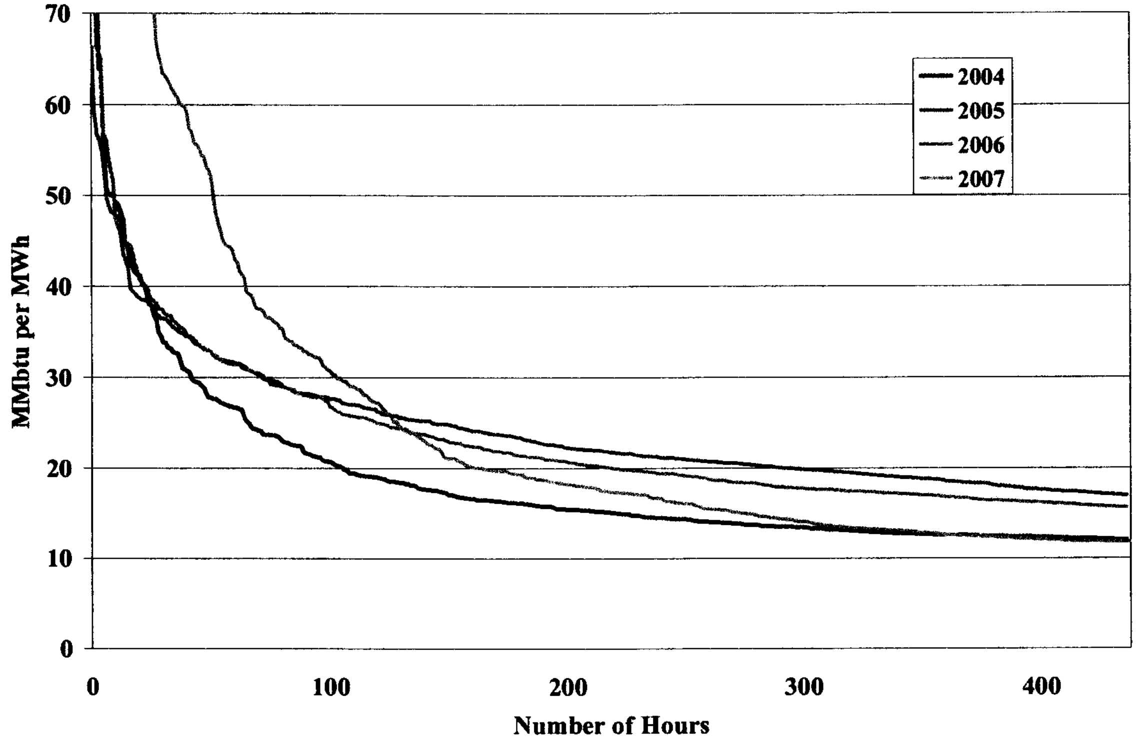
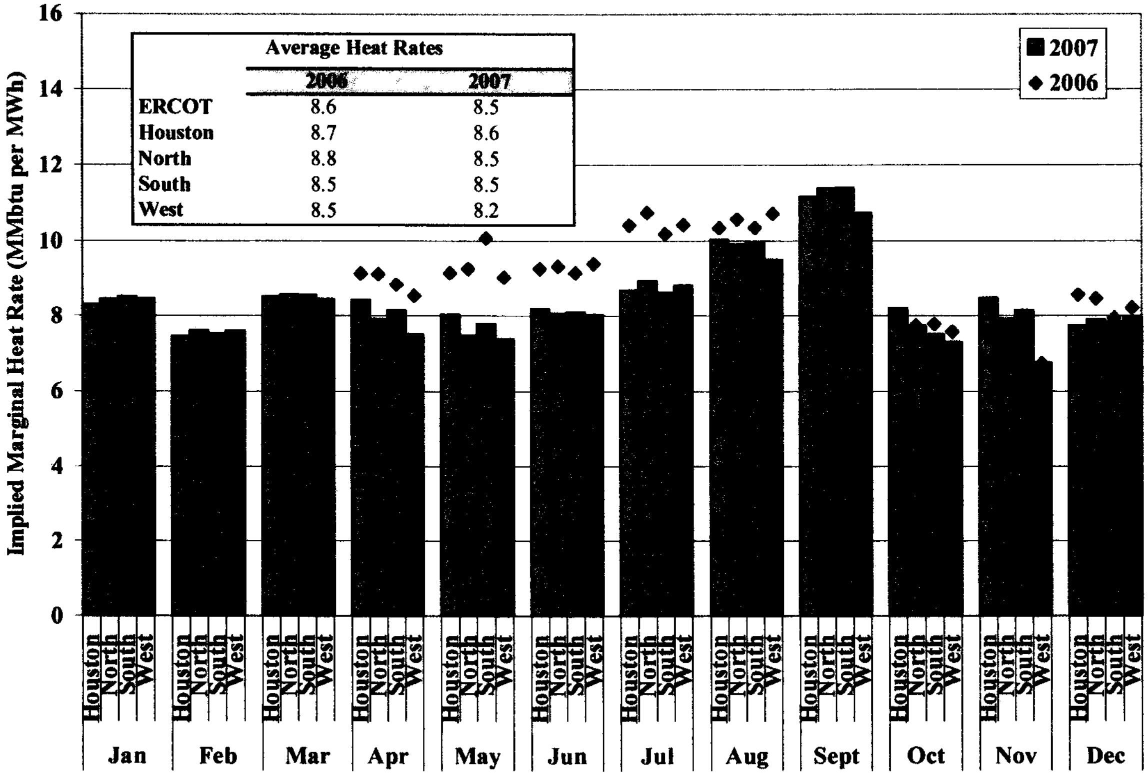
Figure 10: Implied Marginal Heat Rate Duration Curve
Top Five Percent of Hours – 2004 to 2007



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 2006 and 2007. This figure is the fuel price-adjusted version of Figure 1 in the prior sub-section. Adjusted for gas price influence, Figure 11 shows that average implied heat rate for all hours of the year decreased by 1.2 percent from 8.6 in 2006 to 8.5 in 2007.

Figure 11: Monthly Average Implied Marginal Heat Rates 2006 & 2007



On average, the implied heat rate was lower in 2007 than in 2006 for the months of April through August. With the exception of December, the average implied heat rate for the remaining months was higher in 2007 than in 2006. The decreases in implied heat rates during the summer of 2007 relative to 2006 are explained in part due to significantly above average rainfall levels 2007. The higher implied heat rates in September 2007 were due to several days in which non-spinning reserves were deployed and balancing market clearing prices were corrected to significantly higher levels pursuant to the provisions of the ERCOT Protocols.[13]

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

[13] The price correction provisions were adopted in Protocol Revision Request No. 650. The appropriateness of these price correction provisions was addressed in the 2006 ERCOT SOM (2006 ERCOT SOM Report, at 41-42).

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. This will tend to improve the convergence of the forward and real-time prices.

We believe 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 will be willing to pay a premium to purchase energy in the bilateral market.

There are several ways to measure the degree of price convergence between forward and real-time markets. In this section, we measure two aspects of convergence. The first analysis investigates whether there are systematic 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 systematic differences between forward and real-time prices, we examine the difference between the average forward price[14] and the average balancing energy price in each month between 2004 and 2007. This reveals whether persistent and predictable differences exist between forward and real-time prices, which participants should arbitrage over the long-term.

14 Day-ahead bilateral prices are from Megawatt Daily.

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 11 for each month between 2004 and 2007.

Figure 12: Convergence Between Forward and Real-Time Energy Prices 2004 to 2007

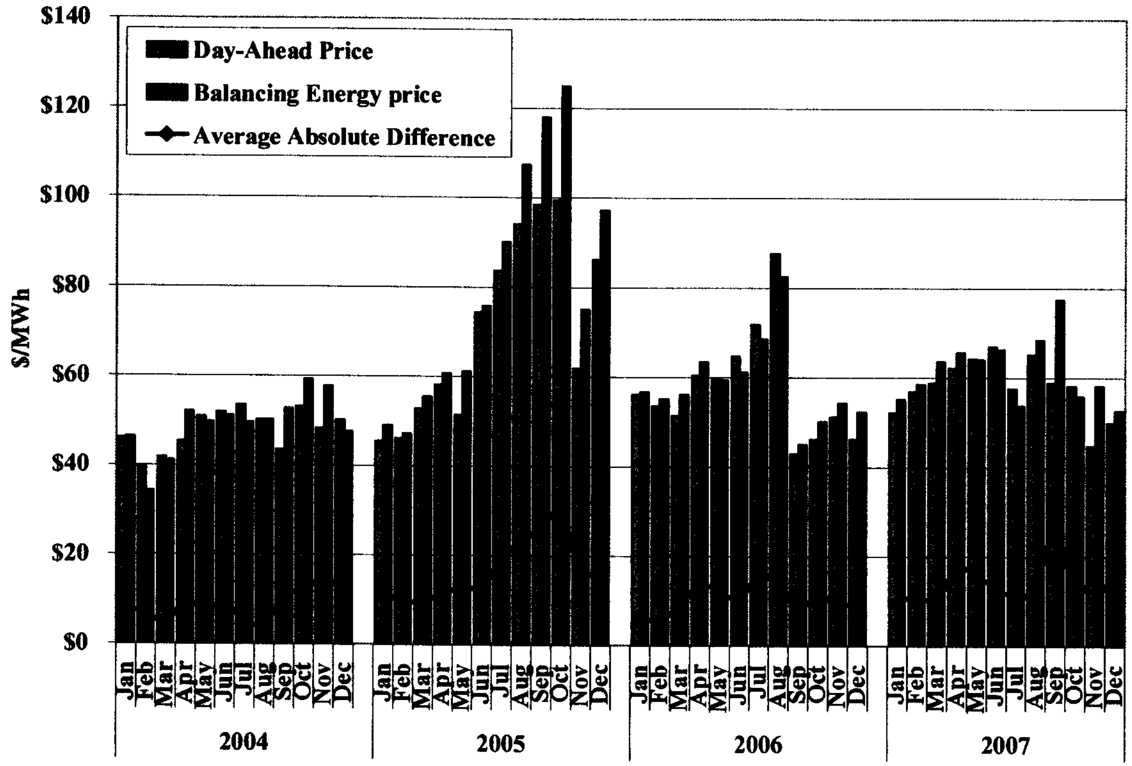


Figure 12 shows price convergence during peak periods (i.e. weekdays between 6 AM and 10 PM). This timeframe matches the definition of peak hours that are commonly traded in the forward market. During most of 2004, the average day-ahead price was consistent with the average balancing energy price. However, starting in September 2004 and continuing through

2005, it became common for the average balancing energy price to exceed the day-ahead price by a significant margin. In 2006, the average day-ahead price again became relatively consistent with the average balancing energy price. In 2007, the average day-ahead prices were also relatively consistent with the average balancing energy price except in the months of September and November. In the month of September there were four days when the price difference was greater than $50, and in the month of November there were two occurrences of price differences greater than $50. The average absolute price difference in September was $29 and the average absolute price difference in November was $16. In most of the months in 2007, the average balancing energy prices were higher than the average day-ahead price.

Figure 12 also shows that the average absolute price difference from 2004 to 2007. The difference (shown by the line) was relatively low during the first eight months of 2004 before rising considerably during the last four months. In 2005, the average absolute difference rose sharply in the summer and fall. In 2006, the average absolute difference dropped closer to the average level observed in 2004. The average absolute difference was $9 in 2004, $17 in 2005 $10 in 2006 and $14 in 2007. As noted above, the average absolute difference measures the volatility of the price differences.

The results in this section indicate that, with the exception of September 2007, convergence between the day-ahead bilateral prices and the balancing energy prices was comparable in 2007 to 2006. It is expected that the implementation of the nodal market with an integrated day-ahead market will result in improved price convergence over that which has been experienced in the zonal market.

4. Volume of Energy Traded in the Balancing Energy Market

The primary purpose of the balancing energy market is the match supply and demand in real-time. 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), the prices in the current market would be identical to the 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 their output and supporting the bilateral sale with balancing energy purchases. Hence, 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 13 shows the average quantities of balancing up and balancing down energy sold by suppliers in each month, along with the net purchases or sales (*i.e.*, balancing up energy minus balancing down energy).

Figure 13: Average Quantities Cleared in the Balancing Energy Market 2003 to 2007

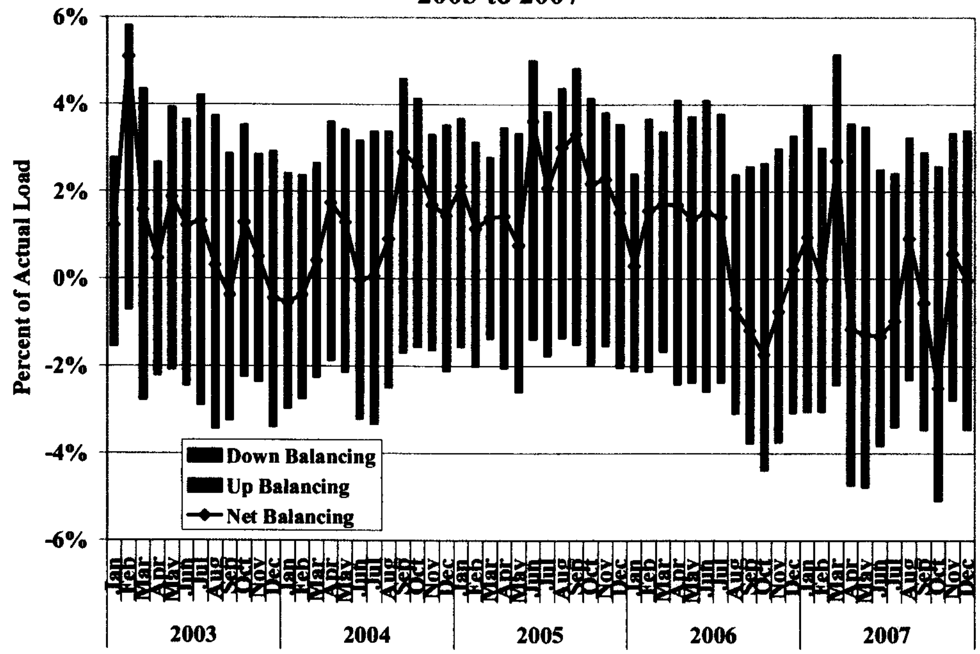


Figure 13 shows that the total volume of balancing up and balancing down energy as a share of actual load increased from an average of 5.6 percent in 2005 to 6.1 percent in 2006 and 6.8 percent in 2007. Starting in August 2006, the average volume of balancing down energy began to increase. In 2007, the average amount of balancing down energy was greater than balancing up energy. Relaxed balanced schedules allow market participants to intentionally schedule more or less than their anticipated load, and to buy or sell 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 spot 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 the balancing energy quantities increased was that large quantities of balancing up and balancing down energy are 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 balancing-up or net balancing-down 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 outside the summer 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 14 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 14 shows the portion of the hours during 2007 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 9 percent of the hours in 2007.

Figure 14: Magnitude of Net Balancing Energy and Corresponding Price 2006 and 2007

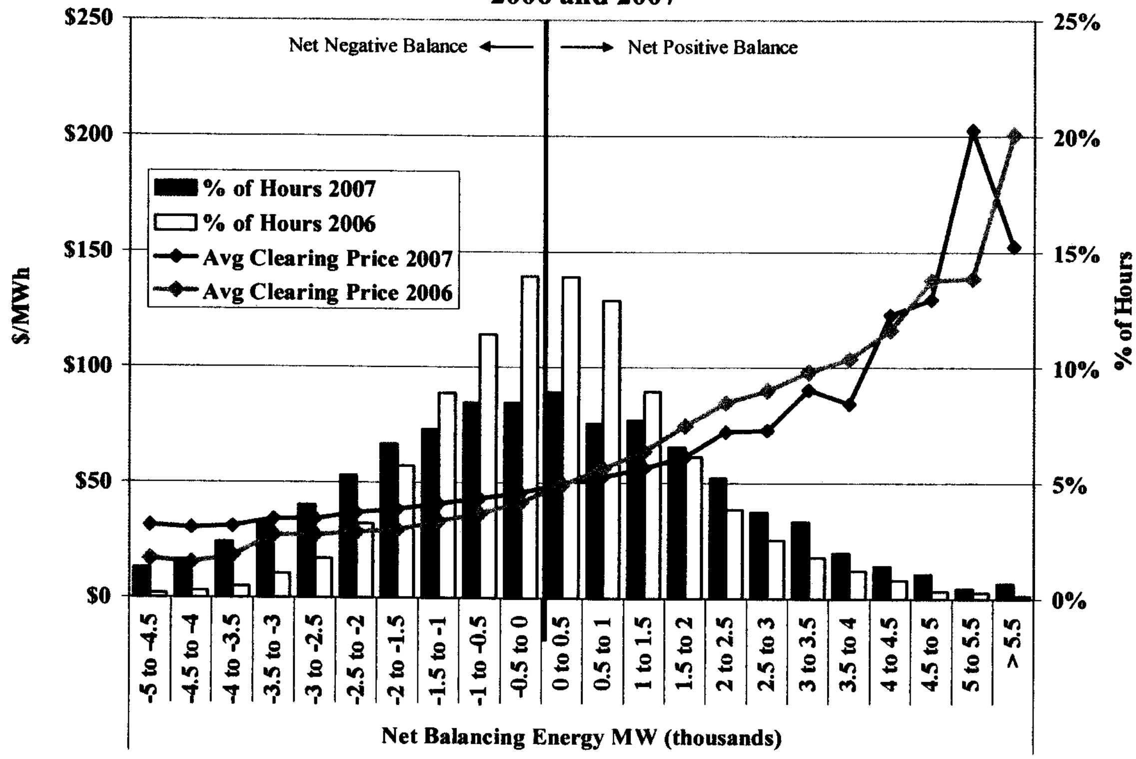


Figure 14 shows a relatively symmetrical distribution of net balancing energy purchases in 2007 centered around zero gigawatts, but the distribution is wider and flatter than 2006. This is consistent with Figure 13 which showed that there were comparable portions of net balancing up and down quantities on average during 2007. In approximately 33 percent of the hourly observations shown, net balancing energy schedules averaged between -1.0 and 1.0 gigawatts. 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. One significant difference from previous years is the drop of energy price at net positive balancing energy deployment levels greater than 5.5 GW. Generally, the occurrences of such significant quantities of balancing energy deployments are representative of times when the available supply (exclusive of reserves) to meet demand is tight. The reasons contributing to this price drop at times of high balancing energy deployments are discussed in subsection I.D.

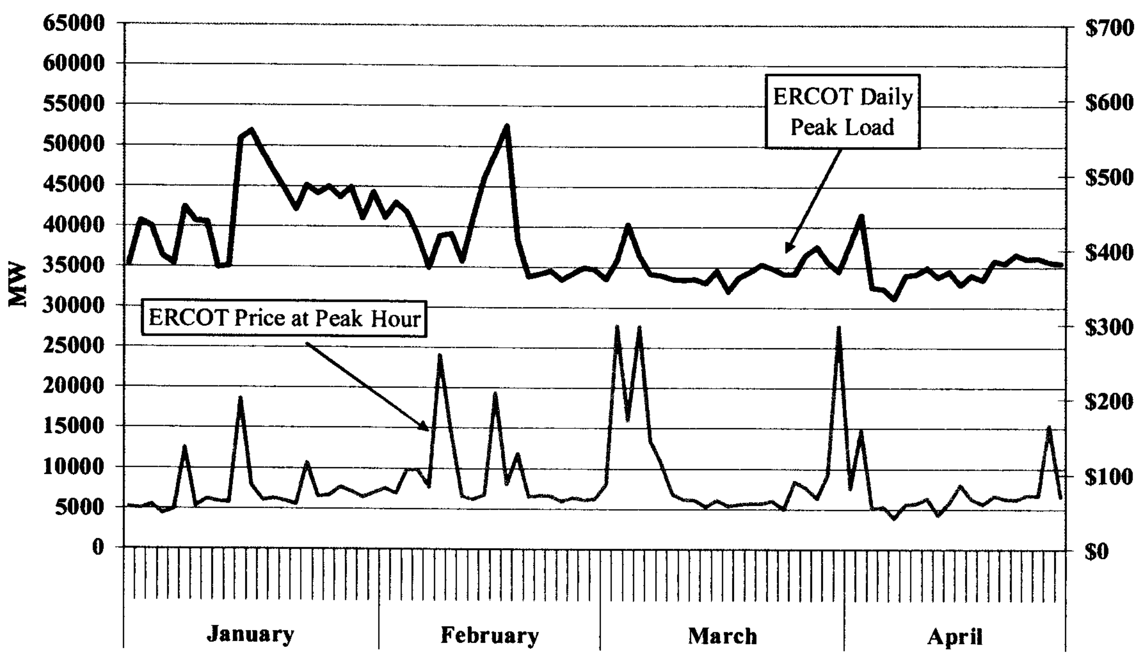
The line plotted in Figure 14 shows the average balancing energy prices corresponding to each level of balancing energy volumes. In an efficiently functioning spot market, there should be little relationship between the balancing energy prices and the net purchases or sales. Instead, one should expect 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 is also consistent with the patterns of prices and volumes in 2005 and 2006. We analyze this relationship more closely in the next sub-section, and in Section II we discuss how scheduling practices and ramping issues explain much of the observed pattern.

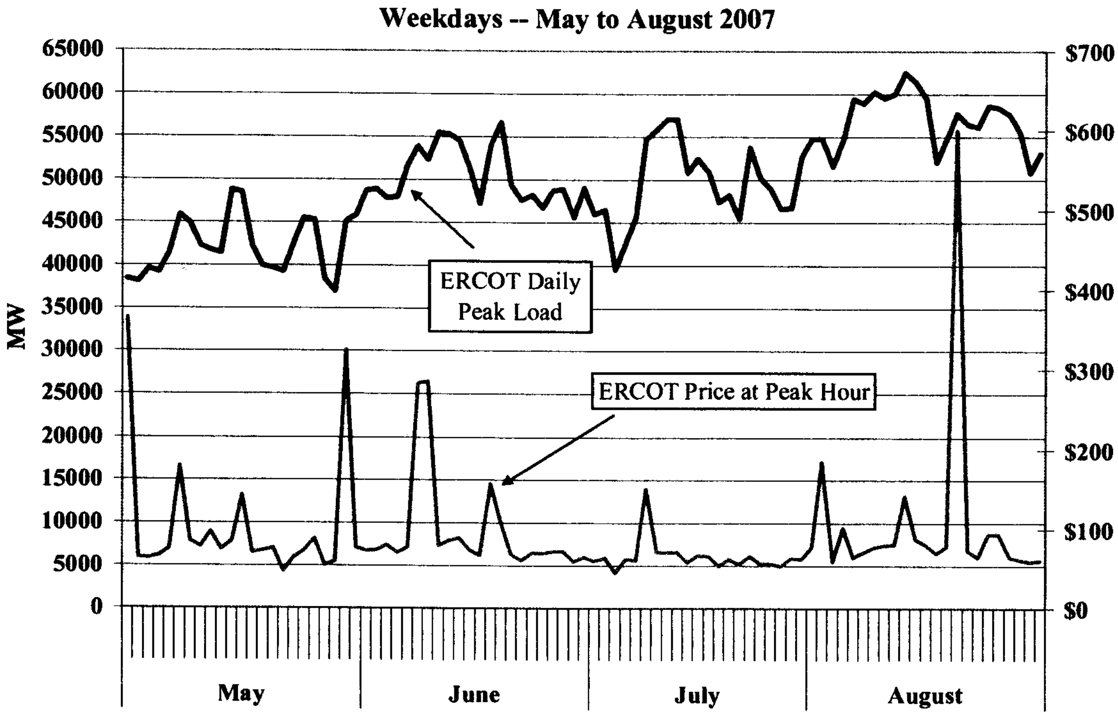
5. Determinants of Balancing Energy Prices

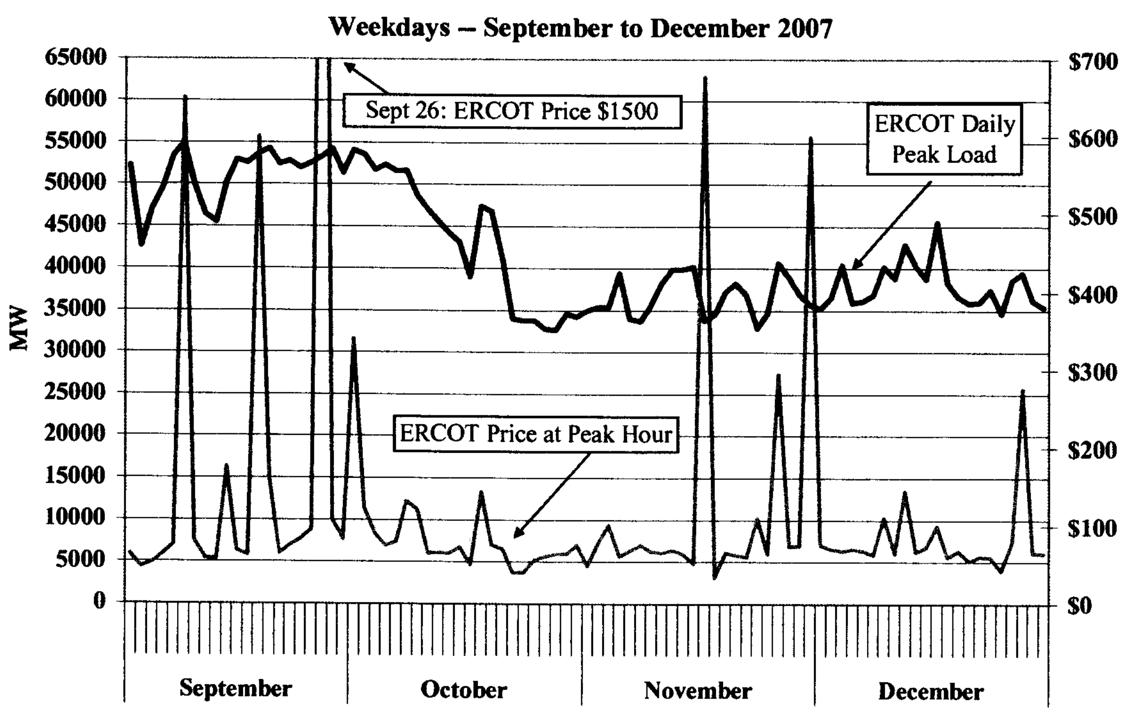
The prior section shows that the level of net sales in the balancing energy market appears to play a significant role in explaining the balancing energy prices. In this section, we examine this relationship in more detail, as well as the role of more fundamental determinants of balancing energy prices,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

Figure 15 shows the average balancing energy price and the actual load in the peak hour of each weekday during 2007.

Figure 15: Daily Peak Loads and Balancing Energy Prices
Weekdays – January to April 2007





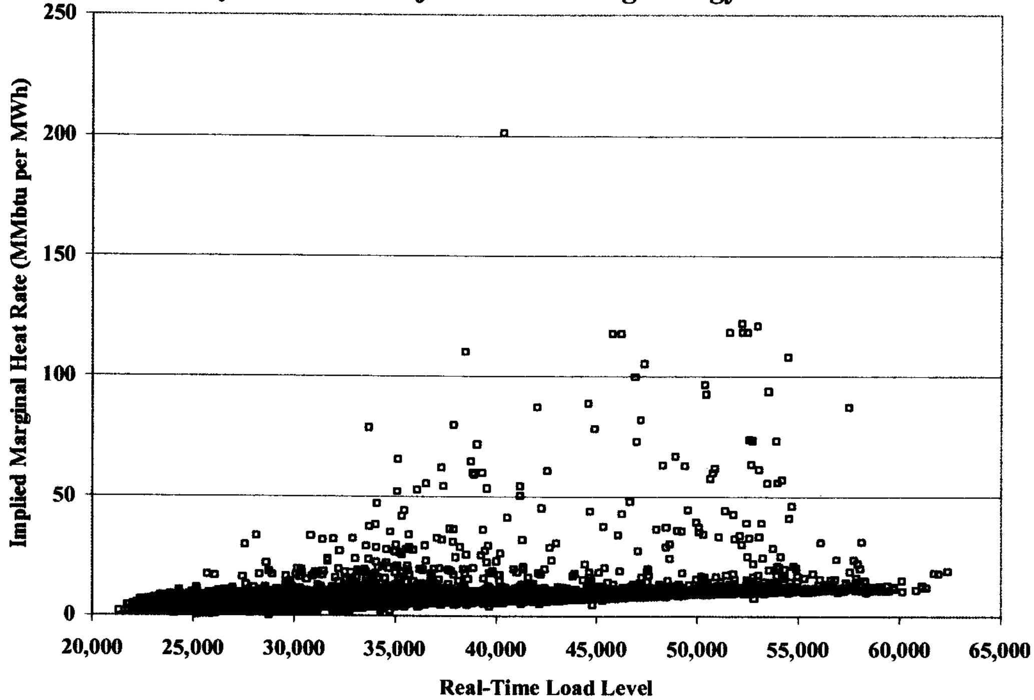


The figures indicates some relationship between high prices (*e.g.*, greater than $200/MWh) and periods when demand is high or rising significantly relatively to the previous days, although the high price occurrences are more common during the shoulder and winter months than during the peak demand summer months.

In an efficient market, we expect for 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. This is primarily due to structural inefficiencies in the balancing energy market that are inherent to the zonal market model, the lack of a centralized unit commitment, load forecast errors, and the fact that the excess online capacity during peak load hours has generally dropped over the last several years.

To further examine the relationship between actual load in ERCOT and balancing energy prices, Figure 16 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 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 16: 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 14 indicates that the net volume of energy purchased in the balancing energy market is often a much stronger determinant of price spikes than the level of demand.

To further examine how the 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 rises during the day from an average of approximately 28 GW at 4 AM to 38 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 17 shows the average load and balancing energy price in each interval from 4 AM through 1 PM in 2007. The price is plotted as a line in the figure while the average load is shown with vertical bars.

Figure 17: Average Balancing Energy Prices and Load by Time of Day Ramping-Up Hours

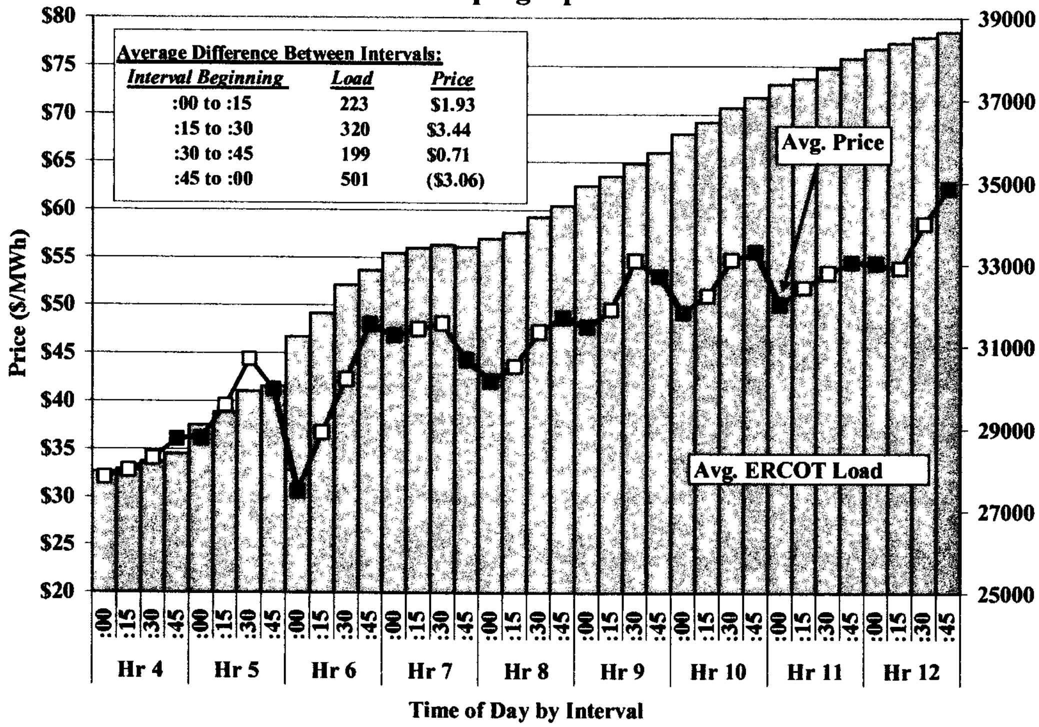


Figure 17 shows that, with the exception of hour 7 and 9, the load steadily increases in every interval and prices generally move upward from about $32 per MWh at 4:00 AM to $62 per MWh at 12:45 PM. If actual load were the primary determinant of energy prices, the balancing energy prices would rise gradually as the actual load rises. However, Figure 17 shows a distinct pattern in the balancing energy prices over the intervals. 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 -$3.06 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 that 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 18 shows this decrease in load by interval, together with the average balancing energy prices for the intervals from 9 PM to 3 AM.

Figure 18: Average Balancing Energy Prices and Load by Time of Day Ramping-Down Hours

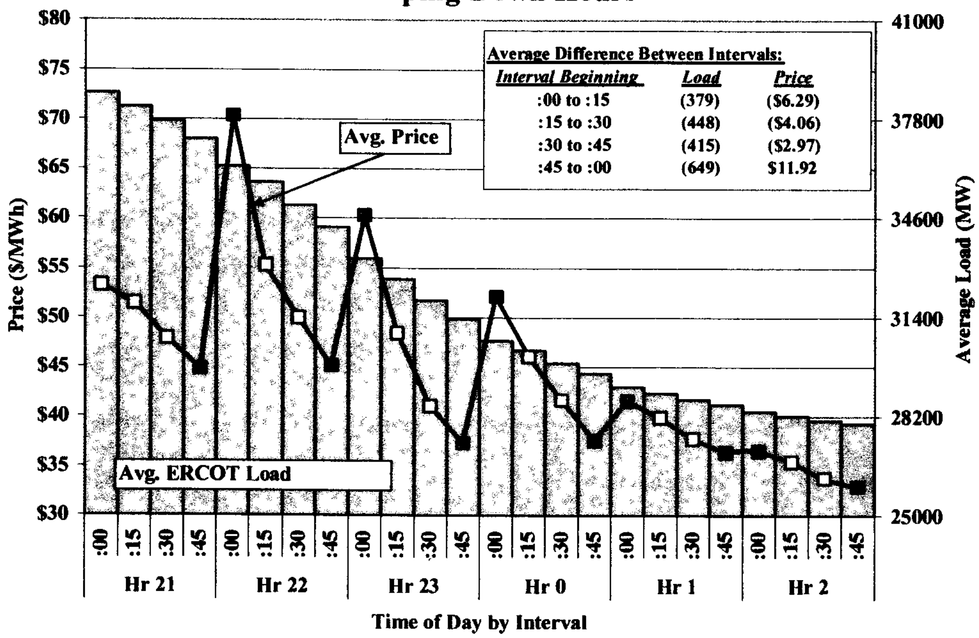


Figure 18 shows that while balancing energy prices decrease over these intervals, they follow a similar pattern as exhibited in the ramping-up hours. The balancing energy price decreases in each interval of the hour before rising substantially in the first interval of the following hour. The balancing energy price increases by an average of $11.92 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 2005 and 2006.[15]

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

15 See 2005 SOM Report and 2006 SOM Report

changes in energy schedules. These scheduling and pricing patterns are examined in detail in Section II below.

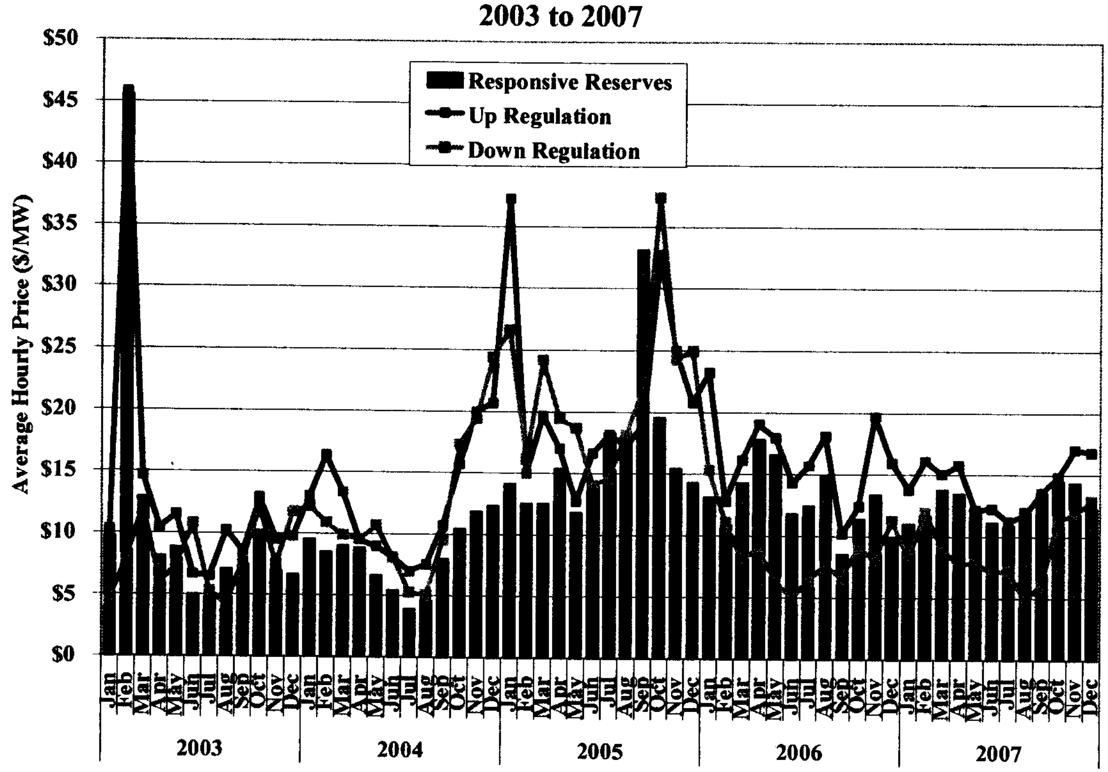
B. Ancillary Services Market Results

The primary ancillary services are up regulation, down regulation, and responsive reserves. ERCOT may also procure non-spinning reserves as needed. 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 2007.

1. Reserves and Regulation Prices

Our first analysis in this section provides a summary of the ancillary services prices over the past five years. Figure 19 shows the monthly average ancillary services prices between 2003 and 2007. Average prices for each ancillary service are weighted by the quantities required in each hour.

Figure 19: Monthly Average Ancillary Service Prices 2003 to 2007



This figure shows that ancillary services prices have generally risen from 2003 to 2005, but that the price levels moderated in 2006 and 2007. Much of these price movements can be attributed to the variations in energy prices that occurred over the same timeframe. Because ancillary services markets are conducted prior to the balancing energy market, participants must include their expected costs 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 level. From 2003 through 2004, regulation down prices were lower than regulation up prices, indicating that the opportunity costs were greater for providers of regulation up. In 2005, the pattern shifted such that regulation down prices were four percent higher on average than regulation up prices. However, in 2006 and 2007, regulation down prices were significantly lower than regulation up prices.

The figure also shows that the prices for up regulation generally exceed prices for responsive reserves. This is consistent with expectations because a supplier must incur opportunity costs to provide both services, while providing up regulation can generate additional costs. These additional costs include (a) the costs of frequently changing output, 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, regulation up providers may have lower opportunity costs than responsive reserves providers to the extent that they are dispatched up to provide regulation.

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 25 percent of hours during 2003, 24 percent of hours during 2004, 23 percent of hours during 2005, 20 percent of hours during 2006, and 14 percent of hours during 2007.

Table 1: Average Hourly Responsive Reserves and Non-Spinning Reserves Prices During Hours When Non-Spinning Reserves Were Procured 2003 to 2007

	2003	2004	2005	2006	2007
Non-Spin Reserve Price	$9.85	$6.83	$25.10	$21.75	$6.07
Responsive Reserve Price	$10.73	$9.10	$28.16	$25.55	$16.74

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.

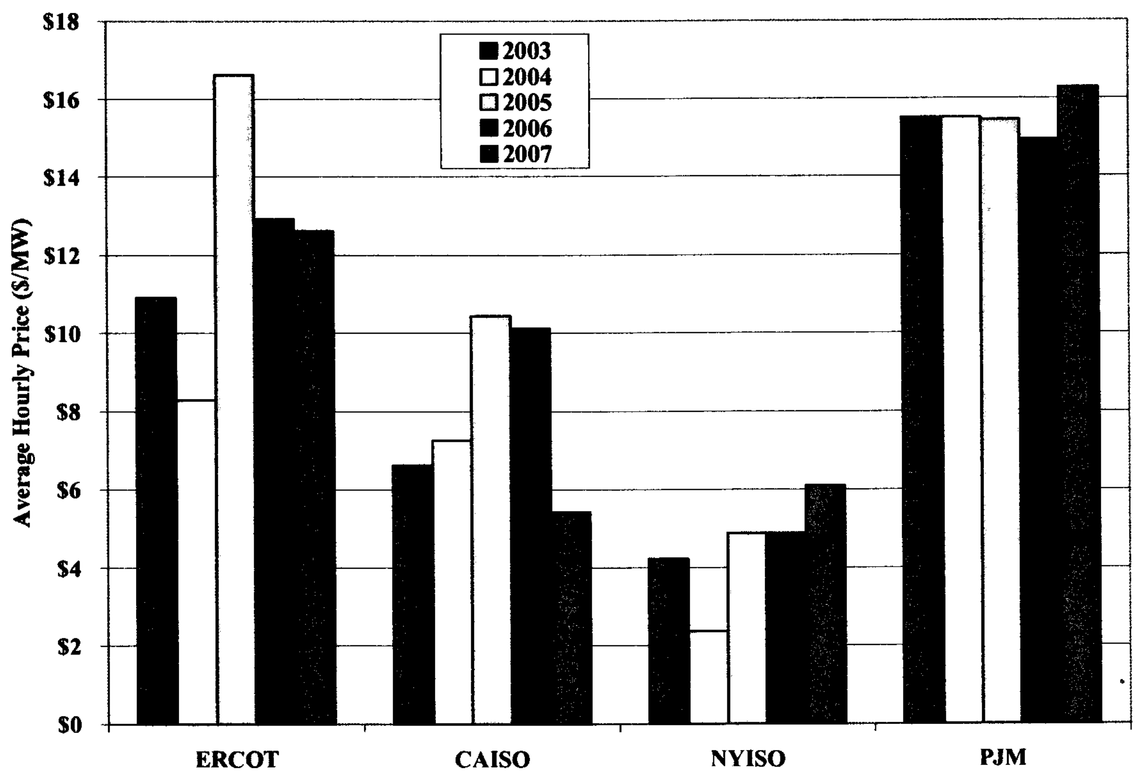
Generators incur two types of costs associated with providing reserves in the ERCOT market. First, reserves providers incur opportunity costs from any profitable sales they forego in the energy market. For generators, this is the same regardless of whether the generator is providing responsive or non-spinning reserves. The second cost that must be considered is the cost of actually being called upon by ERCOT to deploy reserves in real-time. Since generators deployed for reserves are paid for the resulting output at the balancing energy price, there is a risk of being deployed when the balancing energy price is lower than the generator's production costs. While it is also possible for the generator to benefit when the balancing energy price is higher than the generator's costs, this occurs less frequently. Thus, generators providing reserves may run at a loss when they are deployed by ERCOT.

The expected costs of being deployed for reserves are based on the following two factors: (a) the average difference between the resource's production cost and the balancing energy price, and (b) the probability of being deployed. In 2007, about 1.9 percent of the responsive reserves were actually deployed, and 3.1 percent of non-spinning reserves were actually deployed. Therefore, the expected value of the deployment costs may cause the provision of non-spinning reserves to be more costly for some units than responsive reserves.

In general, the purpose of responsive and non-spinning reserves is to protect the system against unforeseen contingencies (*e.g.*, generator outages or load forecast error), rather than for meeting normal load fluctuations. 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 the load requirements. However, in cases when demand is unusually high or unpredictable or the resources projected to be available in real-time may not be sufficient to satisfy the energy demand while meeting the responsive and regulation up reserve requirements, ERCOT will procure non-spinning reserves. This process is a means for ERCOT to implement supplemental generator commitments to increase the supply of energy in the balancing energy market if needed. ERCOT always procures at least 2,300 MW of responsive reserves to ensure adequate protection against the loss of the two largest units.

Responsive reserve prices dropped in 2007 from 2006, but remained higher than the prices observed in 2003 to 2004. Figure 20 shows how the annual average prices in ERCOT from 2003 to 2007 compare to the responsive reserve prices in the California, PJM, and New York wholesale markets. The figure shows that the responsive reserve prices in ERCOT were higher than comparable prices in California, New York, but lower than PJM during 2007.

Figure 20: Responsive Reserves Prices in Other RTO Markets
2003 to 2007



There are a number of reasons why the responsive reserve prices in ERCOT are higher than prices in some of the other regions. First, ERCOT procures substantially more responsive reserves relative to its load than New York, which satisfies a large share of its operating reserve requirements with non-spinning reserves and 30-minute reserves rather than responsive reserves (*i.e.*, 10-minute spinning reserves). However, nearly one half of ERCOT's responsive reserves are satisfied by demand-side resources offered at very low prices, which should serve to offset the fact that ERCOT procures a higher quantity of responsive reserves.

A second reason ERCOT Responsive Reserve prices are higher is because ERCOT (like California and PJM) does not jointly-optimize ancillary services and energy markets. The lack of joint-optimization will generally lead to higher ancillary services prices because participants must incorporate in their offers the potential costs of pre-committing resources to provide reserves or regulation. These costs include the lost profits from the energy market when it would be more profitable to provide energy than ancillary services. Lastly, the offer patterns of market participants can influence these clearing prices. These offer patterns are examined in the next section.

Our next analysis evaluates the variations in regulation prices. The market dispatch model runs every fifteen minutes and produces instructions based on QSE-scheduled energy and balancing energy market offers, while regulation providers keep load and generation in balance by adjusting their output continuously. 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. When demand rises, higher-cost resources must be employed and prices should increase.

Figure 21 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 by the hour of the day. Regulation prices are an average of up and down regulation prices weighted by the quantities of each that are procured.

The figure shows that ERCOT requires approximately 1,230 MW of regulation capability prior to the initial ramping period (beginning at 6 AM). The requirement then jumps up to about 2,000 MW during the steepest ramping hours from 6 AM to 9 AM. The requirement declines to about 1,500 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 rises to approximately 1,800 MW.

Figure 21: Regulation Prices and Requirements by Hour of Day 2007

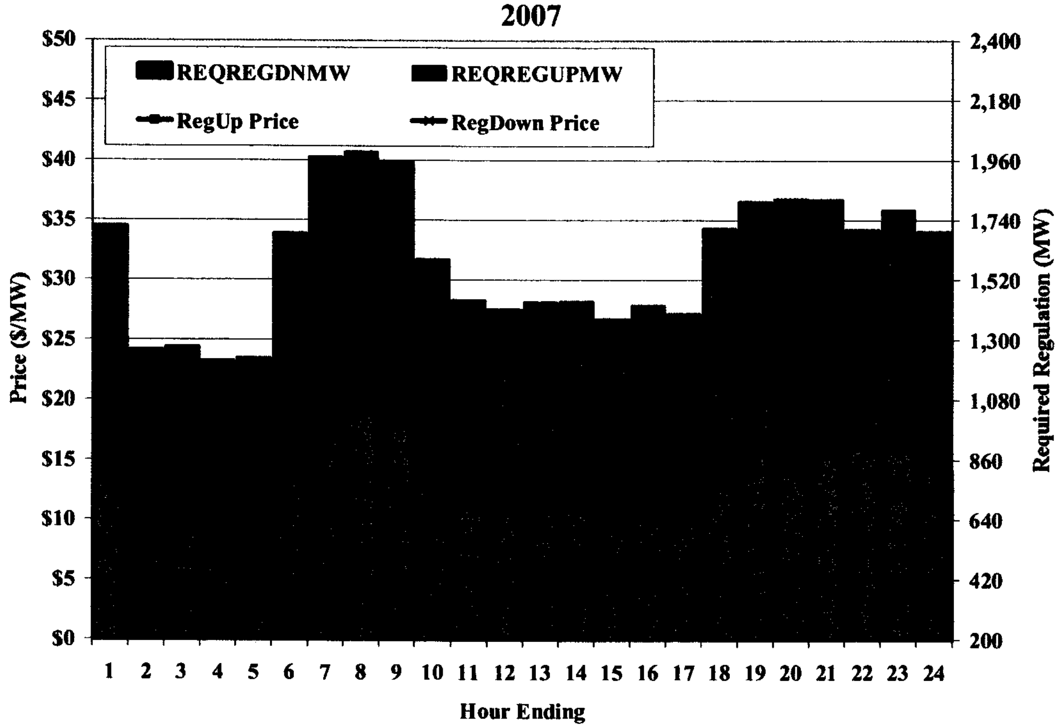


Figure 21 indicates that average regulation prices are generally correlated with the regulation quantity purchased and the typical load pattern in ERCOT. During non-ramping hours, such as overnight and late morning, regulation up and down prices range from $5 to $10 per MW. During the ramping hours in early morning and evening, average regulation up and down prices range from $10 to $23 per MW. In the afternoon hours, regulation up prices range from $10 to $25 and regulation down prices range from $6 to $8 per MW. Regulation up 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 regulation up is lower than in other hours.

Although regulation prices have risen markedly since 2002 due to several factors discussed above, ERCOT has taken significant steps over the same period to reduce regulation market costs. ERCOT has gradually reduced the amount of regulation it procures and uses to keep supply and demand in balance and control frequency on the system. This has directly reduced regulation costs by reducing the quantity scheduled. However, this has also indirectly reduced regulation costs by reducing the clearing prices of regulation. Figure 22 summarizes the average amounts of regulation procured through the auction and/or bilateral arrangements on an annual basis since 2003.

Figure 22: Annual Average Regulation Procurement 2003 to 2007

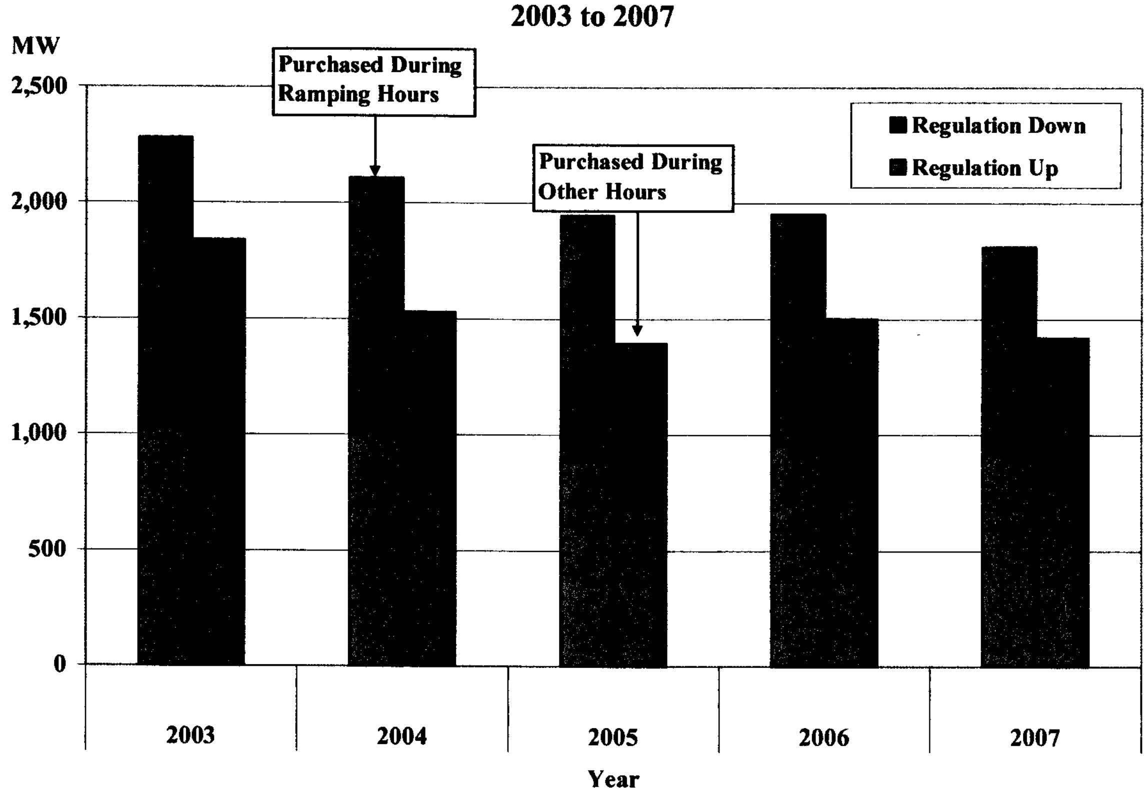


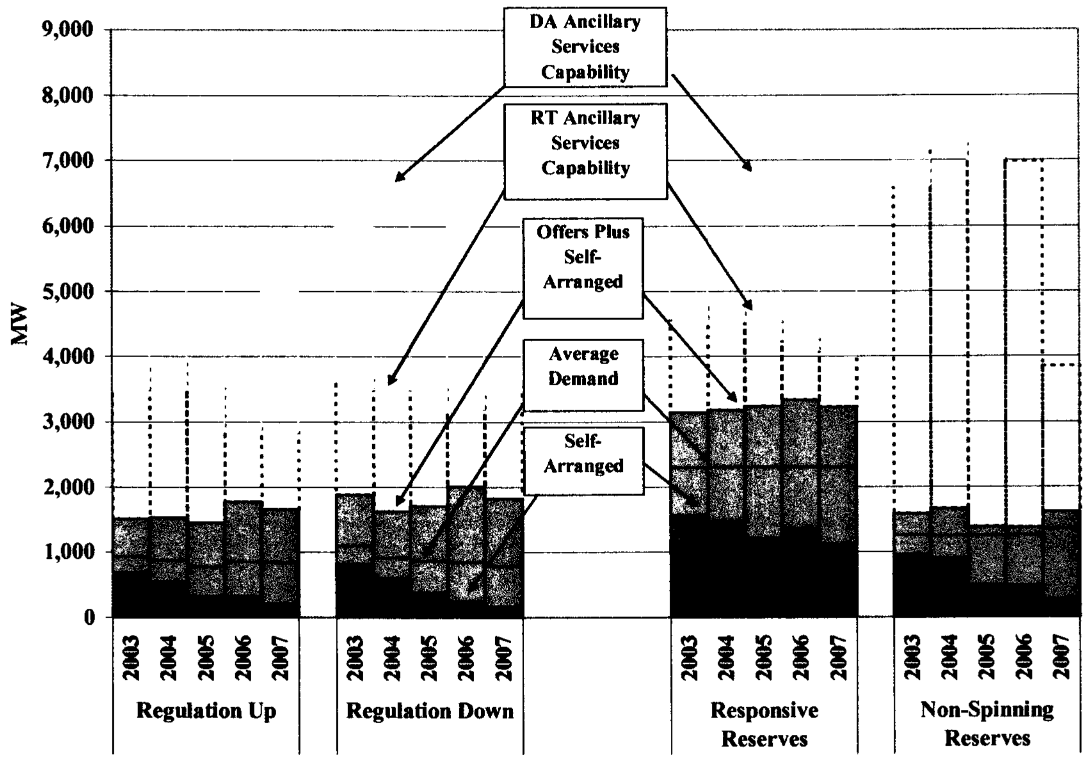
Figure 22 shows that ERCOT has reduced the average regulation quantity scheduled since 2003. The average regulation quantity had steadily declined from 2003-2005, but increased slightly in 2006. In 2007, the average regulation quantities decreased in both the ramping and non-ramping hours compared to 2006. The reduction in average regulation quantities in 2007 is at least partly explained by ERCOT's change in its regulation procurement practices that was implemented in mid-2007. This change allows for a different quantity of regulation to be procured in each hour of each day during a month based upon analysis of historical deployment data, rather than the procurement of fixed quantities over 4 to 5 blocks of hours in each day. The result of this change has been a relative decrease in regulation quantities procured in many hours of each day, with an increase in some hours when regulation demand is the highest. Overall change in the procurement methodology has contributed to a reduction in the average quanttites of regulation procured in 2007.

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 23. 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 23 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 23: Reserves and Regulation Capacity, Offers, and Schedules 2003 to 2007



Note: Non-spinning reserve capability is based on data from generator resource plans. Regulation and responsive reserves capability is based on ERCOT data.

The capability shown in Figure 23 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 23 is not reduced to account for energy produced from each unit, which causes the capability on some resources to be overstated in some hours. Approximately 49 percent of the demand for responsive reserves was satisfied by Loads acting as Resources ("LaaRs"). LaaRs account for only 1,150 MW of the responsive reserves capability shown above, because in 2007 there is a requirement that no more than 50 percent of the 2,300 MW requirement be met with LaaRs.

For non-spinning reserves, Figure 23 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. Hence, the capability that is actually available from a unit in a given hour will generally be less than the amounts shown in this figure because a portion will be used to produce energy.

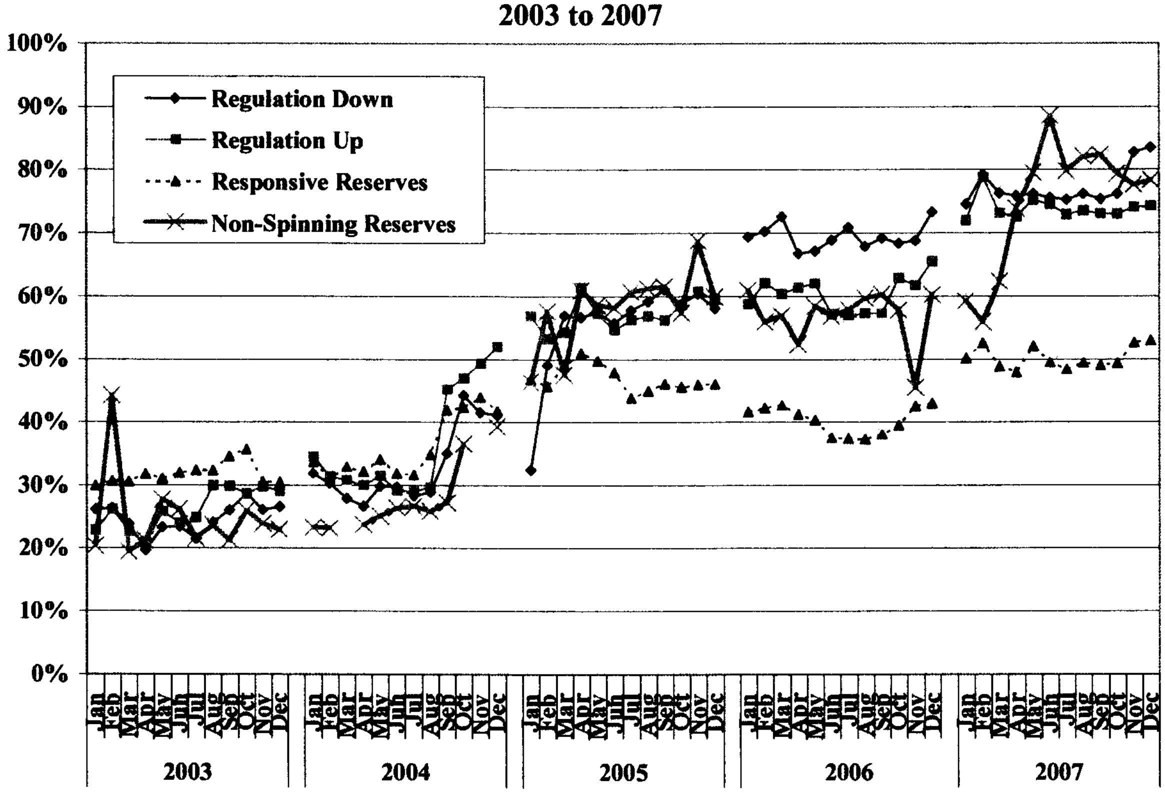
Figure 23 shows that except for responsive reserve in 2006 and 2007, in which about 54 percent and 52 percent respectively of available responsive reserve capacity was offered, less than one-half of each type of ancillary services capability was offered during the year from 2003 to 2007. One explanation 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 associated with this approach. 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.

Figure 23 shows modest changes in the amount of day-ahead ancillary services capability between 2003 and 2007. The installation of several gigawatts of new capacity has contributed to overall capability, while the continued mothballing and retirement of certain units has reduced capability.

Finally, although market participants increasingly rely on the auction market to procure these services, Figure 24 shows that a significant share of these services is still self-supplied. 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 24 shows the share of each type of ancillary service that is purchased through the ERCOT market.

Figure 24: Portion of Reserves and Regulation Procured Through ERCOT 2003 to 2007

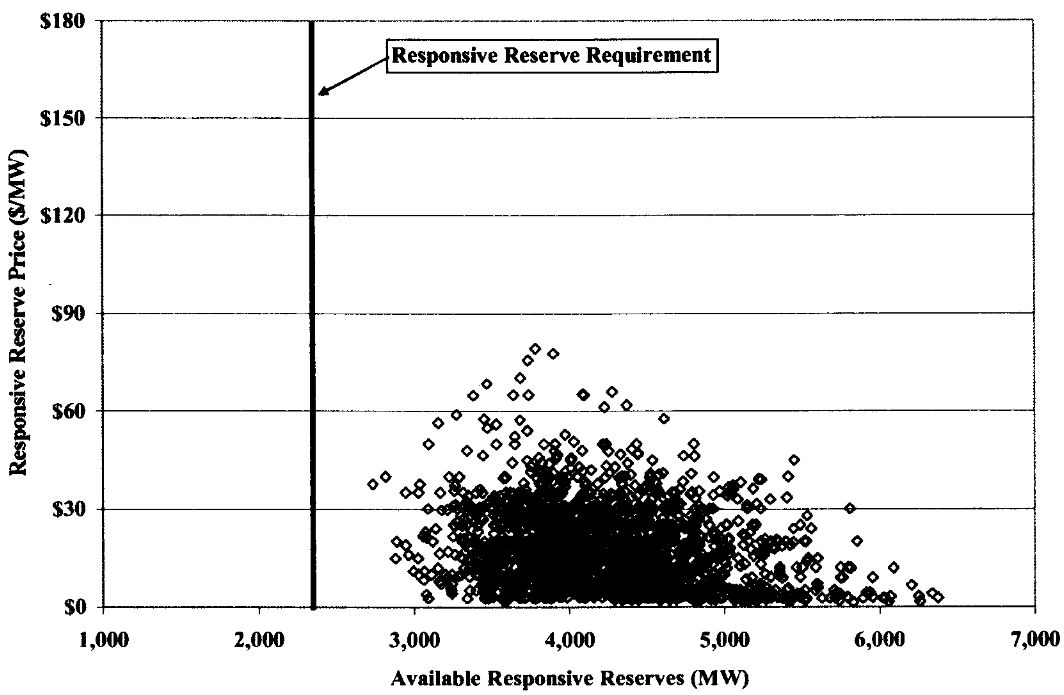


This figure shows that purchases of all ancillary services from the ERCOT markets have generally increased over time, although the purchases of responsive reserve from the ERCOT market have dropped slightly in 2006 (*i.e.*, the quantity of self-arranged responsive reserve has increased slightly). As 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.

The next analysis in this section evaluates the prices prevailing in the responsive reserves market during 2007. Prices in this 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 25 plots the hourly real-time responsive reserves capability against the responsive reserves prices in the peak afternoon hours (2 PM to 6 PM). The capability calculated for this analysis reflects the actual energy output of each generating unit and the actual dispatch point for LaaRs. Hence, units producing energy at their maximum capability will have no available responsive reserves capability and, consistent with ERCOT rules, the responsive reserve that can be provided by each generating unit is limited to 20 percent of the unit's maximum capability. The figure also shows the responsive reserves requirement of 2,300 MW to show the amount of the surplus in each hour.

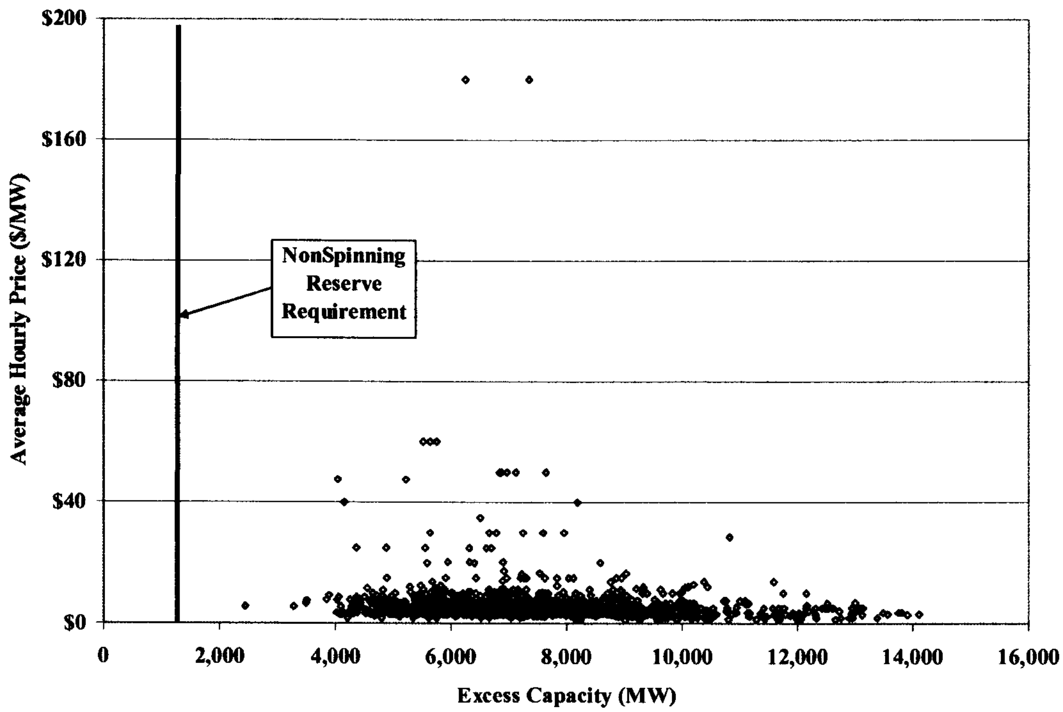
Figure 25: Hourly Responsive Reserves Capability vs. Market Clearing Price Afternoon Peak Hours – 2007



Compared to prior years, this figure indicates a much stronger relationship between the hourly available responsive reserves capability in real time and the responsive reserves prices. In a well functioning-market for responsive reserves, we would expect excess capacity to be negatively correlated with the clearing prices. Additional improvements should result from jointly optimizing the operating reserves and energy markets, which is currently being developed for implementation in the nodal market (day ahead co-optimization, but not real-time).

Non-spinning reserves are 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 26 shows the relationship between excess available non-spinning reserves capability and the market clearing price in the non-spinning reserves auction for the afternoon hours in 2007.

Figure 26: Hourly Non-Spinning Reserves Capability vs. Market Clearing Price All Hours 2007



Like the previous analysis of responsive reserves, the results shown in Figure 26 indicate a stronger correlation between non-spinning reserves prices and the quantity of available reserves capability in real time as compared to the results in prior years. In a well functioning-market for non-spinning reserves, we would expect excess capacity to be negatively correlated with the clearing prices.

C. Net Revenue Analysis

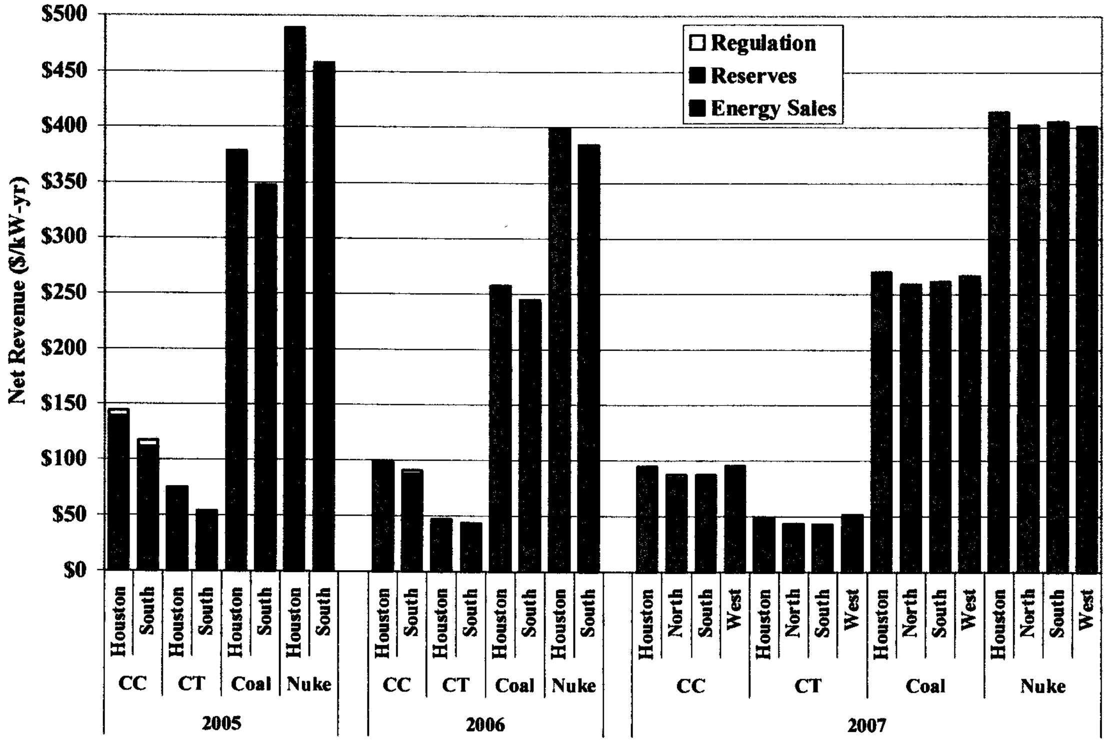
Net revenue is defined as the total revenue that can be earned by a generating unit less its variable production costs. Hence, it is the revenue in excess of short-run operating costs and is available to recover a unit's fixed and capital costs. Net revenues from the energy, operating reserves, and regulation 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 between 2004 and 2007 by various types of generators in each zone.

Figure 27 shows the results of the net revenue analysis for four types of units. 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 27: Estimated Net Revenue 2005 to 2007



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.

The highest net revenues were in the North and Houston zones while lowest net revenue levels were in the South zone. Because the net revenues for the North and West zones in 2005 and 2006 fall within the range of the other zones, we do not show their net revenues in the figure for legibility. Although the analysis indicates that a generator operating in the North zone or in Houston would have earned more net revenue than a generator in the South zone, the relative costs of investment in these zones are also important in determining the most attractive locations for new investment.

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 27 shows that the net revenue fell in 2006 in each zone compared to 2005, and stayed at comparable levels in 2007; however, net revenue remained higher in 2006 and 2007 than in years prior to 2005. 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 is approximately \$70 to \$95 per kW-year. The estimated net revenue for a new gas turbine in 2007 is approximately \$44 per kW-year, which is lower than the estimated net revenue required for new entry. For a new combined cycle unit, the estimated net revenue requirement is approximately \$105 to \$135 per kW-year. The estimated net revenue in 2007 for a new combined cycle unit is approximately \$88 per kW-year, which is also lower than the estimated net revenue required for new entry. The annual revenue requirements above are for new construction. Other types of projects may have substantially lower investment costs, such as projects to upgrade existing facilities, return mothballed units to service or to re-power old sites.

Prior to 2003, 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. The annual fixed costs (including capital carrying costs) are estimated at \$190 to $245 per kW-year for a new coal unit and $280 to $390 per kW-year for a new nuclear unit. Net revenues were at the lower ends of these ranges in 2004, but exceeded them from 2005 to 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 2005 to 2007 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.

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 28 compares estimates of net revenue for each of the auction-based wholesale electricity markets in the U.S.: (a) the ERCOT North Zone, (b) the California ISO, (c) the New York ISO, (d) ISO New England,[16] and (e) the 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.[17]

16 The ISO-New England revised its methodology in 2005 to include estimated revenues from its forward reserves market for the 10,500 BTU/kWh unit. Although this market also existed in 2004, the figures for 2004 do not include forward reserves revenue.

17 The California ISO does not report capacity and ancillary services net revenue separately, so it is shown as a combined block in Figure 28. 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, and, in 2002, the ISO–New England reported net revenues for a 6,800 BTU/kWh combined-cycle unit. The California ISO revised its methodology in 2006 to consider a theoretical new combined-cycle unit to participate in both the Real-time and Day-ahead market, with the net revenues updated from 2004 to 2006.

Figure 28: Comparison of Net Revenue of Gas-Fired Generation between Markets 2005 to 2007

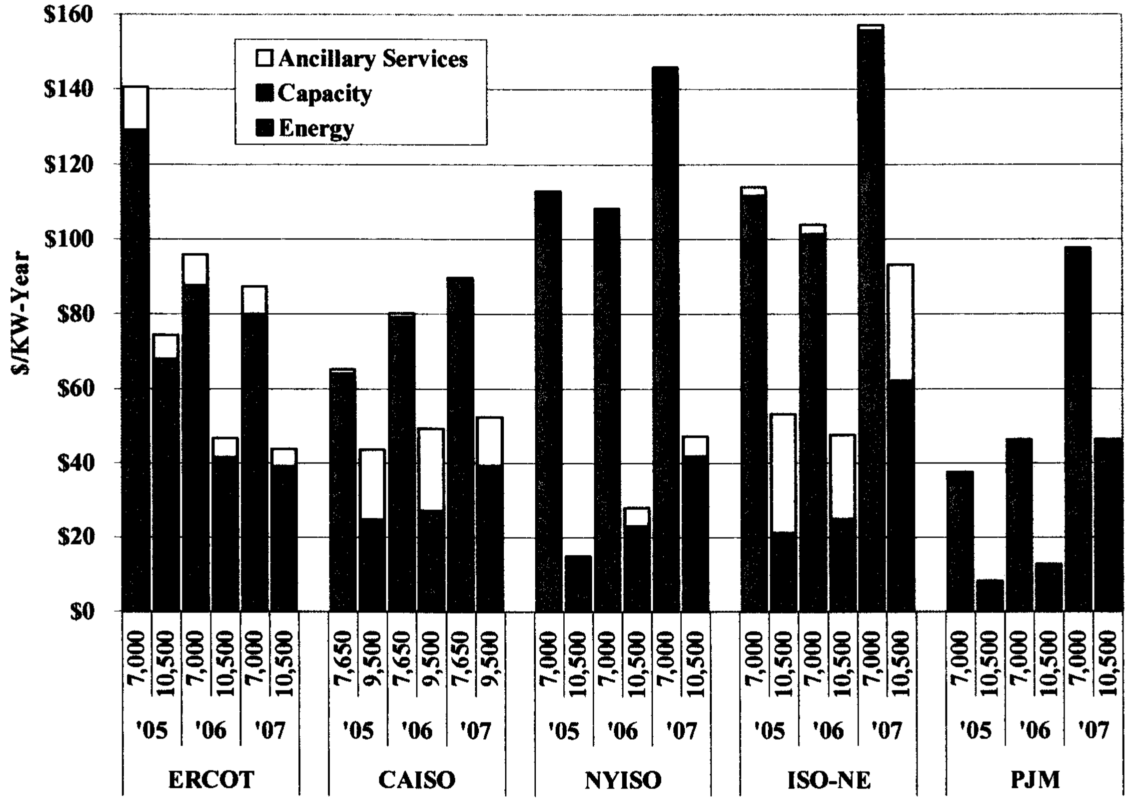


Figure 28 shows that net revenues increased in California, New York, New England and PJM from 2005 to 2007, and decreased in ERCOT. ERCOT is much more dependent on natural gas than the other markets. The decrease in natural gas prices in some of the other regions over this period does not translate as directly into lower electricity prices because natural gas units are displaced in many hours by other types of units. Also, some other markets experienced higher load than previous years such as in PJM, which also led to higher energy price than 2006. Capacity revenue was higher in ISO-NE and PJM due to the recent implementation of capacity market reforms. In PJM, the prior capacity market construct was replaced by the Reliability Pricing Model (RPM) which resulted in higher capacity revenue. In ISO-NE, the implementation of the Forward Capacity Market (FCM) in 2007 also led to an increase in the capacity price. 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.

The net revenue outcomes in the ERCOT markets in 2007 were primarily affected by the following factors:

- Although continuing to decline relative to prior years, planning reserve margins in 2007 were approximately 14.6 percent, which remains above the minimum requirement of 12.5 percent. Excess capacity lowers net revenue by reducing prices, whereas relatively low reserve margins can cause net revenue levels to substantially exceed the annualized cost of a new unit.
- Natural gas prices were relatively flat in 2007 compared to 2006, but remained at levels significantly higher than the years prior to 2005. Thus, net revenue for coal and nuclear units continued to be at levels sufficient to support new entry.
- The effectiveness of the Scarcity Pricing Mechanism was challenged by several operational factors, which are discussed in more detail in the next subsection.
- The competitive performance of the ERCOT market continued to improve in 2007.

In a market with efficient pricing, spot price signals should indicate when and where new generation investment is needed and when existing generation should be retired. Under the nodal market design, it will be important to ensure that the market sends efficient signals for new investment and retirement. This is primarily accomplished in one of two ways:

- A capacity market; and/or
- Shortage pricing provisions to ensure that prices rise appropriately in the energy and ancillary services markets to reflect the true costs of shortages when resources are insufficient to satisfy both the energy and ancillary services requirements.

The PUCT adopted rules in 2006 that define the parameters of an energy-only market. These rules include a Scarcity Pricing Mechanism ("SPM") that provides for a gradual increase in the system-wide offer cap 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 is typically very small. The new rules also eliminated the provisions in the PUCT rules that required *ex post* pricing adjustments during shortage conditions. The next subsection provides a review of the effectiveness of the SPM in 2007.

D. Effectiveness of the Scarcity Pricing Mechanism in 2007

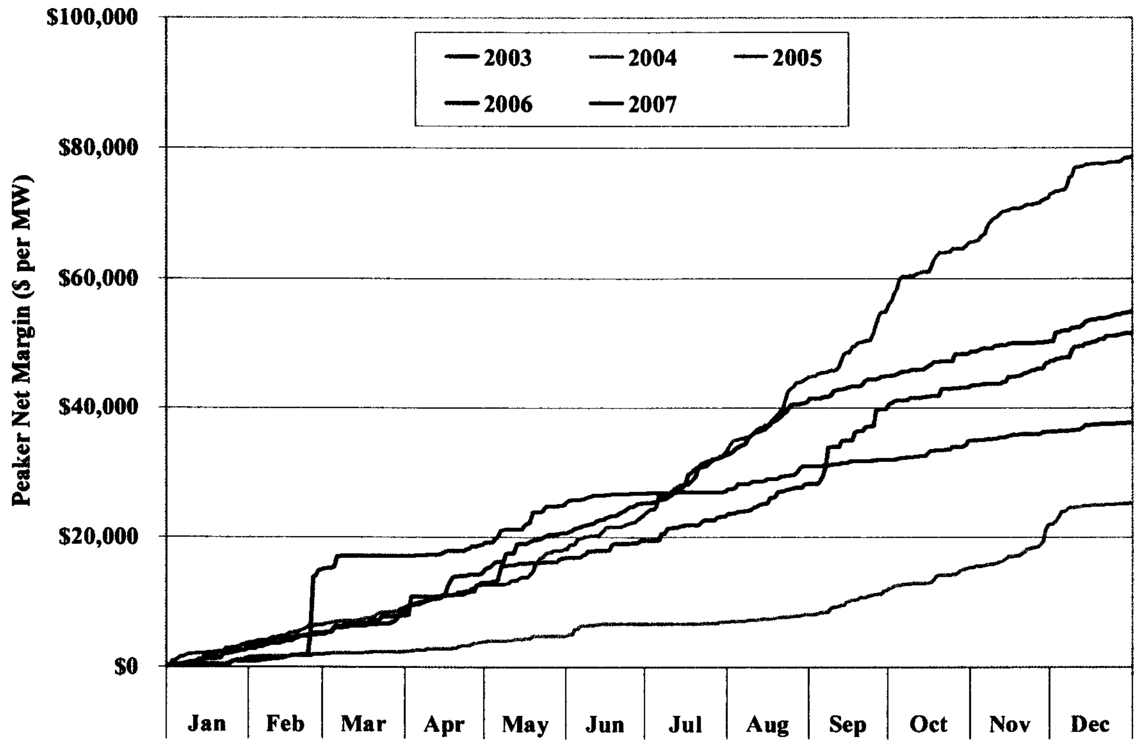
The PUCT's energy-only market rule provides that the IMM may conduct an annual review of the effectiveness of the SPM. This subsection provides an assessment of the results of the first full year of operation under the new rules.

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 29 shows the cumulative PNM that would have been produced for each year from 2002 through 2007.[18]

Figure 29: Peaker Net Margin
2002 to 2007



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 $60 to $85 per kW-year (i.e., $60,000 to $85,000 per MW-year). Thus, as shown in Figure 29 and consistent with the previous findings in this section relating to net revenue, the PNM reached the level sufficient for new entry in only one of the last five years (2005).

There were several factors that challenged the effectiveness of the SPM in 2007, including:

- Frequent out-of-merit deployments by ERCOT during declared short-supply conditions;
- The dependence on market participants to submit offers at or near the offer cap to produce scarcity level prices during legitimate shortage conditions; and

18 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.

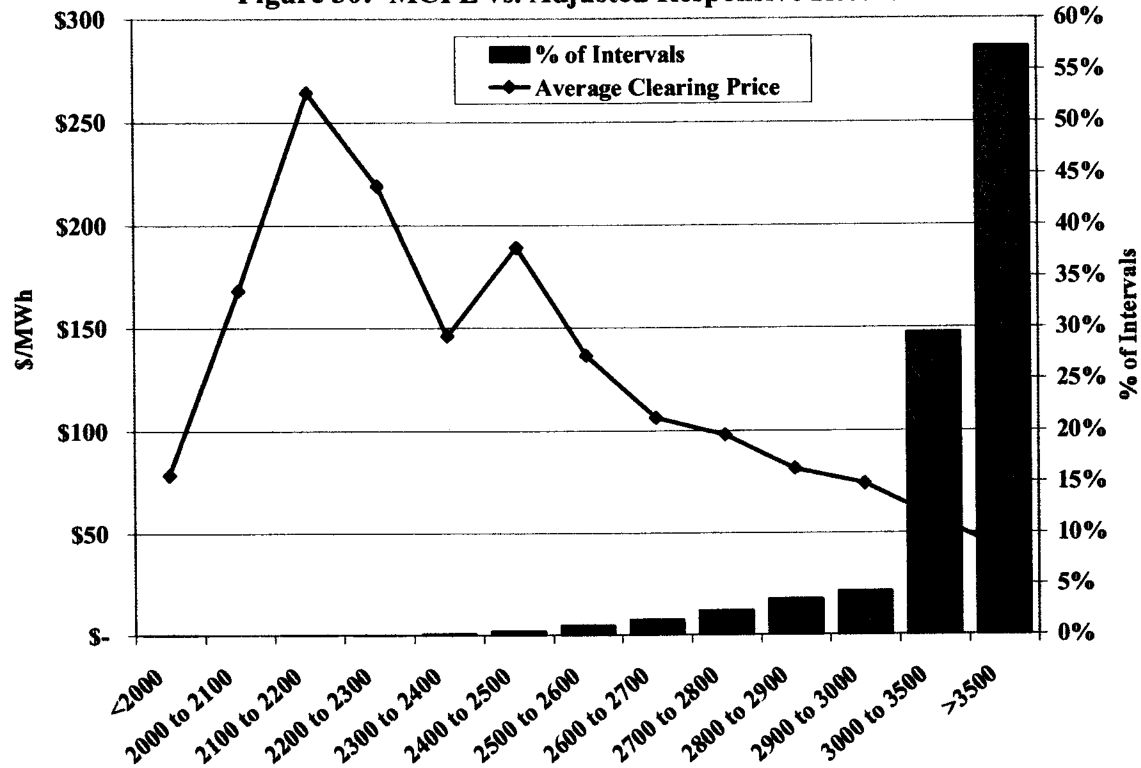
- A strong positive bias in ERCOT's day-ahead load forecast that tended to regularly commit online resources in excess of the quantity required to meet expected demand and operating reserve requirements.

1. Out-of-Merit Deployments during Shortage Conditions

In 2007, ERCOT implemented a new operating procedure 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.

From a reliability perspective, the interim use of the discount factor by ERCOT is understandable, although the long-term objective should be to establish confidence in the maximum ratings reported for each generating unit. In fact, through the implementation of Protocol Revision Request ("PRR") No. 750, an unannounced testing procedure was established in early 2008 that should achieve this objective and result in the eventual elimination of the discount factor. However, 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. Figure 30 illustrates the effect of the use of the discount factor and associated OOM deployments during 2007.

Figure 30: MCPE vs. Adjusted Responsive Reserve



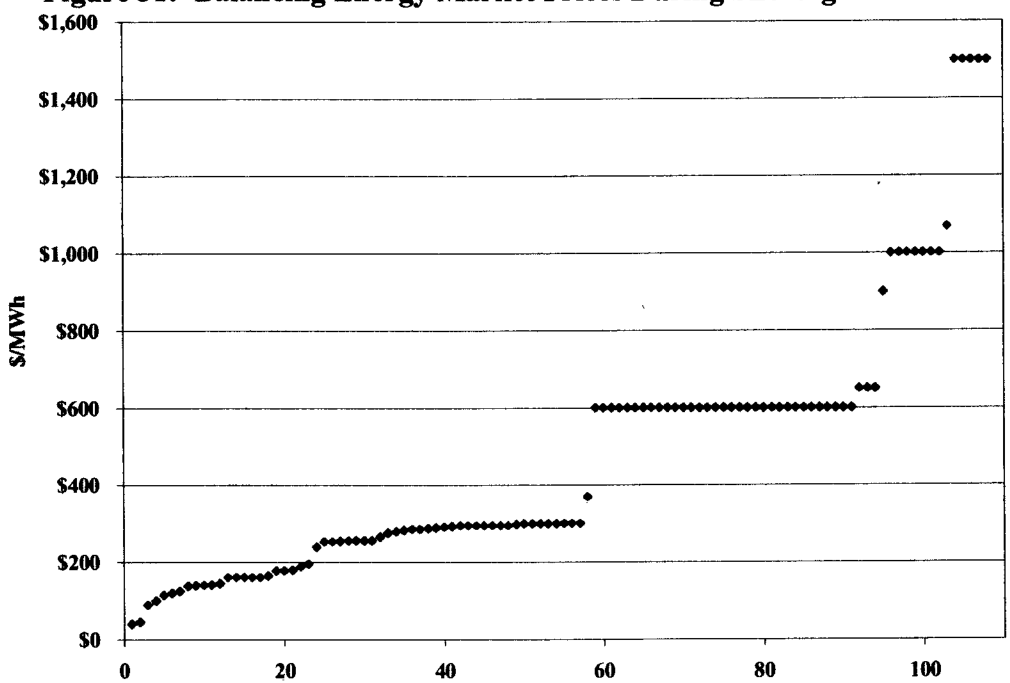
As shown in Figure 30, the average price rose in 2007 as ARR dropped from 3,500 to 2,500 MW. However, once ARR reached 2,500 MW, the average price dropped, which can be attributed to the initial OOM actions taken by ERCOT when ARR reaches 2,500 MW. Prices resumed their increase for ARR levels between 2,100 and 2,400 MW, but dropped significantly at ARR levels less than 2,100 MW. Although only approximately 0.6 percent of the hours in the year (about 50 hours) experienced ARR less than 2,500 MW, it is critical to the success of the energy-only market design and the achievement of long-term resource adequacy objectives that prices be set efficiently during these relatively infrequent 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. Hence, implementation of PRR No. 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.

2. Dependence on High-Priced Offers by Market Participants

As previously discussed, 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 supply of resources is insufficient to simultaneously meet both energy and operating reserve requirements) to provide an appropriate price signal for demand response and new investment when required. 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 smaller market participants. Figure 31 shows the balancing market clearing prices during the 108 15-minute intervals in 2007 when all available balancing energy was exhausted.[19]

Figure 31: Balancing Energy Market Prices During Shortage Intervals



As shown in Figure 31, the prices during these 108 shortage intervals in 2007 ranged from $40 per MWh to the offer cap of $1,500 per MWh (prior to March 1, 2007, the offer cap was $1,000 per MWh). Also evident from the data in this figure are distinct offer thresholds at about $300

19 Intervals with zonal congestion or non-spinning reserve deployments are excluded.

per MWh and at $600 per MWh. Hence, although each of these data points represents identical system conditions in which all available balancing energy was exhausted, the pricing outcomes are widely varied, indicating that relying upon the submission of high priced offers by some market participants to produce scarcity prices during shortage conditions was rather unreliable during 2007.

More reliable and efficient shortage pricing could be achieved by establishing pricing rules that automatically produce scarcity level prices when defined shortage conditions exist on the system. 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, and it would 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.

While such changes would prove difficult with the current zonal systems, we recommend consideration of the future implementation of operating reserve demand curves in the context of the nodal market design to achieve these objectives. Additionally, the future implementation of real-time co-optimization of energy and reserves should also be considered as a nodal market enhancement to further improve the efficient operation of the real-time market.

3. 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 32: Day Ahead Load Forecast Error

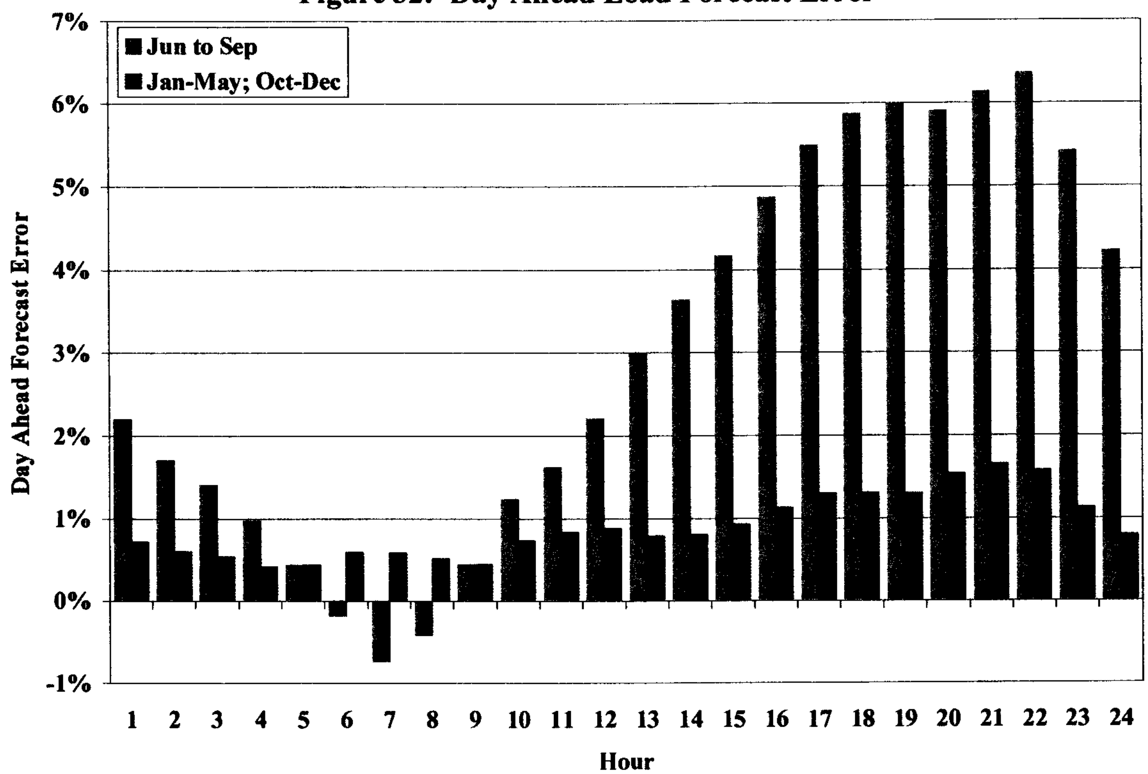


Figure 32 shows the average hourly day-ahead load forecast error for the summer months of June through September, and also for the months of January through May and October through December. In this figure, positive values indicate a day-ahead load forecast that was greater than the actual real-time load. These data indicate a positive bias (*i.e.*, over-forecast) in the day-ahead load forecast over almost all hours in 2007, 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 445 MW for the non-summer months, and an average over-forecast of 2,650 MW for the four summer months.

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.[20] Thus, we recommend that ERCOT review the causes of the positive bias in its day-ahead load forecast.

In conjuction, with the day-ahead load forecast review, ERCOT should explore potential changes to its reserve procurement policies and its day-ahead and supplemental unit commitment procedures in an effort to enhance the efficiency of its unit commitment processes while still satisfying reliability requirements. Additionally, although not a significant issue for most of 2007, this review should include the effects of the considerable increase in the installed wind generation capacity in the ERCOT region during the last quarter of 2007 and in 2008 and beyond, as the substantial addition of more unpredictable and uncontrollable resources has significant implications related to efficient and reliable unit commitment and real-time operations.

4. Recommended Modifications to the SPM

The issues described in this subsection influence the effectiveness of the SPM, but their resolution does not require changes to the SPM as set forth in PUCT rules. However, we do recommend one change to the SPM that would require a modification to the existing rules.

In the PUCT rules, the price that is used to calculate the peaker net margin is measured as the price at an ERCOT-wide hub. Essentially, this is an average price for the ERCOT market. When there is congestion on the system, prices across the ERCOT market will differ, with the import-constrained areas experiencing higher prices than the export-constrained areas. Hence, from the perspective of providing the price signal to attract new entry, a more relevant measure is a regional price that can more precisely measure where that price signal has been provided. The addition of new capacity in generally import-constrained areas not only serves to help alleviate the magnitude of congestion, but also contributes to achieving the system-wide adequacy objectives.

Therefore, we recommend that the price that is used in the peaker net margin calculation in the PUCT's SPM rules be modified to be a set of regional prices, and that the cumulative peaker net

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

margin be calculated as the highest cumulative regional value. Once the annual cumulative peaker net margin threshold set forth in the PUCT rules is reached for any of the defined regions, the transition from the high system offer cap to the low system offer cap would occur for all regions for the duration of the annual SPM cycle.

In the zonal market, the appropriate regions would be the congestion management zones. In the nodal market, the areas represented by the defined nodal load zones may be valid regional definitions, although other reasonable regional definitions could be considered.

II. Scheduling and Balancing Market Offers

In the ERCOT market, QSEs submit balanced load and energy schedules prior to the operating hour. These forward schedules are initially submitted in the day ahead and can be subsequently updated during the adjustment period up to sixty minutes before the operating hour. QSEs are also required to submit a resource plan that indicates the units that are expected to be on-line and satisfying their scheduled energy obligations. Under ERCOT's relaxed balanced schedules policy, the load schedule is not required to approximate the QSE's projected load. When a QSE's load schedule is less than its actual real-time load, its generation is under-scheduled and it will purchase its remaining energy requirements in the balancing energy market at the balancing energy price. Likewise, when a QSE's load schedule is greater than actual load, its generation is over-scheduled and it will sell the residual in the balancing energy market at the balancing energy price.

The QSE schedules and resource plans are the main supply and demand components of the ERCOT market. In this section, we evaluate certain aspects of the QSE schedules and resource plans and we draw conclusions about balancing energy prices, market participants' behavior, and the efficiency of the market design.

This section analyzes a number of issues, beginning with load scheduling by QSEs. The analysis focuses on the degree to which load schedules depart from actual load levels. Our second analysis focuses on the balancing energy market and, in particular, how scheduling patterns affect balancing energy deployments and prices. The third analysis evaluates the rate of participation in the balancing energy market.

A. 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 scheduled load be reflective of the actual load of a QSE. Additionally, a QSE may balance some or all of 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 33 shows a scatter diagram that plots the ratio of the final load schedules to the actual load level during 2007. The ratio shown in the figure will be greater than 100 percent when the final load schedule is greater than the actual load.

Figure 33: Ratio of Final Load Schedules to Actual Load All ERCOT

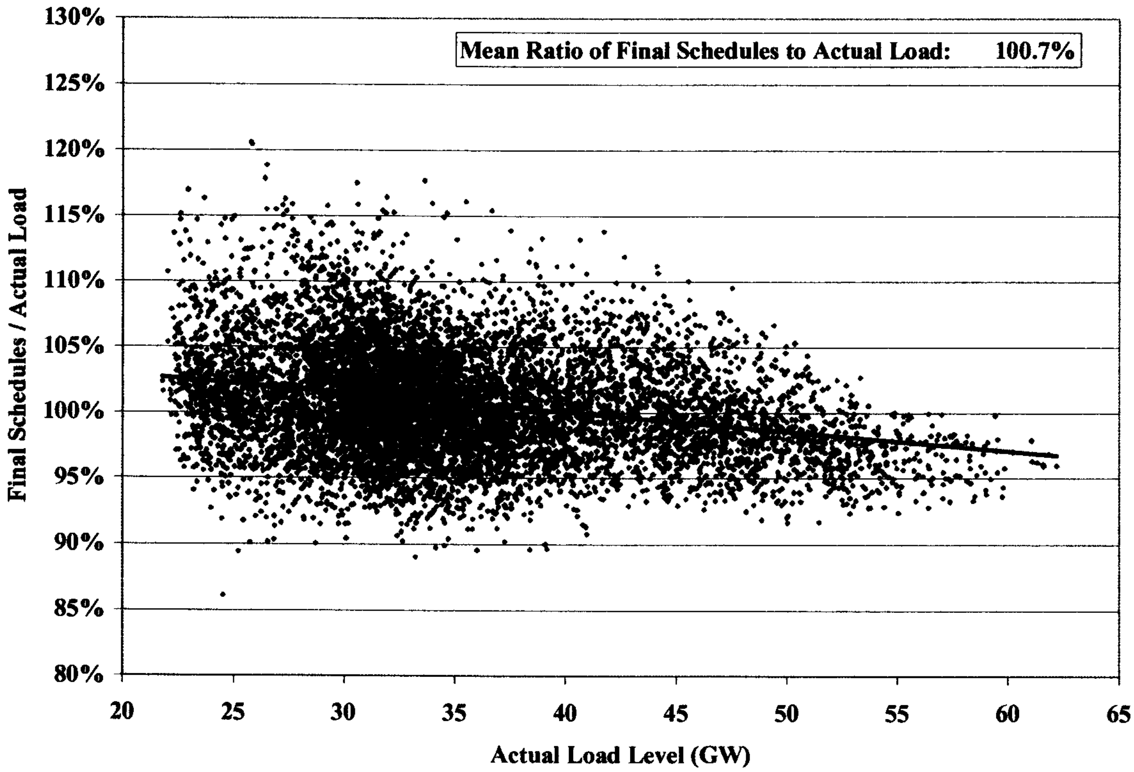


Figure 33 shows that final load schedules generally come very close to actual load in the aggregate, as indicated by an average ratio of the final load schedules to actual load of 100.7 percent. However, the figure also includes a trend line indicating that the ratio of final load schedules to actual 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 2006, 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 34: Average Ratio of Final Load Schedules to Actual Load by Load Level All Zones

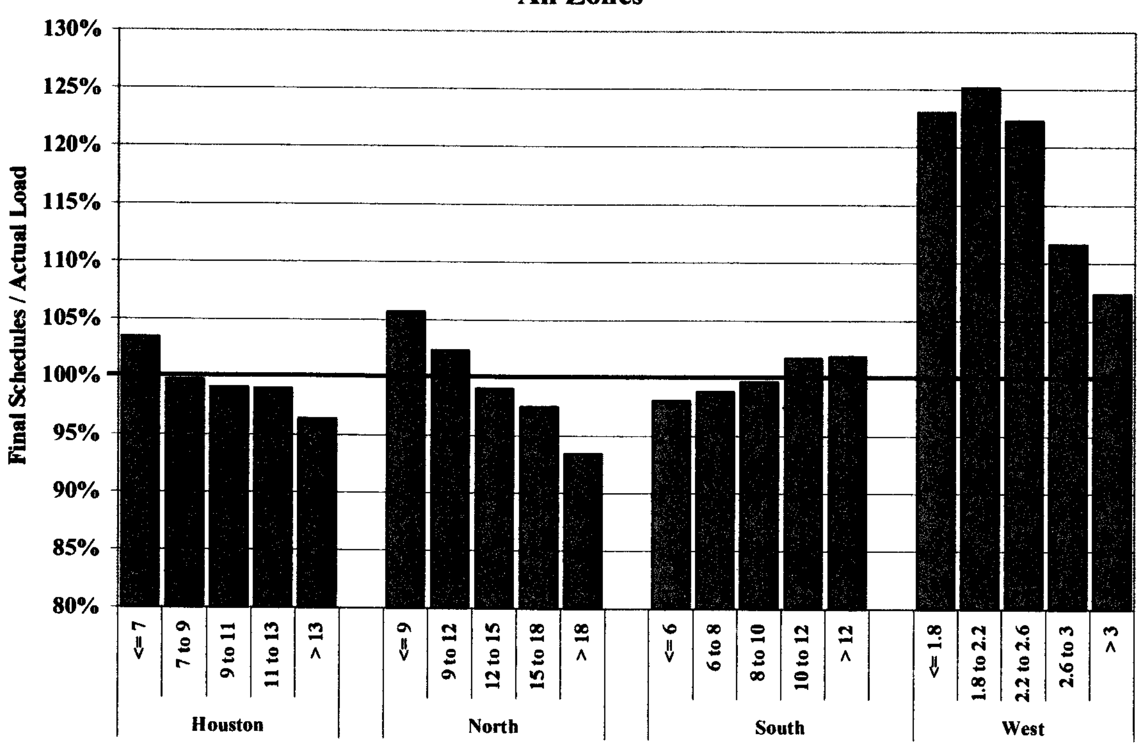


Figure 34 is a further analysis of final load schedules that shows the ratio of final load schedules to actual load evaluated at five different load levels for each of the ERCOT zones.

Figure 34 shows that:

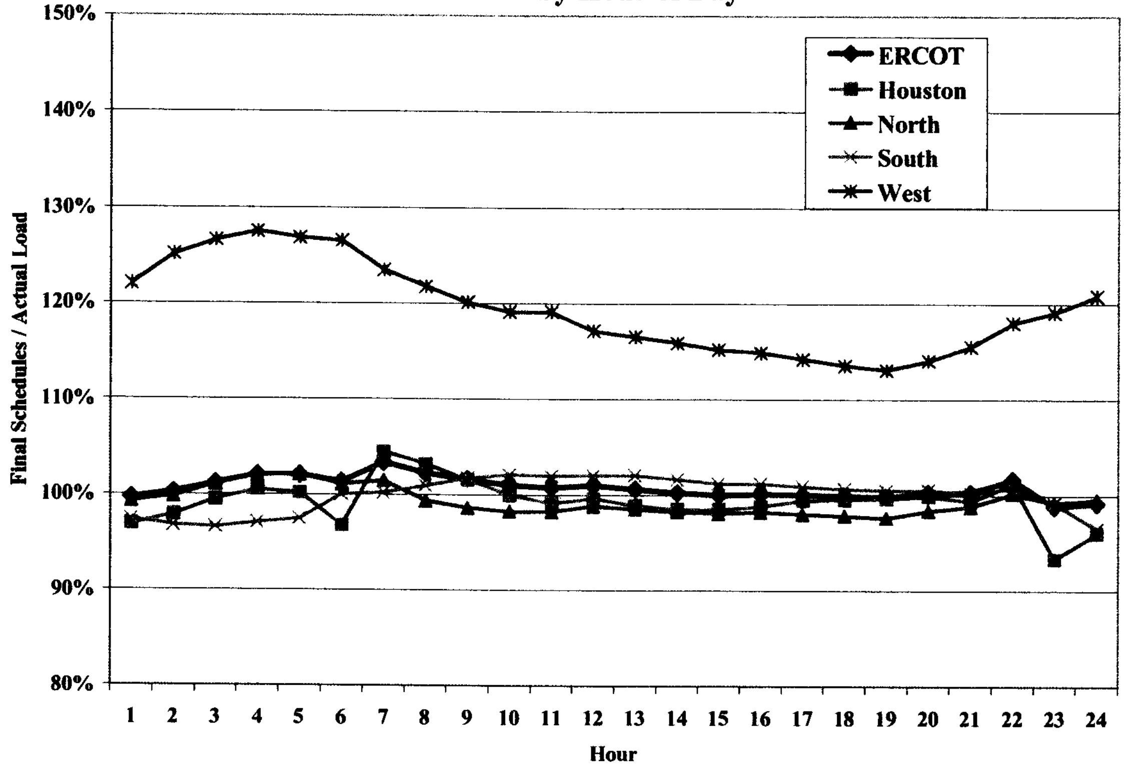
- The final schedule quantity decreases in three of the four zones as actual load increases. In contrast, the schedules in the South zone increase slightly as actual load increases.
- The West Zone is generally over-scheduled, although the ratios decline as load increases.
- Houston is under-scheduled at most load levels, but the level of under-scheduling is lower than in 2006. In 2006, the under-scheduling levels ranged from 4 percent at lower load levels up to 8 percent at high load levels. In 2007, the range is from 0.2 percent to 3.6 percent.

The result of these scheduling patterns is that the QSEs in Houston are net buyers of balancing energy to the extent that they do not offer generation in the balancing energy market to cover their deficits. In contrast, QSEs in the South Zone, to a lesser degree, are net sellers of balancing energy. Thus, the net importing zones seem to under-schedule while the net exporting zones over-schedule. 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. It 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. This poses no operational concerns and is a mechanism by which some suppliers may more fully utilize their portfolio.

To further analyze load scheduling, Figure 35 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 35: 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. At hour ending 7, the ERCOT-wide ratio increases to 103 percent. In the other hours, the ERCOT-wide ratio ranges between 99 and 102 percent. The higher ratio in the West zone is most likely explained by the increases in wind capacity in 2007 where the wind is scheduled as a price taker in the West zone, and by the trading of "seller's choice" bilateral contracts that often designate the West zone as the point of delivery and for which some of the transactions are scheduled as a price taker in the West 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 35 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.

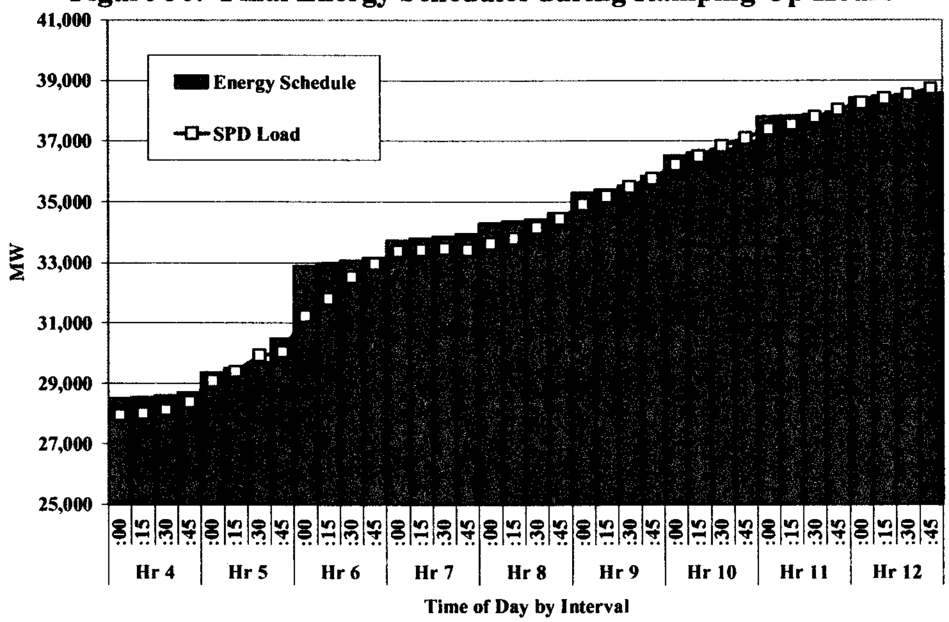
B. Balancing Energy Market Scheduling

In the previous section, we analyzed balancing energy prices 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 section, 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 36 shows average energy schedules and actual load for each interval from 4 AM to 1 PM during 2007.

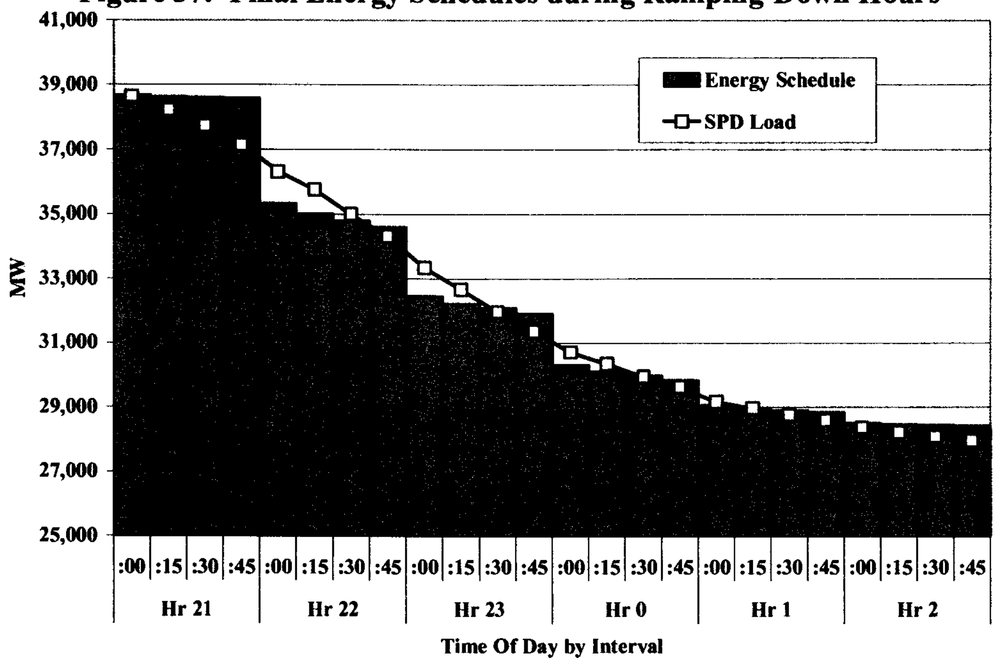
In general for ERCOT as a whole, energy schedules that are less than the actual load result in balancing energy purchases while energy schedules higher than actual load result in balancing energy sales. On average, load increases from approximately 28 GW to almost 39 GW in the nine hours shown in Figure 36. The average increase per 15-minute interval is approximately 330 MW, although the rate of increase is greatest from 5:45 AM to 7:00 AM and relatively flat from 7:00 AM to 8:30 AM. This "hump" in the 6 AM to 8 AM timeframe is due, primarily, to the fact that the daily peak occurs in the morning during certain times of year. However, a small hump persists around 6 AM throughout the year.

Figure 36: Final Energy Schedules during Ramping-Up Hours



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 large increases from the fourth interval to the first interval of the next hour. For instance, the average energy schedule increases by over 2.4 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 several hundred megawatts in the subsequent three intervals. The same scheduling patterns exist in the ramping-down hours. Figure 37 shows average energy schedules and load for each interval from 9 PM to 3 AM during 2007.

Figure 37: Final Energy Schedules during Ramping-Down Hours



On average, load drops from approximately 39 GW to less than 28 GW in the six hours shown in Figure 37. The average decrease per 15-minute interval is approximately 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 during the ramping-up hours, energy schedules change (decrease) in relatively large steps at the top of each hour. For instance, the average energy schedule drops nearly 3.3 GW from the last interval before 10 PM to the interval beginning at 10 PM.