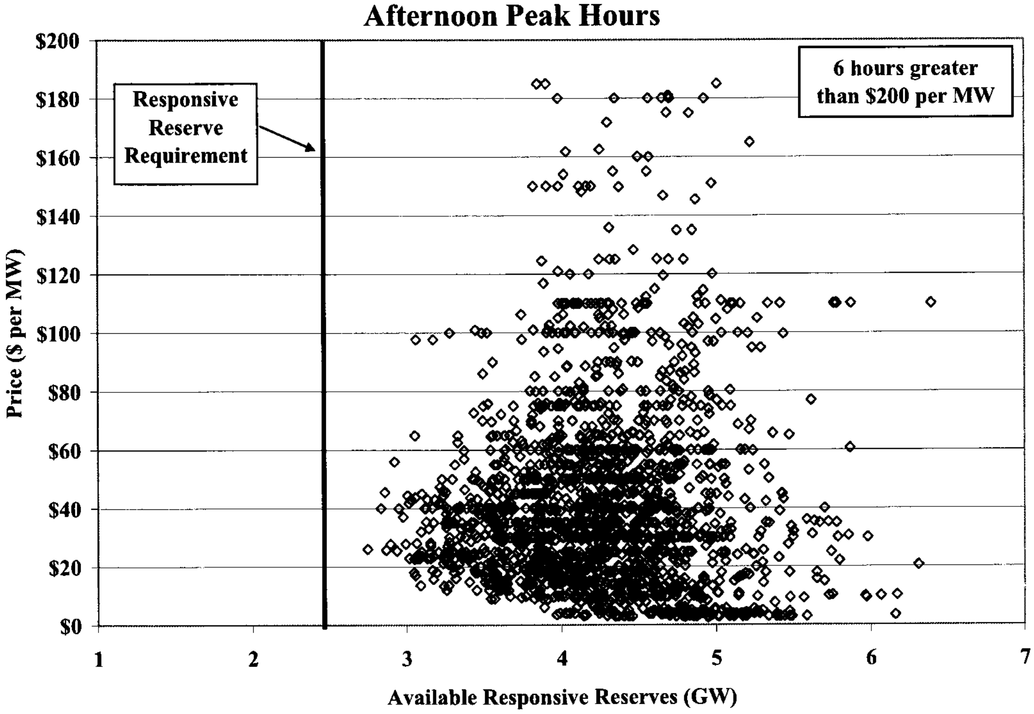


As market participants have gained more experience with the ERCOT markets, larger portions of the available reserves and regulation capability have been offered into the market, thereby increasing the market's liquidity. Nevertheless, Figure 25 shows that a fair share of these services is still self-supplied,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

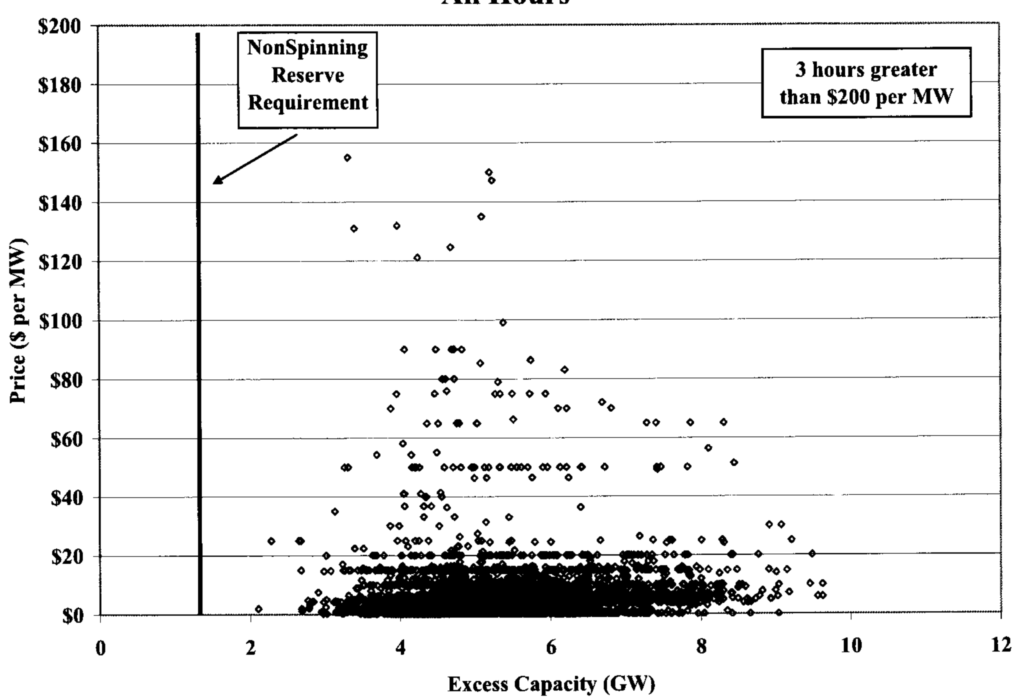
Figure 26: Hourly Responsive Reserves Capability vs. Market Clearing Price Afternoon Peak Hours



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

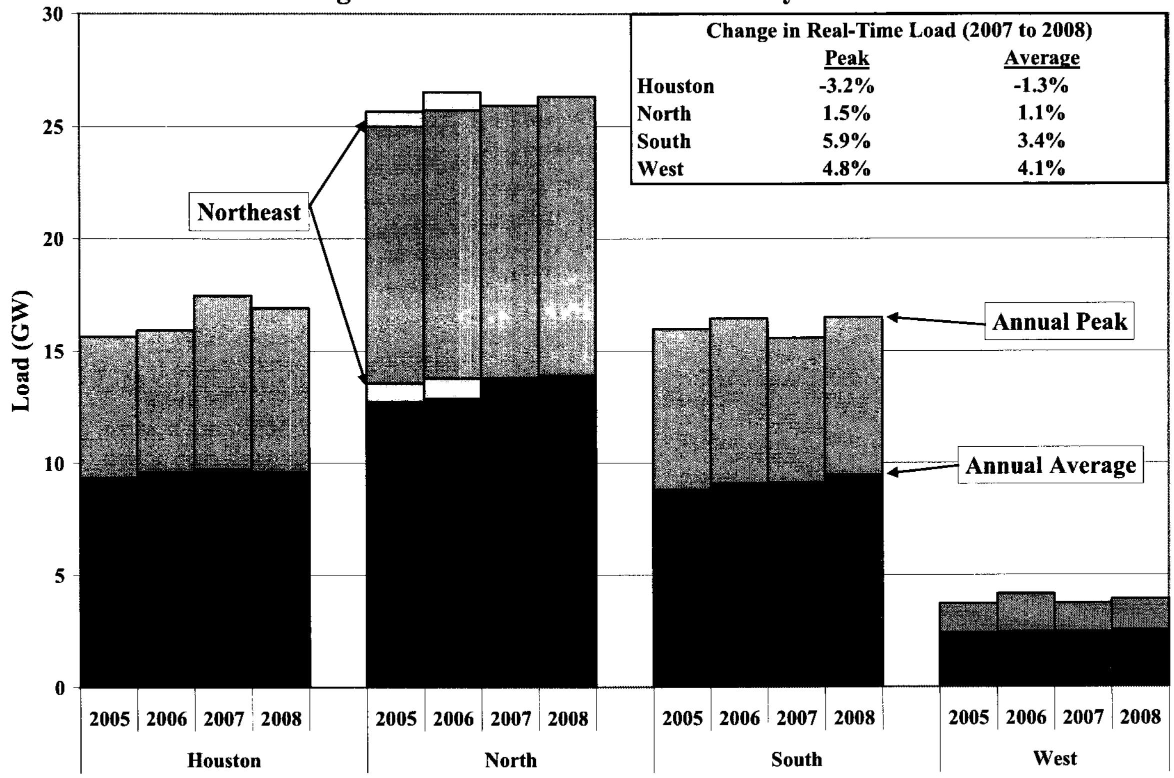
There are two important dimensions of load that should be evaluated separately. First, the changes in overall load levels from year to year can be shown by tracking the changes in average load levels. This metric will tend to capture changes in load over a large portion of the hours during the year.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

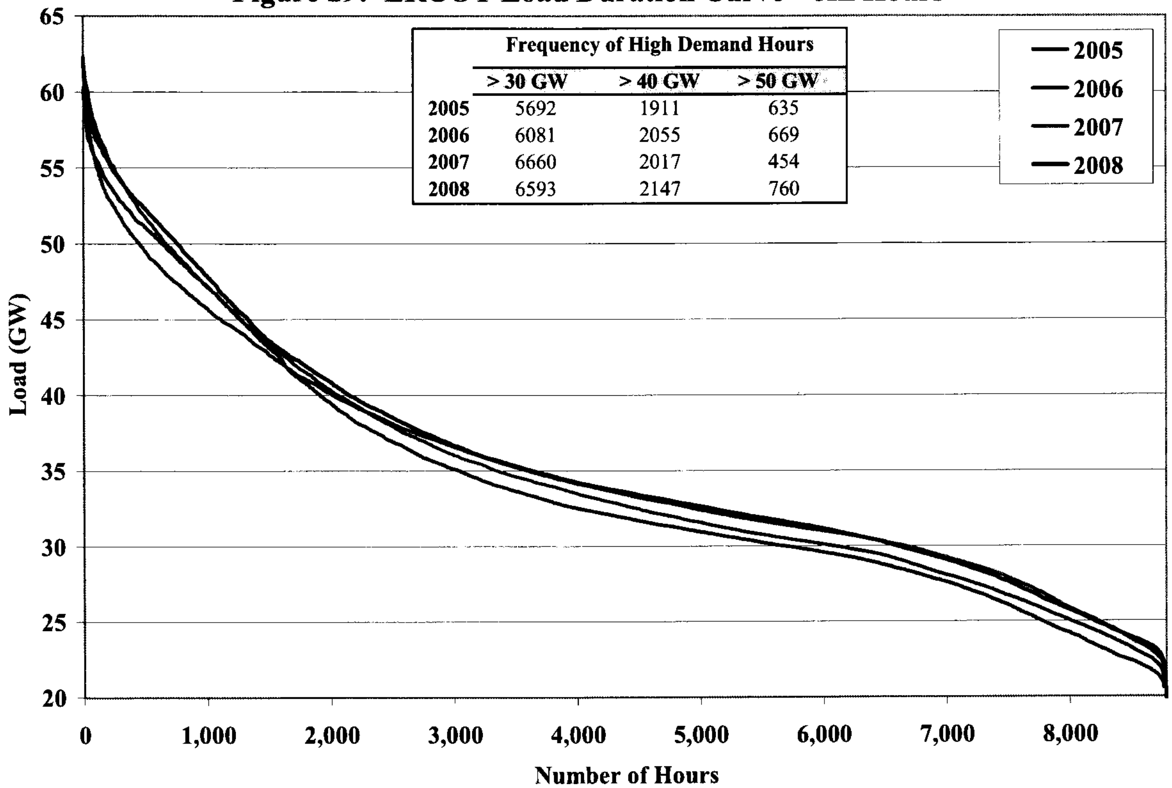
in different hours for different zones. As a result, the sum of the non-coincident peaks for the zones was greater than the annual ERCOT peak load.

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.

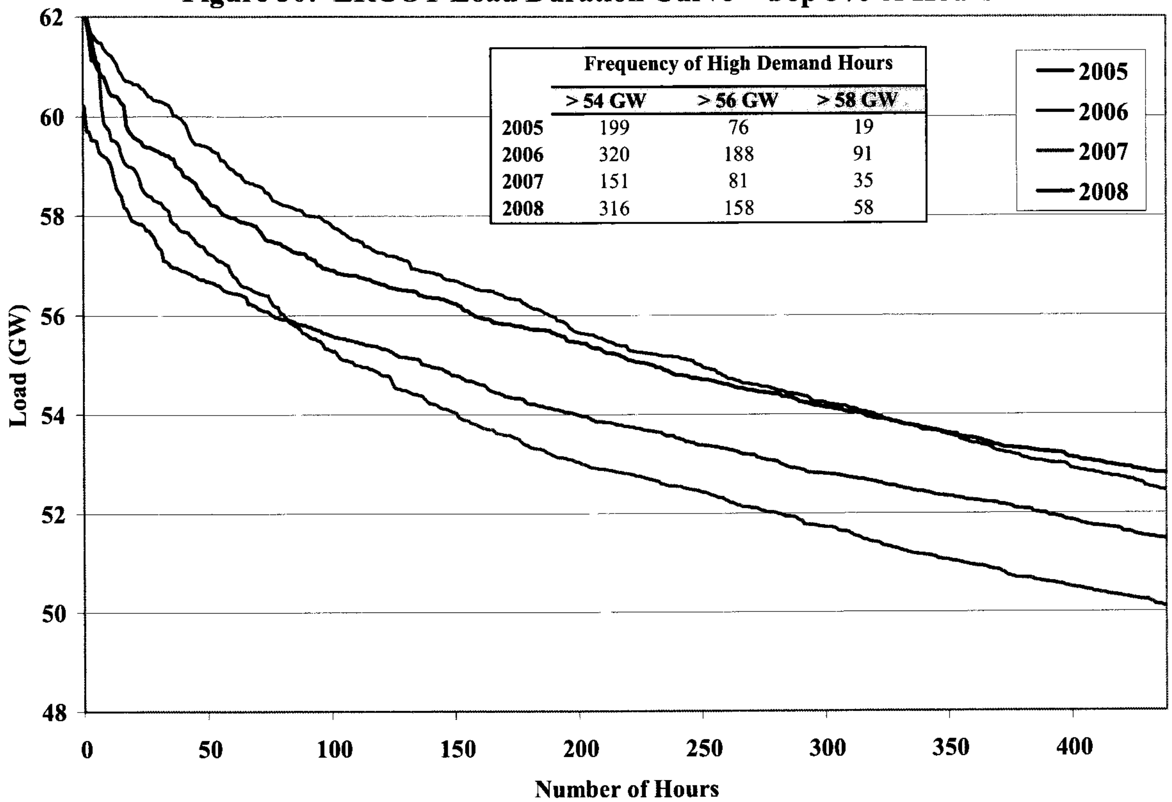
Figure 29: ERCOT Load Duration Curve – All Hours



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



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

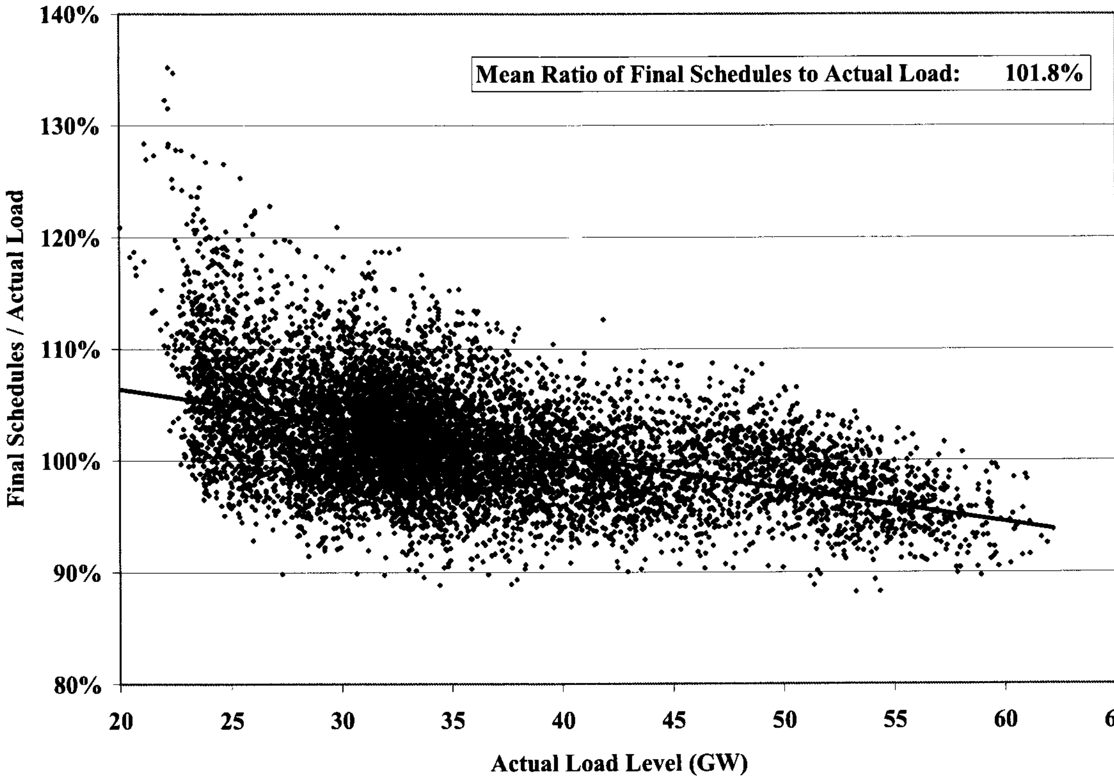


Figure 31 shows that final load schedules generally come very close to actual load in the aggregate, as indicated by an average ratio of the final load schedules to actual load of 101.8 percent. However, the figure also includes a trend line indicating that the ratio of final load schedules to actual 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

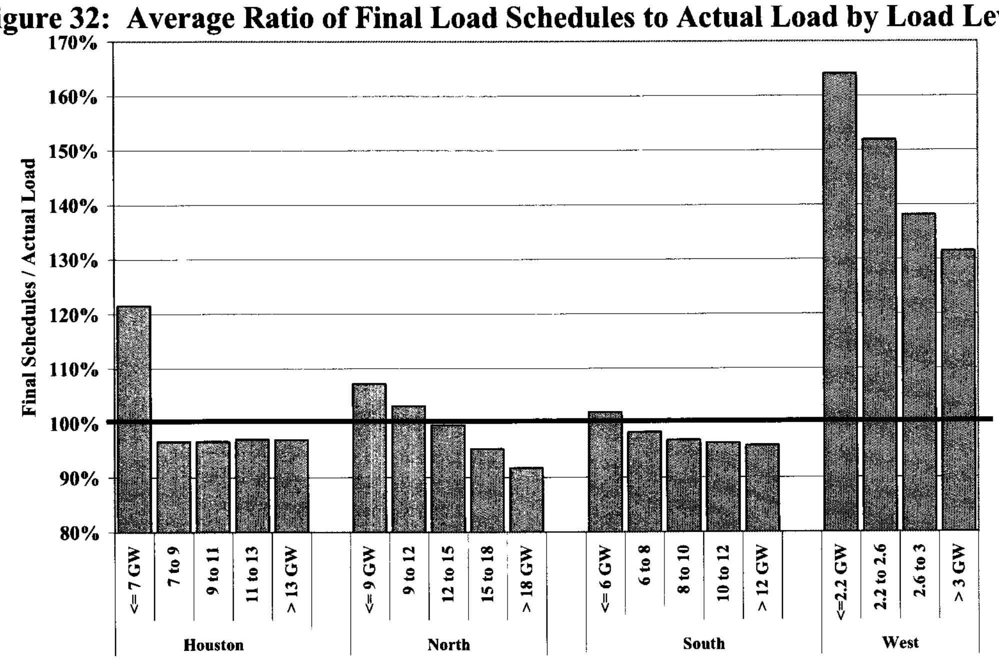


Figure 32 shows the ratio of final load schedules to actual load evaluated at five different load levels for each of the ERCOT zones. Figure 32 shows that:

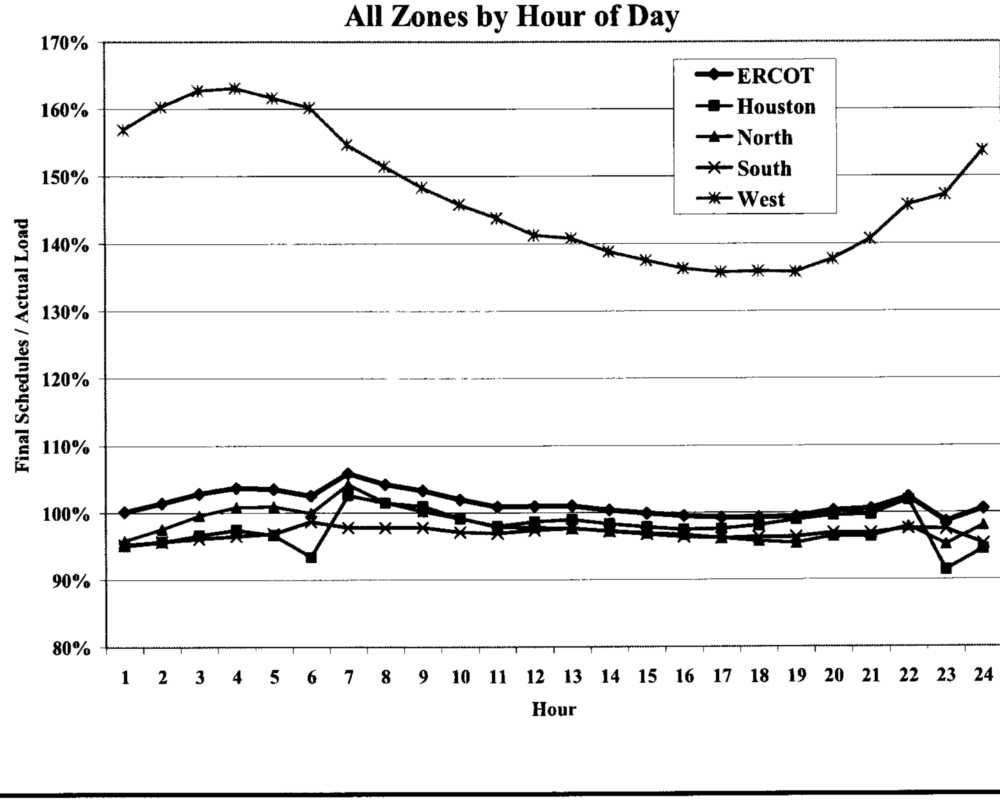
- The final schedule quantity decreases in three of the four zones as actual load increases.
- The West Zone is generally over-scheduled, although the ratio declines as load increases.

- 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



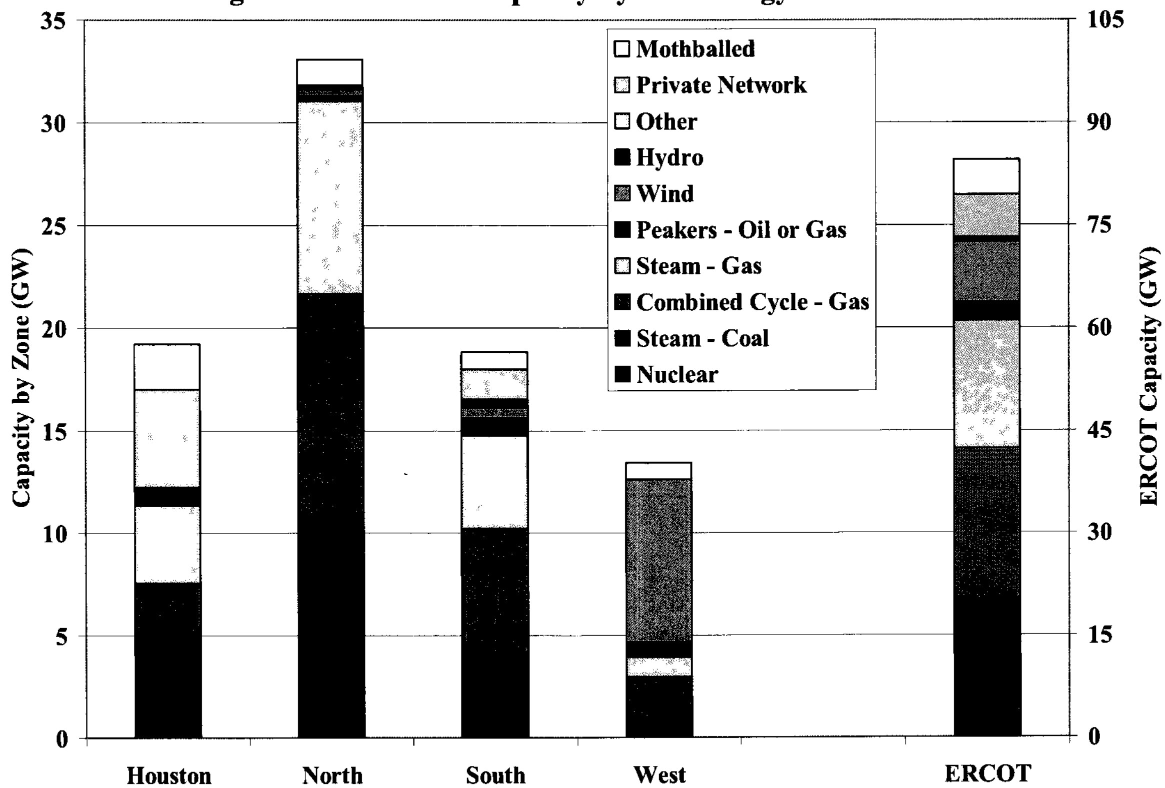
This figure shows that on an ERCOT-wide basis, final schedules are close to actual load in most of the hours during the day. The ERCOT-wide ratio increases to 106 percent at hour ending 7 and decreases to 99 percent and at hour ending 23. In the other hours, the ERCOT-wide ratio ranges between 95 and 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.

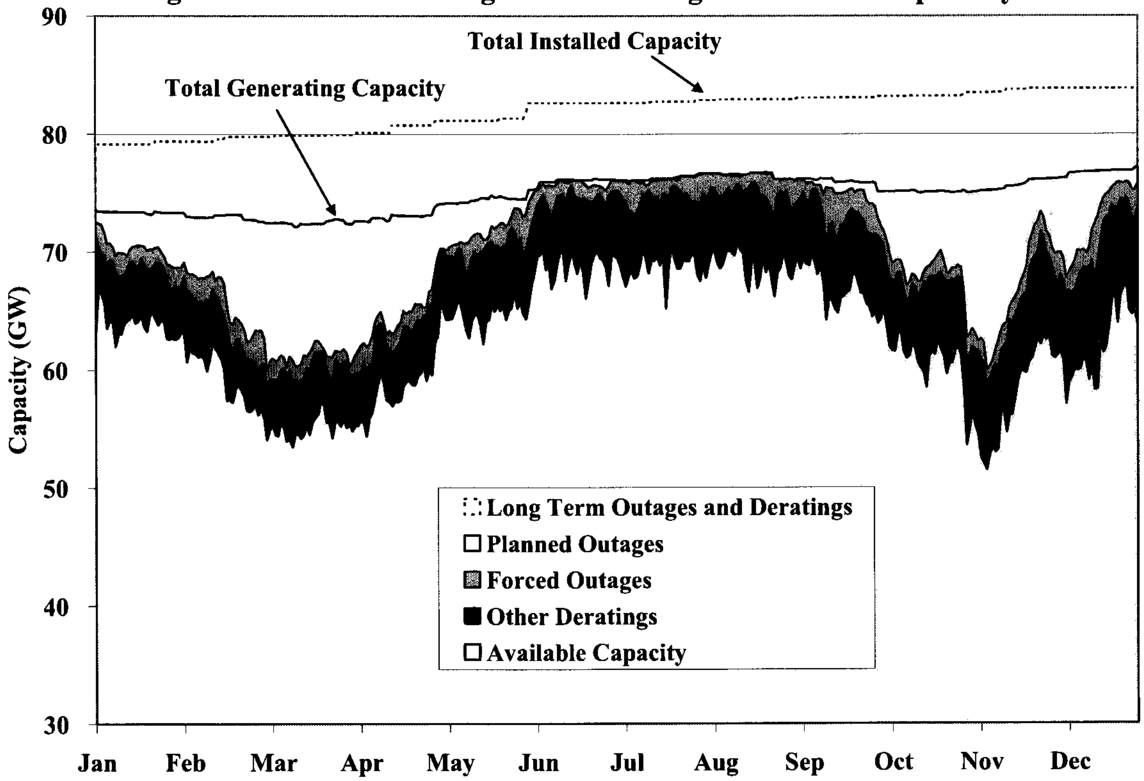
The distribution of capacity among the ERCOT zones is similar to the distribution of demand with the exception of the large amount of wind capacity in the West Zone. The North Zone accounts for approximately 39 percent of capacity, the South Zone 22 percent, the Houston Zone 23 percent, and the West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

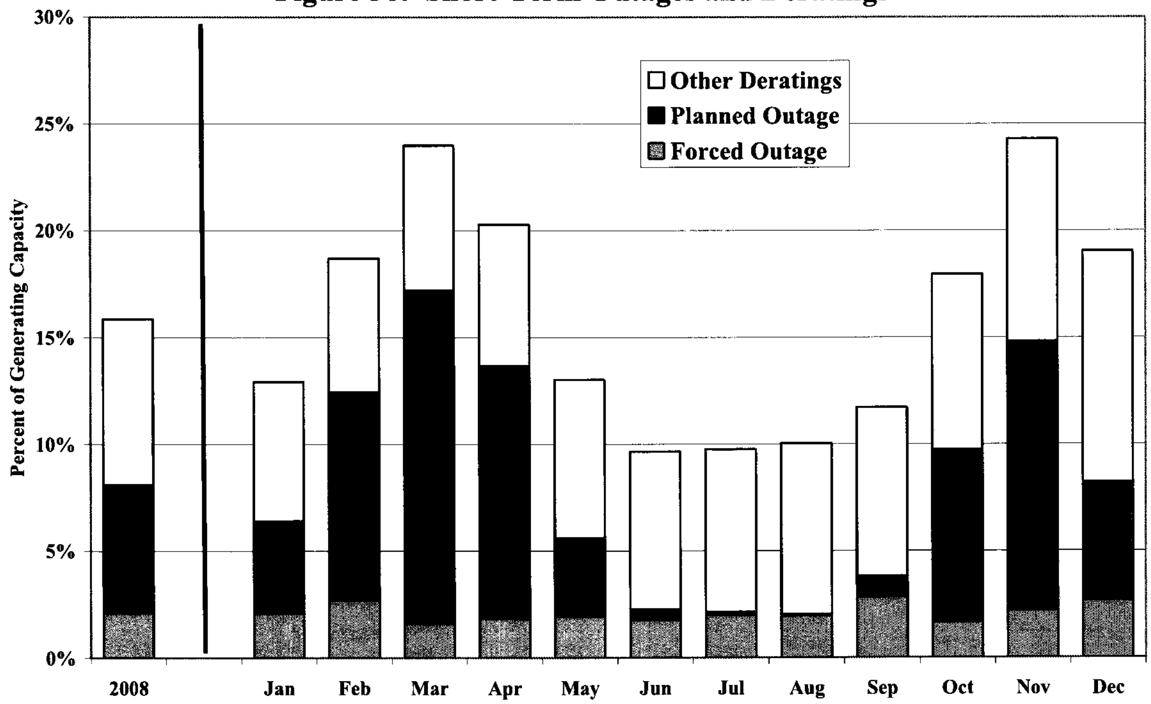
- Resources out-of-service for extended periods due to maintenance requirements.

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