POTOMAC ECONOMICS

2007 STATE OF THE MARKET REPORT
FOR THE
ERCOT WHOLESALE ELECTRICITY MARKETS

POTOMAC ECONOMICS, LTD.

Independent Market Monitor for the
ERCOT Wholesale Market

August 2008

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

This report reviews and evaluates the outcomes of the ERCOT wholesale electricity markets in 2007. It includes assessments of the incentives provided by the current market rules and procedures, and analyses of the conduct of market participants. We find improvements in a number of areas over the results in prior years that can be attributed to changes in the market rules or operation of the markets. Our analysis also indicates that the market performed competitively in 2007. However, the report generally confirms prior findings that the current market rules and procedures are resulting in systematic inefficiencies. This report also assesses the effectiveness of the scarcity pricing mechanism pursuant to the provisions of Public Utility Commission of Texas ("PUCT") Substantive Rule 25.505(g).

Many of these findings can be found in five previous reports we have issued regarding the ERCOT electricity markets.[1] These reports included a number of recommendations designed to improve the performance of the current ERCOT markets. Many of these recommendations were considered by ERCOT working groups and some were embodied in protocol revision requests ("PRRs"). Most of the remaining recommendations will be addressed by the introduction of a nodal market design.

The wholesale market should function more efficiently under the nodal market design by: providing better incentives to market participants, facilitating more efficient commitment and dispatch of generation, and improving ERCOT's operational control of the system. The congestion on all transmission paths and facilities will be managed through market-based mechanisms in the nodal market. In contrast, under the current zonal market design, most transmission congestion is resolved through non-transparent, non-market-based procedures.

Under the nodal market, unit-specific dispatch will allow ERCOT to more fully utilize the generating resources than the current market, which frequently exhibits shortage prices when the generating capacity is not fully utilized. Finally, the nodal market will produce price signals that

[1] "ERCOT State of the Market Report 2003", Potomac Economics, August 2004 (hereafter "2003 SOM Report"); "2004 Assessment of the Operation of the ERCOT Wholesale Electricity Markets", Potomac Economics, November 2004 (hereafter "Assessment of Operations"); "ERCOT State of the Market Report 2004", Potomac Economics, July 2005 (hereafter "2004 SOM Report"); "ERCOT State of the Market Report 2005", Potomac Economics, July 2006 (hereafter "2005 SOM Report"); and "ERCOT State of the Market Report 2006", Potomac Economics, August 2007 (hereafter "2006 SOM Report").

provide incentives to build new generation where it is most needed for managing congestion and maintaining reliability. In the long-term, these enhancements to overall market efficiency should translate into substantial savings for consumers.

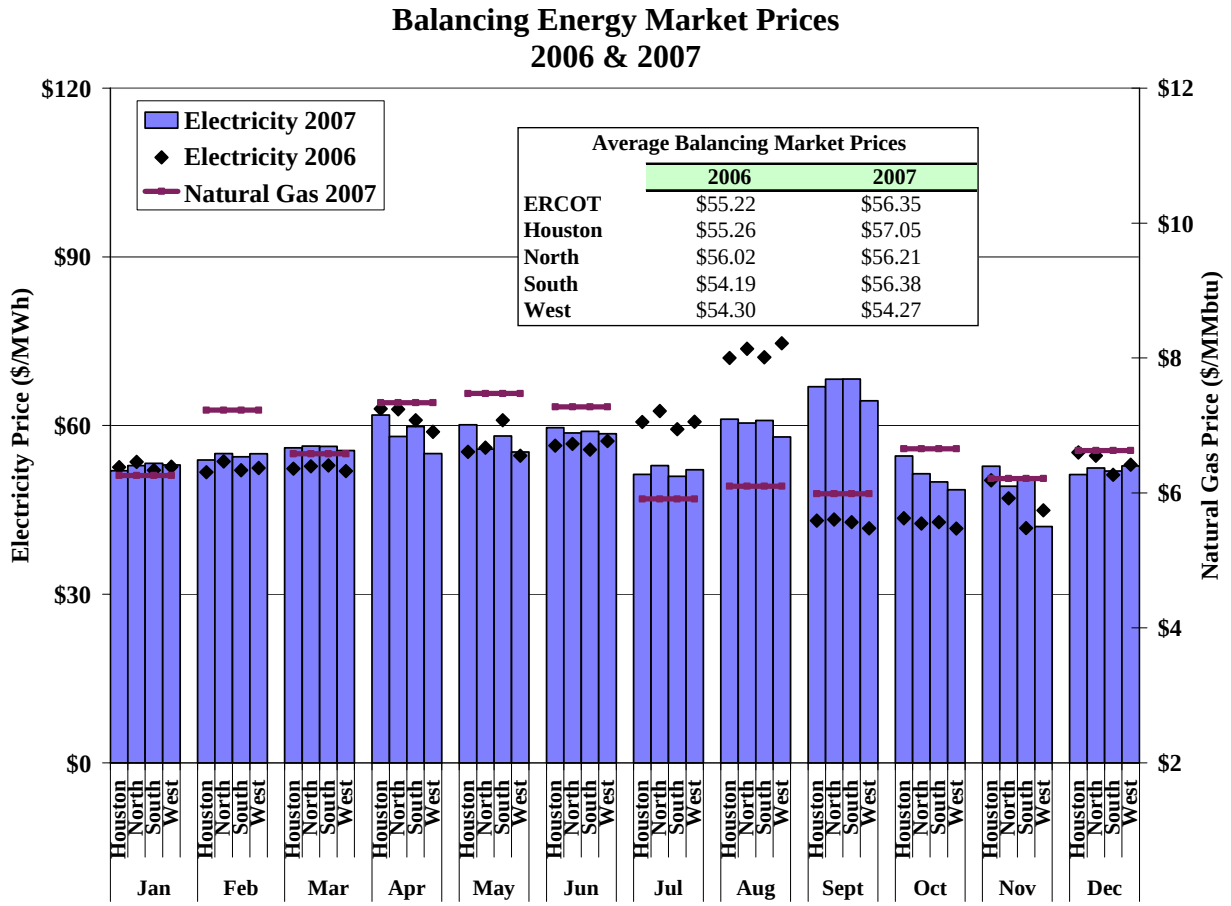
A. Review of Market Outcomes

1. Balancing Energy Prices

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

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

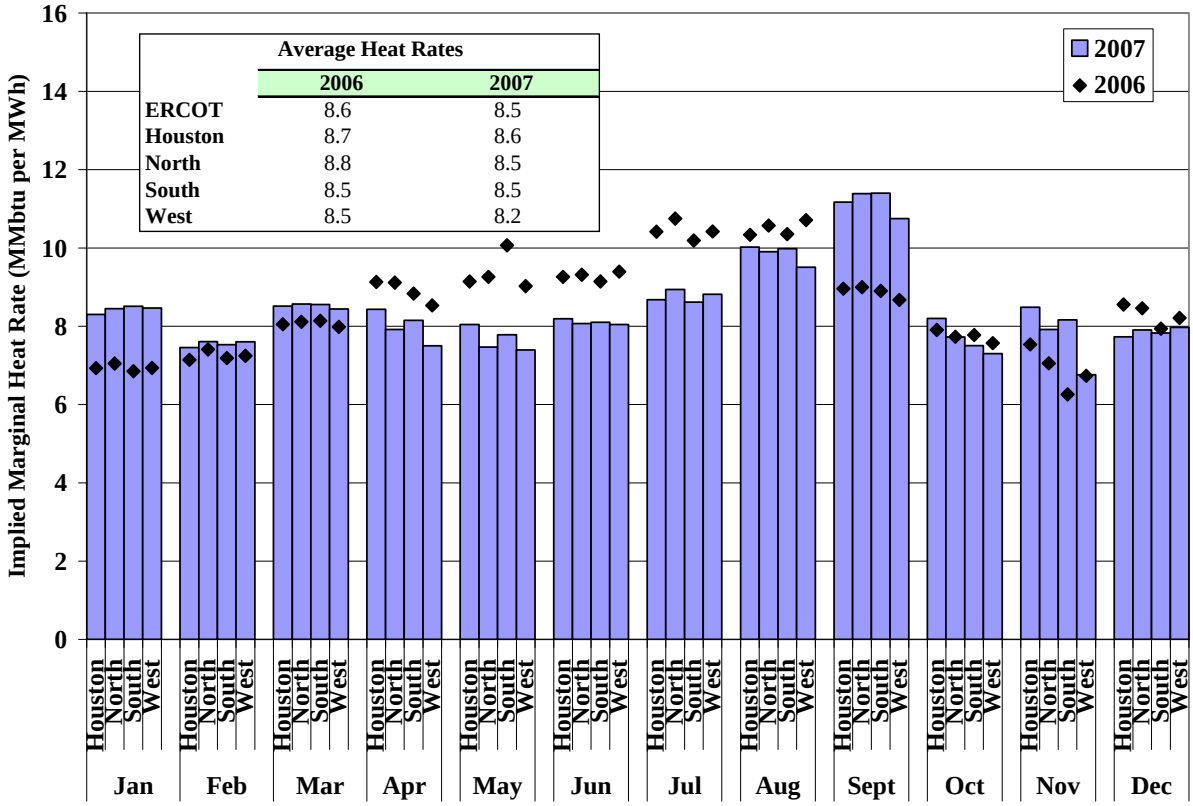
As shown in the following figure, balancing energy market prices were 2 percent higher in 2007 than in 2006, with September 2007 showing the largest increase from the same month in 2006. The average natural gas price in 2007 increased 4 percent over 2006 levels, with monthly changes ranging from a 25 percent increase in September ($4.81/MMBtu in September 2006 and $5.99/MMBtu in September 2007) to an 18 percent decrease in Janauary ($7.59/MMBtu in January 2006 and $6.26/MMBtu in January 2007). Natural gas is typically the marginal fuel in the ERCOT market. Hence, the movements in wholesale energy prices from 2006 to 2007 were largely a function of natural gas price levels.



Although natural gas price fluctuations are the dominant factor driving electricity prices in the ERCOT wholesale market, fuel prices alone do not explain all of the price changes. At least three other factors contributed to price changes in 2007. First, as discussed in Section III of this report, ERCOT peak demand and installed capacity were relatively flat in 2007, and energy production increased only slightly in 2007 compared to 2006. In contrast to prior years with increasing demand and decreasing supply, the static supply and demand characteristics from 2006 to 2007 contributed to comparable wholesale pricing outcomes over the course of these two years. Second, the balancing energy offer cap was raised to $1,500 on March 1, 2007, whereas the offer cap was $1,000 in 2006. The increased offer caps are intended to produce higher prices during shortage conditions. However, as discussed in Section I, this mechanism was not always effective in achieving this intended outcome. Finally, the overall competitive performance of the market exhibited continued improvement in 2007, which will tend to lower prices and is examined in detail in Section V. The following figure presents ERCOT balancing energy market

prices adjusted for natural gas price fluctuations to better highlight variations in electricity prices not related to fuel costs.

Monthly Average Implied Marginal Heat Rate
2006 & 2007



Adjusted for gas price influence, the above figure 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.[2] 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 higer levels pursuant to the provisions of the ERCOT Protocols.[3]

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

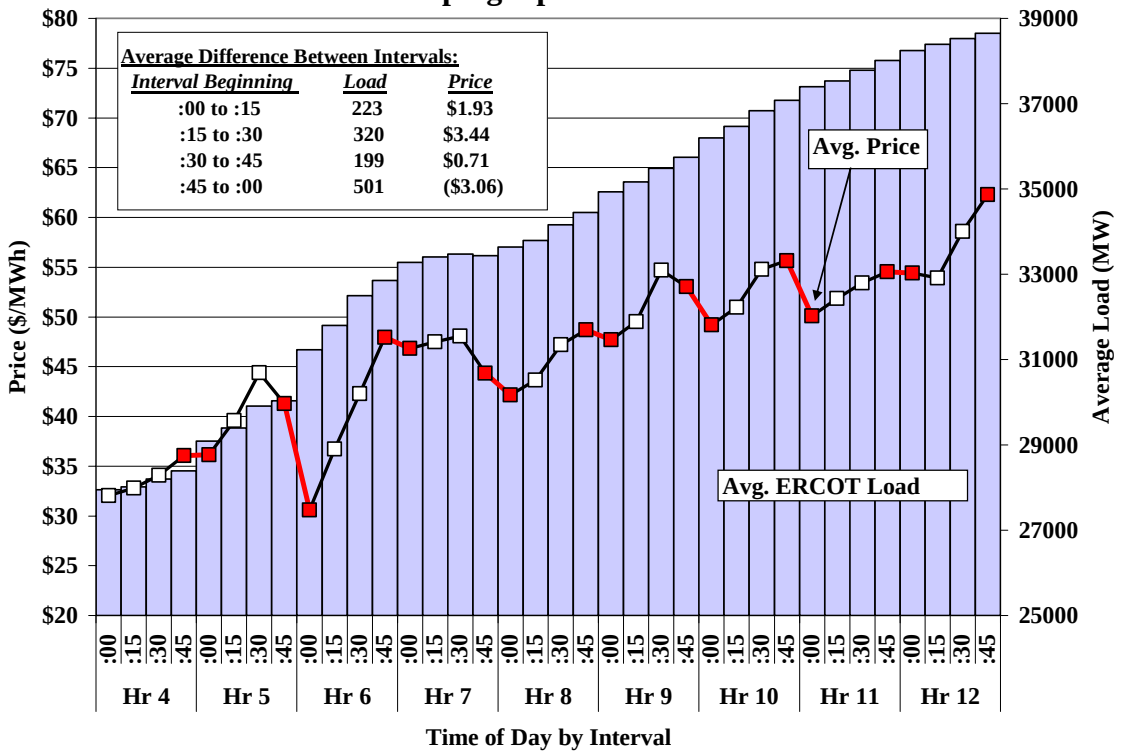
3 The price correction provisions were adopted in Protocol Revision Request No. 650. The appropriateness

The report evaluates two other aspects of the balancing energy prices: 1) the correlation of the balancing energy prices with forward electricity prices in Texas, and 2) the primary determinants of balancing energy prices. Natural market forces should push forward market prices to levels consistent with expectations of spot market prices. Forward prices were relatively consistent with balancing energy prices on the vast majority of days in 2007, although the introduction of the nodal market that includes an integrated day-ahead market should improve the convergence between day-ahead and real-time energy prices.

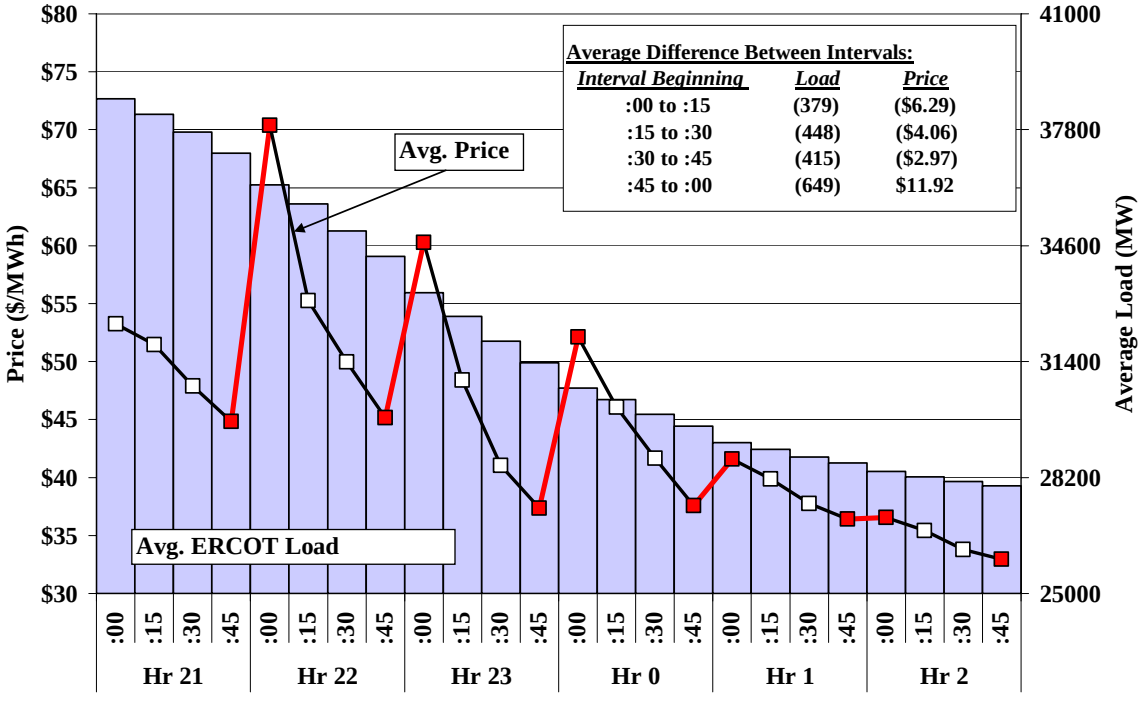
As discussed in prior reports, we continue to observe in 2007 a clear relationship between the net balancing energy deployments and the balancing energy prices. This is not expected in a well-functioning market. This relationship is partly due to the hourly scheduling patterns of most of the market participants. The energy schedules change by large amounts at the top of each hour while load increases and decreases smoothly over time. This creates extraordinary demands on the balancing energy market and erratic balancing energy prices, particularly in the morning when loads are increasing rapidly and in the evening when loads are decreasing rapidly.

of these price correction provisions was addressed in the 2006 SOM Report (2006 SOM Report, at 41-42).

Average Balancing Energy Prices and Load by Time of Day Ramping-Up Hours – 2007



Average Balancing Energy Prices and Load by Time of Day Ramping-Down Hours – 2007

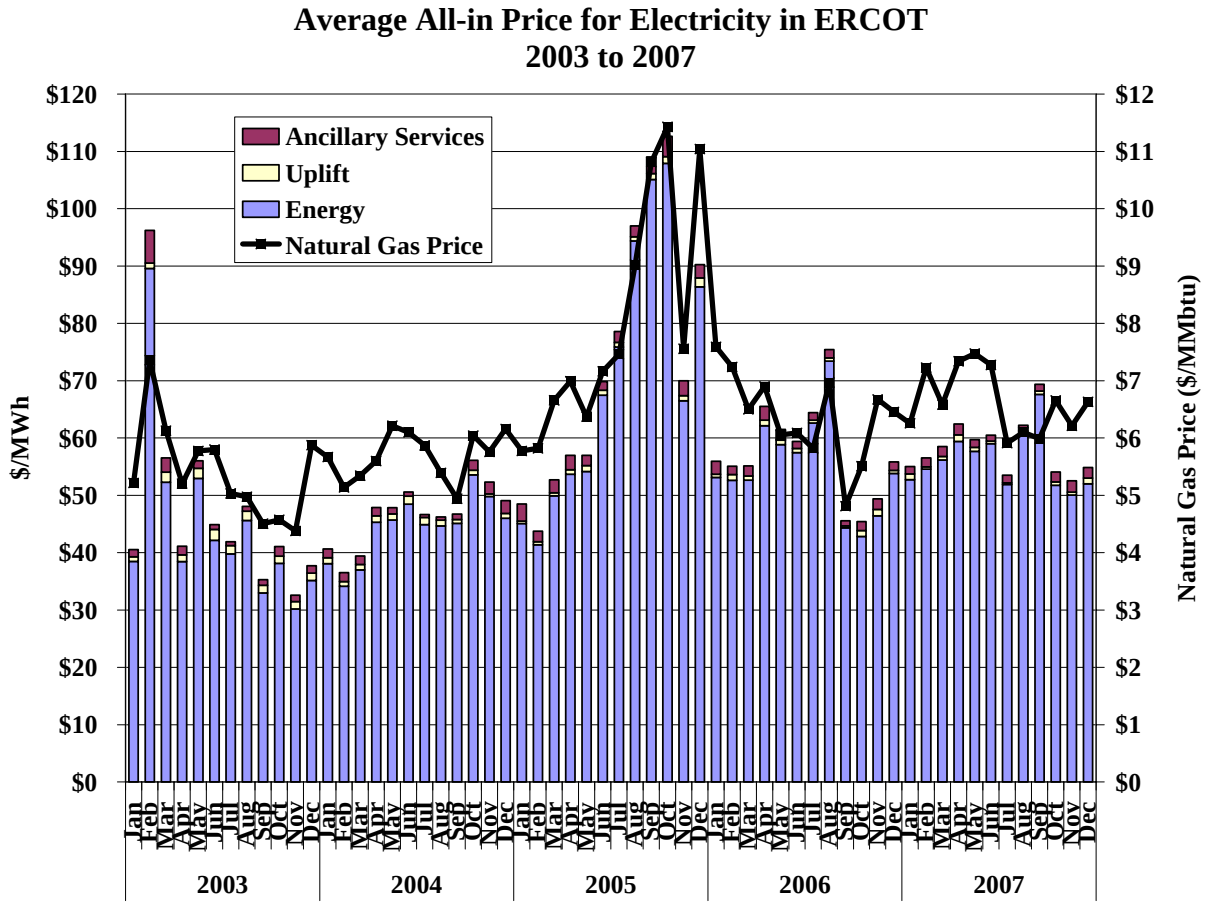


The previous two figures summarize these erratic price patterns by showing the balancing energy prices and actual load in each 15-minute interval during the morning "ramping-up" hours and evening "ramping-down" hours. These pricing patterns raise significant efficiency concerns regarding the operation of the balancing energy market. Moreover, this pattern has been consistently observed for several years and is likely to continue until changes are made to the market rules.[4] In prior reports, we have made several recommendations to address the issue under the current zonal design, although most have not been implemented because of the effort to timely implement the nodal market. The nodal market will provide for a comprehensive solution to the operational issues described in this and prior reports.

2. All-In Electricity Prices

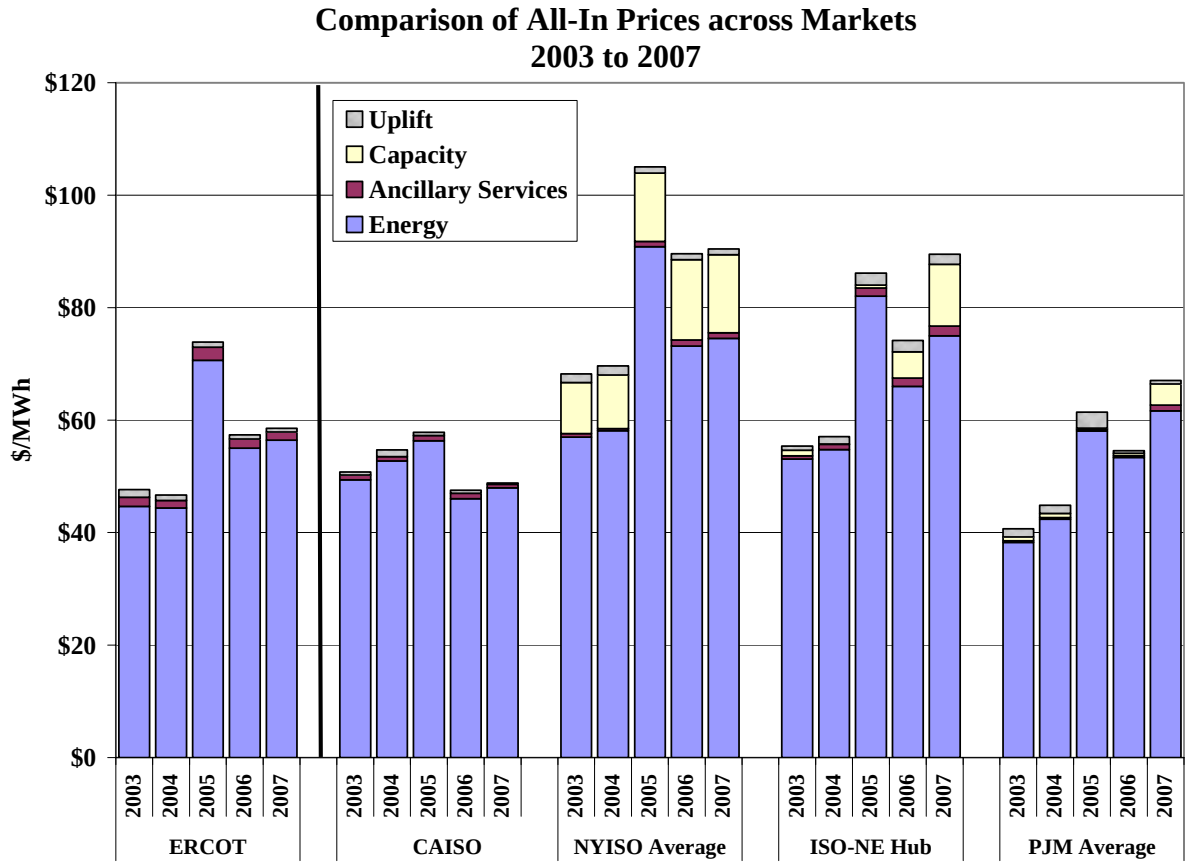
In addition to the costs of energy, loads incur costs associated with operating reserves, regulation, and uplift. The uplift costs include payments for out-of-merit capacity ("OOMC"), Replacement Reserve ("RPRS") out-of-merit energy ("OOME"), and reliability must run agreements ("RMR"), but excluding administrative charges such as the ERCOT fee. These costs, regardless of the location of the congestion, are borne equally by all loads within ERCOT. We calculated an average all-in price of electricity that includes balancing energy costs, ancillary services costs, and uplift costs. The monthly average all-in energy prices for the past four years are shown in the figure below along with a natural gas price trend.

[4] See 2003 SOM Report, Assessment of Operations, 2004 SOM Report, 2005 SOM Report and 2006 SOM Report.



The figure indicates that natural gas prices were the primary driver of the trends in electricity prices from 2003 to 2007. Natural gas prices increased in 2005 by an average of more than 41 percent from 2004 levels while the all-in price for electricity increased by 63 percent. In 2006, the natural gas price dropped by an average of 20 percent from 2005 levels and the all-in price for electricity decreased by 23 percent. In 2007, the natural gas price increase by an average of 4 percent from 2006 levels and the all-in price for electricity increased by 0.5 percent.

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



Wholesale electricity markets in the U.S. experienced substantial increases in energy prices from 2002 to 2003 and from 2004 to 2005 due to increased fuel costs. In 2006, energy prices in the U.S. dropped in every region due to decreased fuel costs. In 2007, the all-in prices increased in all the above five regions, with relatively small increases in ERCOT, California and New York, and more significant increases in New England and PJM.

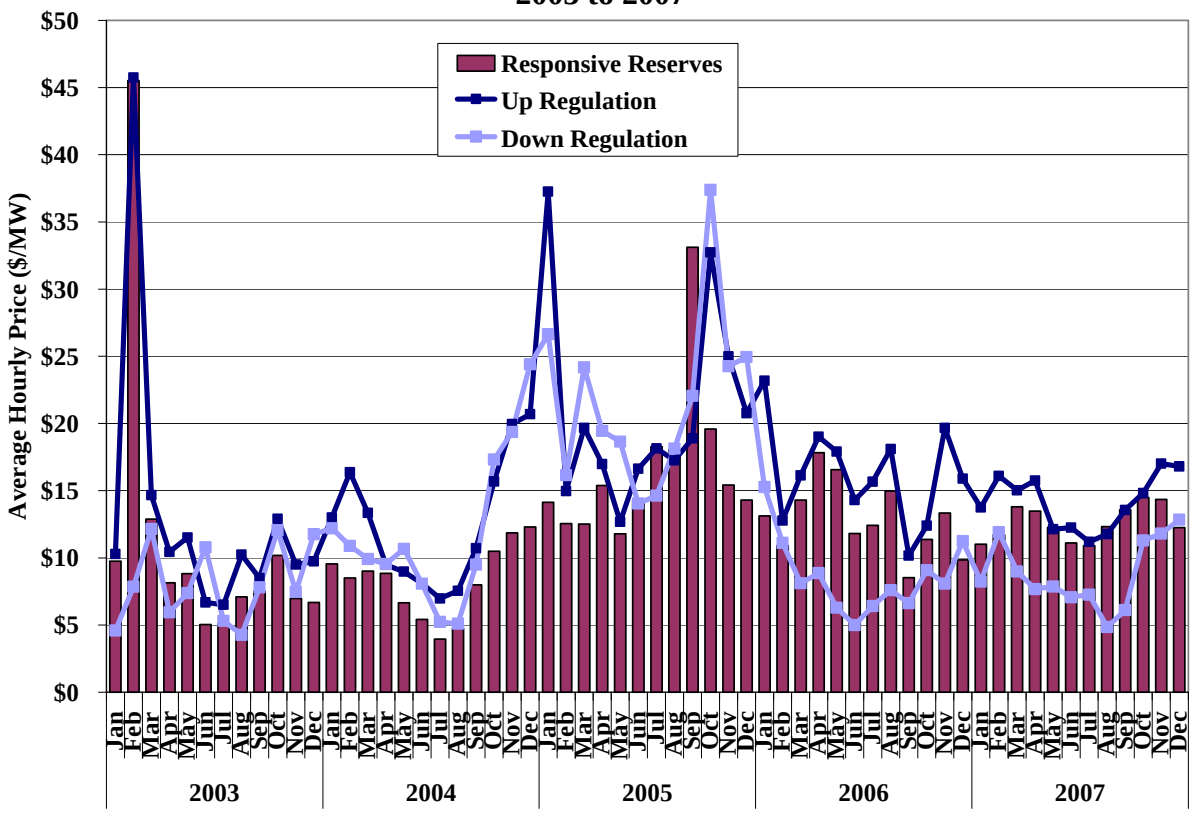
3. Ancillary Services Markets

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.

Ancillary services prices were comparable in 2006 and 2007, with both years showing modest increases over the levels prevailing in 2003. This is consistent with long-term trends in natural gas and electricity prices, and significantly below the price levels experienced in 2005. 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 responsive reserves and regulation can incur opportunity costs when they reduce the output from economic units to make the capability available to provide these services. The following figure shows the monthly average prices for regulation and responsive reserve services from 2003 to 2006.

Monthly Average Ancillary Service Prices
2003 to 2007



Although ancillary services prices have generally risen over the last several years, the impact has been partly mitigated by reductions in the required quantities of regulation. In 2002, ERCOT required approximately 3,000 MW of combined up and down regulation. By 2007, the requirement was reduced to an average of 1,800 MW during ramping hours and 1,420 MW during non-ramping hours. This has *directly* reduced regulation costs by reducing the overall quantity scheduled, either through bilateral arrangements or through the day-ahead auction. This has also *indirectly* reduced regulation costs by reducing the clearing prices of regulation that would have prevailed under higher demand levels for regulation. 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 quantities of regulation procured in 2007.

In this report, we compare the amounts of capacity scheduled to provide operating reserves to the quantities of capacity that are actually available in real time. In general, we find that the capacity available to provide reserves in real time far exceeds the quantities scheduled to meet the operating reserves requirements. This highlights issues relating to the efficiency of the ERCOT markets, which are expected to improve with the implementation of the nodal market.

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

4. Net Revenue Analysis

The next analysis of the outcomes in the ERCOT markets in 2007 is the analysis of "net revenue". Net revenue is defined as the total revenue that can be earned by a new generating unit less its variable production costs. It represents the revenue that is available to recover a unit's fixed and capital costs. Hence, this metric shows the economic signals provided by the market for investors to build new generation or for existing owners to retire generation. In long-run equilibrium, the markets should provide sufficient net revenue to allow an investor to break-even on an investment in a new generating unit, including a return of and on the investment.

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

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

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

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

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

The report estimates the net revenue that would have been received in 2005 to 2007 for four types of units: a natural gas combined-cycle generator, a simple-cycle gas turbine, a coal-fired steam turbine with scrubbers, and a nuclear unit. Net revenue was insufficient to support new entry for gas-fired units in 2007, although the net revenue for gas-fired units in 2007 remained significantly higher than years prior to 2005. As in 2005 and 2006, net revenue for coal and nuclear units remained above the levels required to support new entry. 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 Section I.D.
- 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, the Modified Competitive Solution Method – a mechanism that, per PUCT rules, required *ex post* reductions to the clearing price when all available energy was exhausted – was eliminated by the new rules.

5. Effectiveness of the Scarcity Pricing Mechanism in 2007

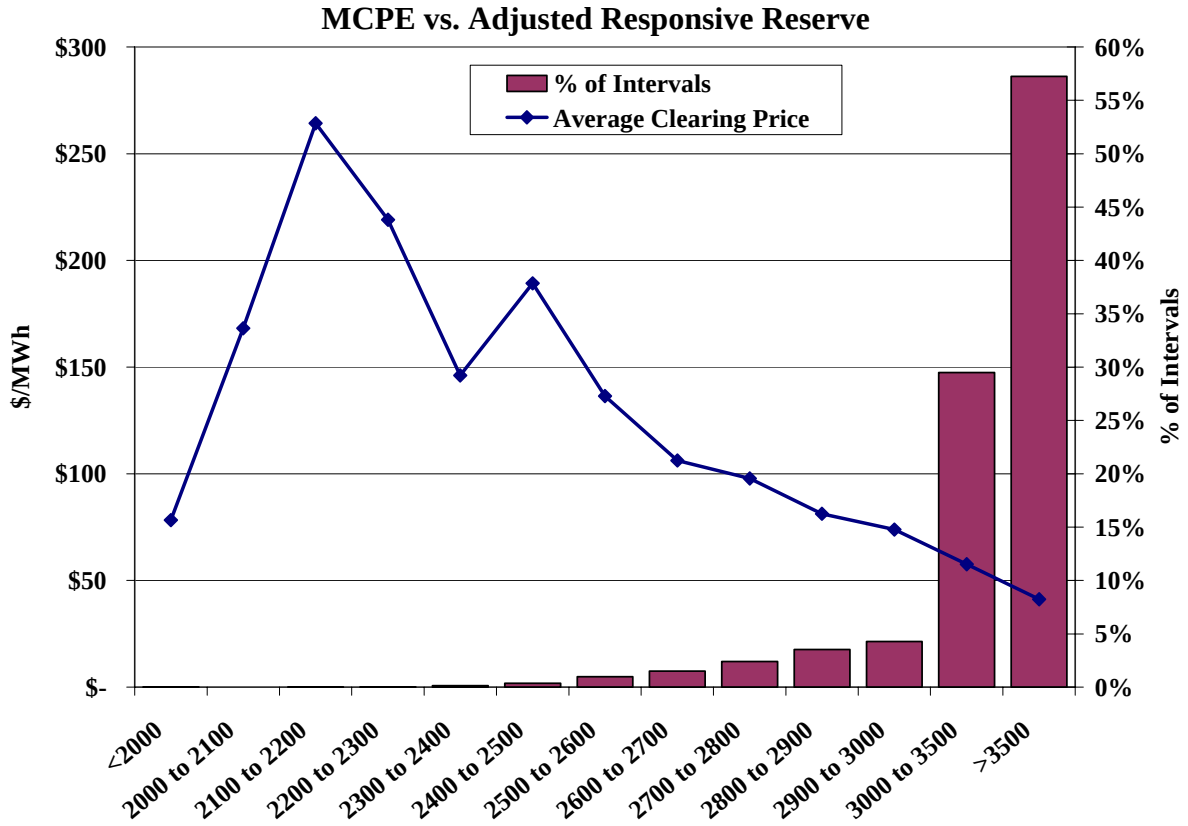
Unlike markets with a long-term capacity market where fixed capacity payments are made to resources across the entire year regardless of the relationship of supply and demand, the objective of the energy-only market design is to allow prices to rise significantly higher during legitimate shortage conditions (*i.e.*, when the supply of resources is insufficient to simultaneously meet both energy and operating reserve requirements) 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.

The Scarcity Pricing Mechanism ("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. Consistent with the results of the net revenue analysis, 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 ("OOM") 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
- 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.

The following figure illustrates the relationship between the balancing energy price and the amount of adjusted responsive reserve ("ARR"), which is a measure of the market operating reserve margin or shortage condition. ERCOT begins taking short-supply actions when ARR decreases below 2,500 MW, and declares an alert when ARR decreases below 2,300 MW. As ARR decreases to toward these levels and below, a gradual and ultimately very sharp increase in price should result if the scarcity pricing mechanism is effective. However, as can be seen from the following figure, frequent OOM deployments had the effect of depressing the price under these shortage conditions.

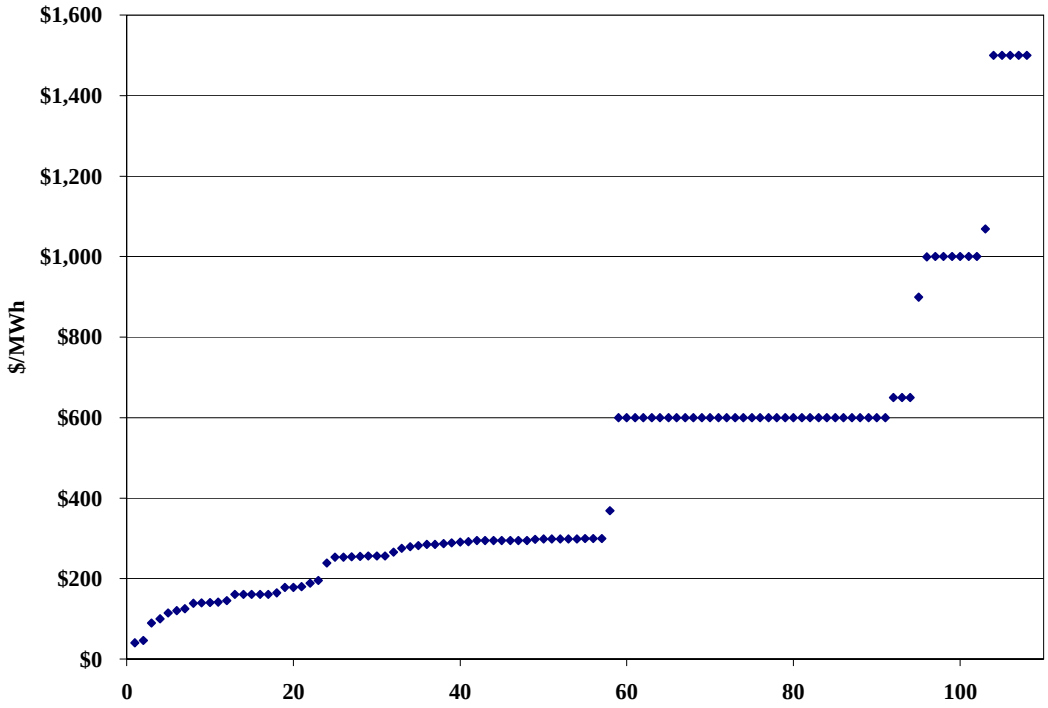


As shown in the figure above, the average price rose in 2007 as ARR dropped from 3,500 to 2,500 MW. However, once ARR reached 2,500 MW, the average price dropped, which can be attributed to the initial OOM actions taken by ERCOT when ARR reaches 2,500. Prices resumed their increase for ARR levels between 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.

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. The following figure shows the balancing market clearing prices during the 108 15-minute intervals in 2007 when all available balancing energy was exhausted.

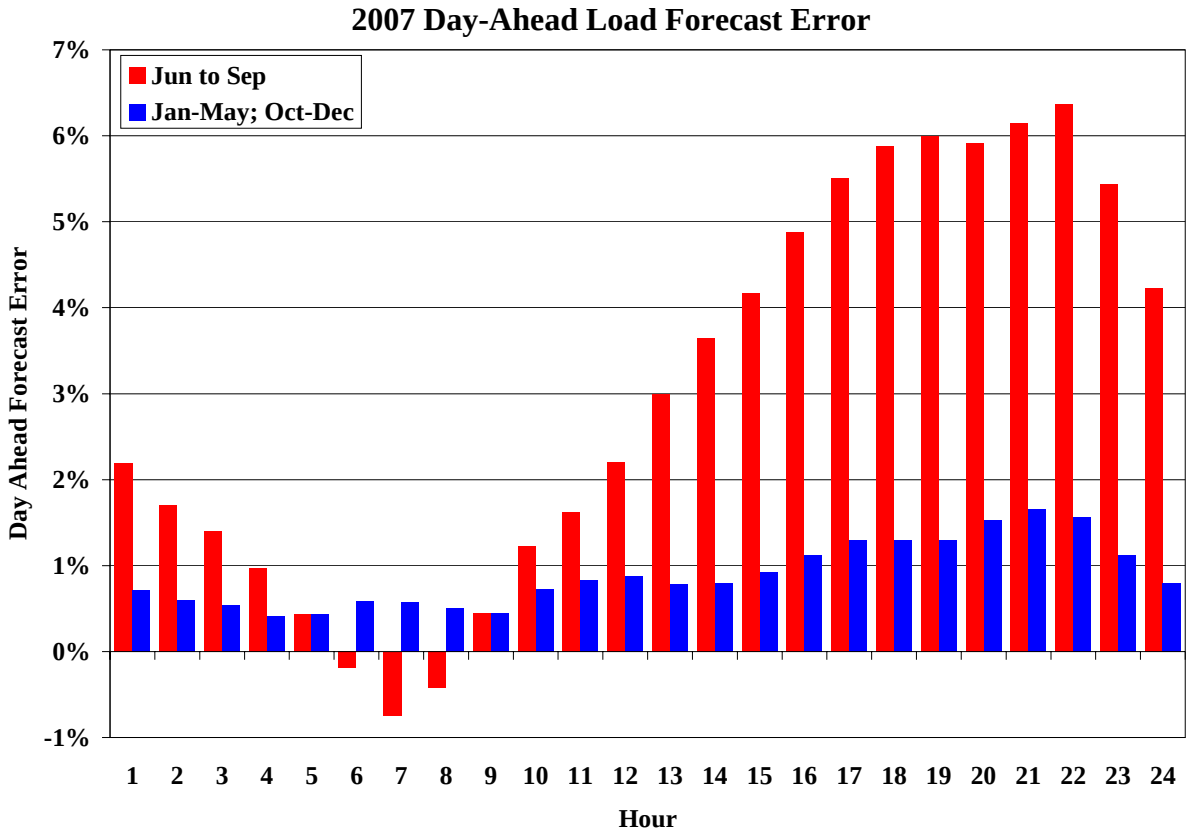
Balancing Energy Market Prices During Shortage Intervals



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

Along with the factors above, the existence of a strong and persistent positive bias in the day-ahead load forecast in 2007 has the effect of producing an inefficient over-commitment of resources and depressing real-time prices relative to a more optimal unit commitment. The

following figure shows the ERCOT day-ahead load forecast error by hour in 2007, with the summer and non-summer months presented separately.



Because of the inefficiencies associated with a persistently high day-ahead load forecast, we recommend that ERCOT review the causes of the positive bias in its day-ahead load forecast, and explore potential changes to its reserve procurement policies and its day-ahead and supplemental unit commitment procedures.

B. Balancing Energy Offers and Schedules

QSEs play an important role in the current ERCOT markets. QSEs must submit balanced schedules so that the quantity of generation scheduled matches the quantity of load scheduled prior to real-time. However, there is no requirement for the scheduled load to match the forecast of real-time load. When actual real-time load exceeds the energy scheduled prior to real-time, the remaining load is served by energy purchased in the balancing energy market. Conversely, when scheduled energy exceeds actual real-time load, load serving entities sell their excess to the balancing energy market. QSEs submit balancing energy offers to increase or decrease their

energy output from the scheduled energy level. The balancing-up offers correspond to the unscheduled output from the QSEs' online and quick-start resources.

In addition to the forward schedules and offers, QSEs submit resource plans that provide a non-binding indication of the generating resources that the QSE will have online and producing energy to satisfy its energy schedule and ancillary services obligations. The report evaluates the effects on the balancing energy market of the QSEs' schedules, offers, and resource plans.

1. Hourly Schedule Changes

One of the most significant issues affecting the ERCOT balancing energy market is the changes in energy schedules that occur from hour to hour, particularly in hours when loads are changing rapidly (*i.e.*, "ramping") in the morning and evening. The report shows that:

- In these ramping hours, the loads are generally moving approximately 300 to 500 MW each 15-minute interval.
- Although QSE's can modify their schedules each interval, most only change their schedules hourly, resulting in schedule changes averaging 1,000 to 4,000 MW in these hours (and sometimes significantly larger).
- The inconsistency between the changes in schedules and actual load in these hours places an enormous burden on the balancing energy market, resulting in the erratic pricing patterns shown above.

Several changes have been recommended in prior reports to address this issue, most of which will not be implemented because of the transition to the nodal market. The issues that these recommendations were designed to address should be resolved by the implementation of unit-specific dispatch under the nodal market design.

2. Portfolio Offers in the Balancing Energy Market

The report evaluates the portfolio offers submitted by QSEs in the balancing energy market, including both the quantity and ramp rate of the offers (the amount of the offer that can be deployed in any single 15-minute interval).

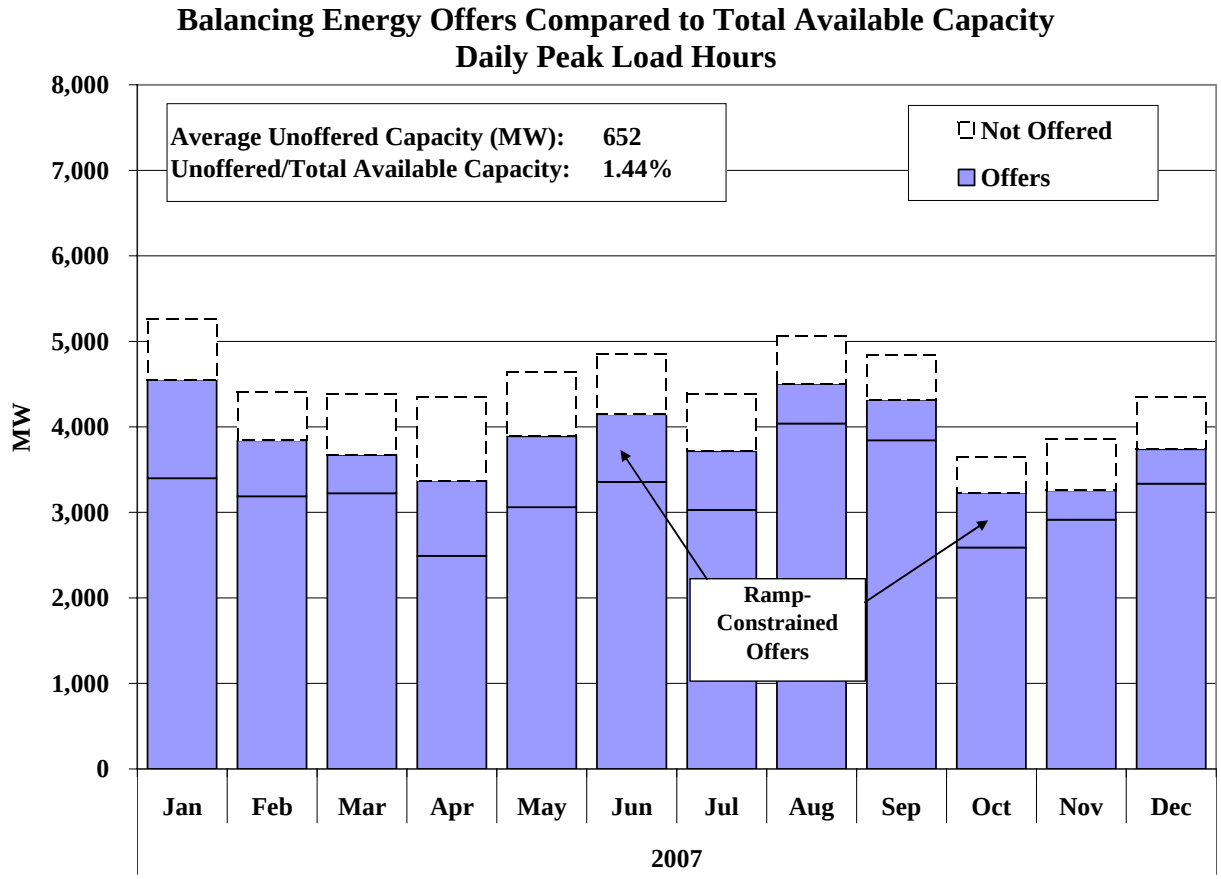
The volatility of the balancing energy prices in each interval is primarily related to the balancing energy deployments. However, this volatility can be exacerbated when the portfolio ramp rates are binding. Portfolio ramp rates are constraints QSEs submit with their balancing energy offers to limit the quantity of balancing up or balancing down 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 unnecessary price spikes (as well as large negative prices). There are three aspects of the current market design that inhibit QSEs from fully utilizing the ramp capability of their portfolio. These are: (1) portfolio ramp rates; (2) portfolio level rather than unit level dispatch; and (3) lack of coordination between energy schedules and ramping. These issues were discussed in detail in the 2005 SOM Report. The operational implications associated with these issues continued in 2007 and will likely continue until the current zonal market design is replaced. However, each of these issues will be significantly ameliorated or eliminated with the implementation of the nodal market.

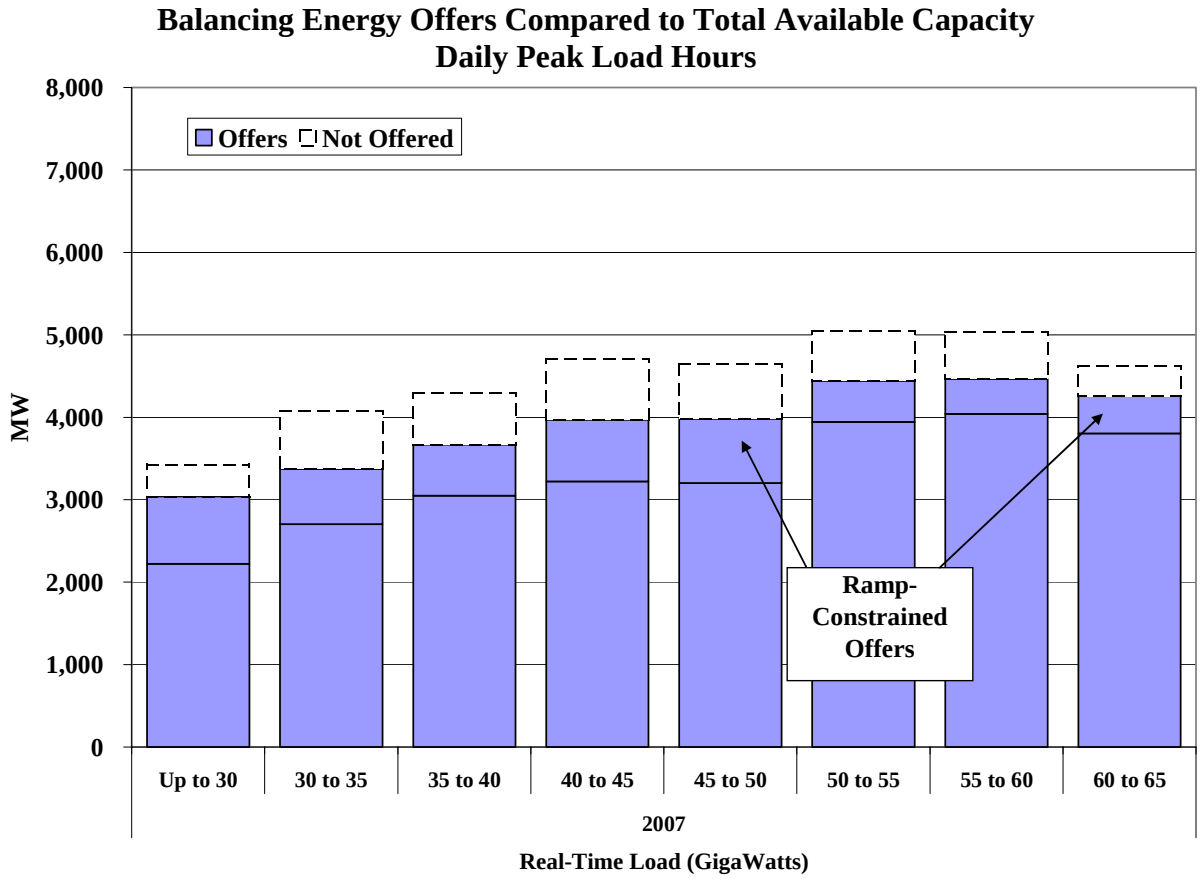
3. Balancing Energy Market Offer Patterns

We also evaluate balancing energy offer patterns by analyzing the rate at which capacity is offered. The figure below shows the average amount of capacity offered to supply balancing up service relative to all available capacity.



The figure above shows only slight variation in 2007 over time in quantities of energy available and offered to the balancing energy market. As discussed in more detail in the 2005 and 2006 ERCOT SOM Reports, there are various structural impediments associated with the zonal market model that serve to explain the residual quantity of un-offered capacity that persists from month-to-month.

Un-offered energy can raise competitive concerns to the extent that it reflects withholding by a dominant supplier that is attempting to exercise market power. To investigate whether this has occurred, the figure below shows the same data as the previous figure, but arranged by load level for daily peak hours in 2007. Because prices are most sensitive to withholding under the tight conditions that occur when load is relatively high, increases in the un-offered capacity at high load levels would raise competitive concerns.



This figure indicates that in 2007, the average amount of capacity available to the balancing market increased gradually up to 60 GW of load and then declined at higher levels. The decline in balancing energy available at higher load levels is associated with the fact that scheduled generation increases at higher load levels, thereby leaving less residual capacity available to be offered as balancing energy. As indicated in the figure, the quantity of un-offered capacity does not change significantly as load levels increase.

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

C. Demand and Resource Adequacy

1. Installed Capacity and Peak Demand

Since electricity cannot be stored, the electricity market must ensure that generation matches load on a continuous basis. Thus, one critical issue for a wholesale electricity market is whether sufficient supplies exist to satisfy demand under peak conditions. In 2007, the load served by ERCOT reached a peak of over 62 GW. The total load level increased about 0.7 percent in 2007 from 2006. Changes in the peak demand levels are very important because they are a key determinant of the probability and frequency of shortage conditions, although daily unit commitment practices, load uncertainty and unexpected resource outages are also contributing factors.

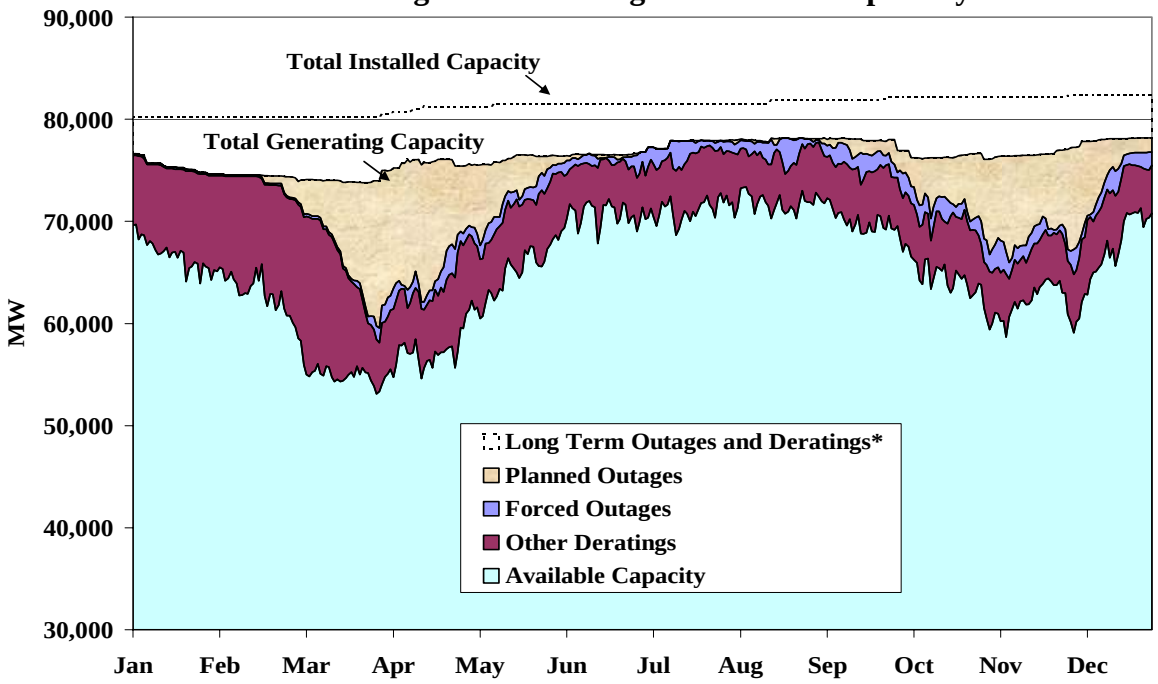
More broadly, peak demand levels and the capability of the transmission network are the primary factors that determine whether the existing generating resources are adequate to maintain reliability. The report provides an accounting of the current ERCOT generating capacity, which is dominated by natural gas-fired resources. These resources account for 70 percent of generation capacity in ERCOT as a whole, and 85 percent in the Houston Zone.

ERCOT has more than 80 GW of installed capacity. This includes import capability, resources that can be switched to the SPP, and Loads acting as Resources ("LaaRs"). However, significant amounts of this are not kept constantly in service. ERCOT estimates that about 5 GW was mothballed during 2007 and a large amount of capacity is used to satisfy cogeneration demands rather than to produce electricity. Furthermore, ambient temperature restrictions increase during the summer months when demand is highest, leading to substantial deratings. Although ERCOT had sufficient capacity to meet load and ancillary services needs during the 2007 peak, it is important to consider that electricity demand will continue to grow and that a significant number of generating units in Texas will soon reach or are already exceeding their expected lifetimes. Without significant capacity additions, these factors may cause the resource margins in ERCOT to diminish rapidly over the next three to five years. This reinforces the importance of ensuring that efficient economic signals are provided by the ERCOT market.

2. Generator Outages and Commitments

Despite adequate installed capacity, resource adequacy must be evaluated in light of the resources that are actually available on a daily basis to satisfy the energy and operating reserve requirements in ERCOT. A substantial portion of the installed capability is frequently unavailable due to generator deratings.

Short and Long-Term Deratings of Installed Capability



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

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

A derating is the difference between the installed capability of a generating resource and its maximum capability (or "rating") in a given hour. Generators can be fully derated (rating equals 0) due to a forced or planned outage. However, it is very common for a generator to be partially derated (*e.g.*, by 5 to 10 percent) because the resource cannot achieve its installed capability level due to technical or environmental factors (*e.g.*, ambient temperature conditions). The previous figure shows the daily available and derated capability of generation in ERCOT.

The figure shows that long-term outages and other deratings fluctuated between 7 and 22 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, as expected.

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

The report finds that the excess on-line capacity during daily peak hours on weekdays averaged 3,020 MW in 2007, which is approximately 8 percent of the average load in ERCOT. The overall trend in excess on-line capacity also indicates a movement toward more efficient unit commitment across the ERCOT market; however, the current market structure is still based primarily upon a decentralized unit commitment process whereby each participant makes independent generator commitment decisions that are not likely to be optimal. Further contributing to the suboptimal results of the current unit commitment process is that the decentralized unit commitment is reported to ERCOT through non-binding resource plans that form the basis for ERCOT's day-ahead planning decisions. However, these non-binding plans can be modified by market participants after ERCOT's day ahead planning process has concluded. Consequently, ERCOT frequently takes additional actions to ensure reliability that may be more costly and less efficient. Hence, the introduction of a day-ahead energy market with centralized Security Constrained Unit Commitment ("SCUC") that is financially binding

under the nodal market design promises substantial efficiency improvements in the commitment of generating resources.

3. Load Participation in the ERCOT Markets

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

During 2007, 2,050 MW of capability were qualified as LaaRs. The amount of responsive reserves provided by LaaRs has gradually increased from about 900 MW at the beginning of 2004 and stood at 1,985 MW at the end of 2005. In 2007, LaaRs were permitted to supply up to 1,150 MW of the responsive reserves requirement. Although the participants with LaaR resources are qualified to provide non-spinning reserves and up balancing energy in real-time, LaaR participation in the non-spinning reserve and, balancing energy market was negligible in 2007.[5] This is not surprising because the value of curtailed load tends to be relatively high, and providing responsive reserves offers substantial revenue with very little probability of being deployed. In contrast, resources providing non-spinning reserves are 70 times more likely to be deployed. Hence, most LaaRs will have a strong preference for providing responsive reserves over non-spinning reserves or balancing energy.

The clearing price for responsive reserves provided by LaaRs is set by the marginal generator, although the quantity of LaaRs willing to supply responsive reserves at the clearing price typically exceeds the demand (*i.e.*, 1,150 MW). The design of this market encourages inefficient behavior by QSEs that want to sell responsive reserves from their demand resources and results in inefficient prices in the responsive reserve market.

To improve the efficiency of responsive reserves pricing and incentives for suppliers, we recommend that ERCOT set separate prices for the two types of responsive reserves. The best

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

way to accomplish this would be by having two responsive reserves constraints in the ancillary services auction: (i) that the responsive reserves procurement (including bilateral schedules) be greater than or equal to 2,300 MW and (ii) that the responsive reserves procurement from LaaRs (including bilateral schedules) be less than or equal to 1,150 MW. The clearing price paid to generators would be equal to the shadow price of the first constraint only, while the clearing price paid to LaaRs would be equal to the shadow price of the first constraint minus the shadow price of the second constraint.

ERCOT stakeholders considered this change in 2006 and, due to resource constraints, decided not to implement it in the current market and instead drafted a protocol revision to implement it in the nodal market. However, this protocol revision failed to receive the necessary two-thirds vote at the ERCOT Technical Advisory Committee in 2007; thus, there is currently no plan to implement any of the changes described above for the RRS market. As previously discussed, the current mechanism for selecting providers and determining clearing prices for responsive reserves is inefficient and leads to excessive reliability costs for consumers. Therefore, we recommend that these changes be reconsidered for implementation in the nodal market design.

D. Transmission and Congestion

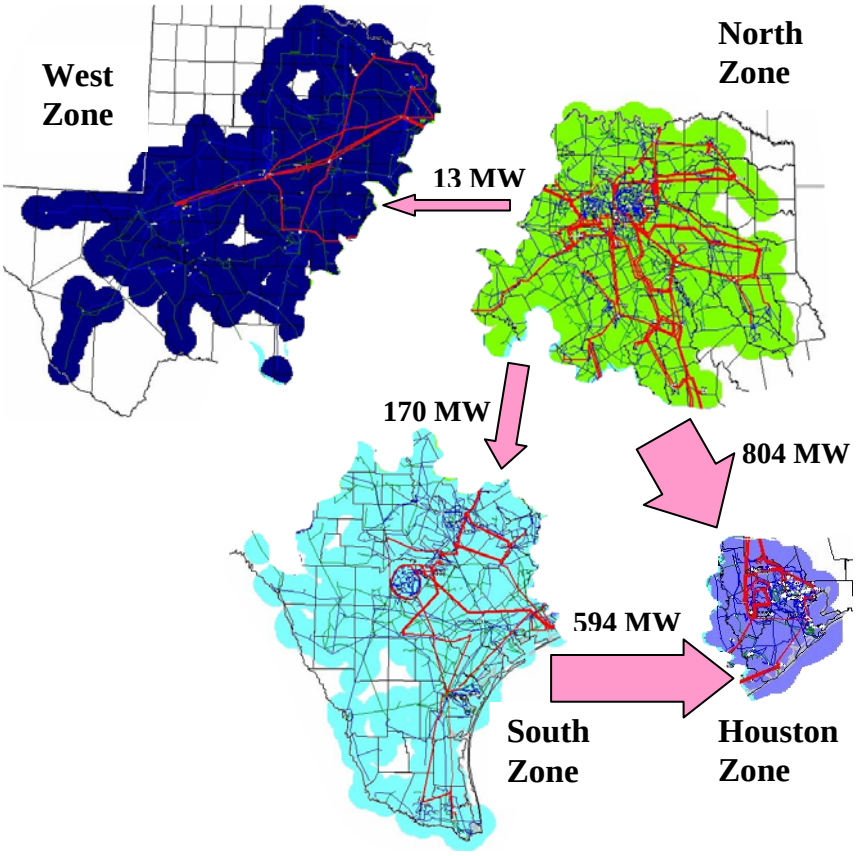
One of the most important functions of any electricity market is to manage the flows of power over the transmission network, limiting additional power flows over transmission facilities when they reach their operating limits. In ERCOT, constraints on the transmission network are managed in two ways.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

1. Electricity Flows between Zones and Interzonal Congestion

The balancing energy market uses the Scheduling, Pricing, and Dispatch ("SPD") software which dispatches energy in each zone to serve load and manage congestion between zones. The

SPD model embodies the market rules and requirements documented in the ERCOT protocols. To manage interzonal congestion, SPD uses a simplified network model with four zone-based locations and five transmission interfaces. The transmission interfaces are referred to as Commercially Significant Constraints ("CSCs"). The following figure shows the average flows modeled in SPD during 2007 over each of these CSCs.

Average Modeled Flows on Commercially Significant Constraints
2007



Note: In the figure above, CSC flows are averaged taking the direction into account. So one arrow shows the average flow for the North-to-West CSC was 13 MW, which is equivalent to saying that the average for the West-to-North CSC was *negative* 13 MW.

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

- The simplifying assumptions made in the SPD model can result in modeled flows that are considerably different from actual flows.
- A considerable quantity of flows between zones occurs over transmission facilities that are not defined as part of a CSC. When these flows cause congestion, it is beneficial to create a new CSC to better manage congestion over that path.

- Based on modeled flows, Houston is a significant importer while the North Zone and the South Zone export significant amounts of power.
- The physical flow vs. physical limit analysis reveals that the physical limits sometimes differ significantly from the actual flows.

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

The levels of interzonal congestion increased considerably to $114 million in 2007, which reflects an increase of $45 million from 2006. This increase was the result of more frequent congestion on the North-to-Houston, North-to-West, and West-to-North CSCs, as well as increased shadow price caps.[6]

To account for the fact that the modeled flows can vary substantially from the actual physical flows (due to the simplifying assumptions in the model), ERCOT operators must adjust the modeled limits for the CSC interfaces to ensure that the physical flows do not exceed the physical limits. This process results in highly variable limits in the market model for the CSC interfaces.

2. Transmission Congestion Rights and Payments

Participants in Texas can hedge against congestion in the balancing energy market by acquiring Transmission Congestion Rights ("TCRs") between zones which entitle the holder to payments equal to the difference in zonal balancing energy prices. Because the modeled limits for the CSC interfaces vary substantially, the quantity of TCRs defined over a congested CSC frequently exceeds the modeled limits for the CSC. When this occurs, the congestion revenue collected by ERCOT will be insufficient to satisfy the financial obligation to the holders of the TCRs and the revenue shortfall is collected from loads through uplift charges. The aggregate shortfall increased considerably to $61 million in 2007, up from $7 million in 2006. This increase was

6 A shadow price is the economic value of a constraint that is reflected in the zonal prices. The cap prevents the shadow price from rising above the cap.

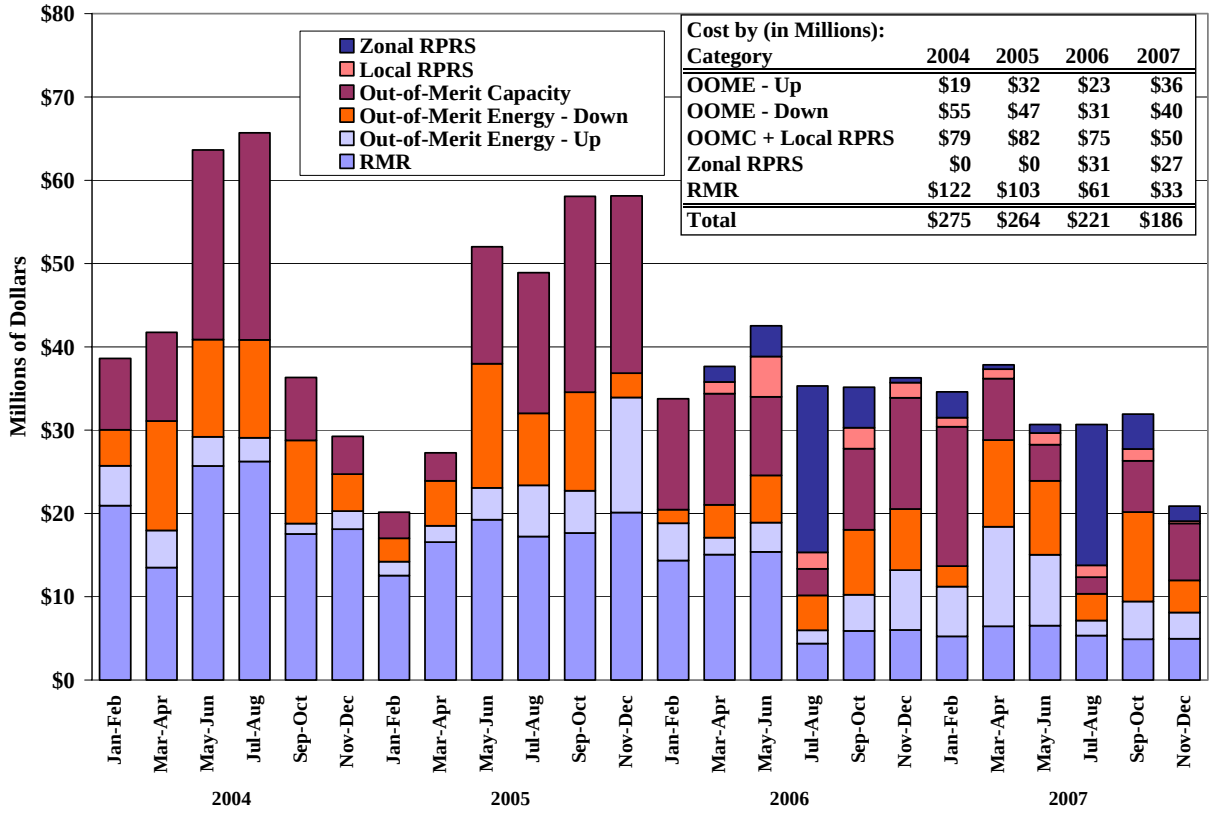
primarily due to increased interzonal congestion in 2007 and decreased accuracy in the quantity of TCRs sold in the monthly auction, especially for the West-to-North and North-to-West CSC.

In a perfectly efficient system with no uncertainty, the average congestion cost in real-time should equal the auction price of the congestion rights. In the real world, however, we would expect only reasonably close convergence with some fluctuations from year to year due to uncertainties. In 2006, market participants over-estimated the value of congestion on the South to North, South to Houston, and North to Houston CSCs. In 2007, market participants still over-estimated the value of congestion on the South to North and South to Houston CSCs, but significantly under-estimated the value of congestion on the North to Houston, North to West and West to North CSCs. The auction values correlate closely with actual congestion values from prior years, indicating that market participants have difficulty in accurately estimating future congestion costs.

3. Local Congestion and Local Capacity Requirements

ERCOT manages local (intrazonal) congestion using out-of-merit dispatch ("OOME up" and "OOME down"), which causes units to depart from their scheduled output levels. When not enough capacity is committed to meet local reliability requirements, ERCOT sends OOMC instructions for offline units to start up to provide energy and reserves in the relevant local area. ERCOT also enters into RMR agreements with certain generators needed for local reliability that may otherwise be mothballed or retired. When these units are called out-of-merit order, they receive revenues specified in the agreements rather than standard OOME or OOMC payments. The following figure shows the out-of-merit energy and capacity costs, including RMR costs, from 2004 to 2007.

Expenses for Out-of-Merit Capacity and Energy
2004-2007



Cost by (in Millions): Category	2004	2005	2006	2007
OOME - Up	$19	$32	$23	$36
OOME - Down	$55	$47	$31	$40
OOMC + Local RPRS	$79	$82	$75	$50
Zonal RPRS	$0	$0	$31	$27
RMR	$122	$103	$61	$33
Total	$275	$264	$221	$186

The results in the figure above show that overall uplift costs for RMR units, OOME units, and OOMC/Local RPRS units were relatively consistent between 2004 and 2005. The costs decreased by $74 million in 2006 from $264 million to $221 million, a reduction of 16 percent. In 2007, there was a further decrease from $221 million to $186 million, a reduction of 16 percent. There were substantial reductions to RMR cost due to the expiration of RMR agreements in 2007, which accounts for $28 million of the $35 million decrease from 2006 to 2007. Total OOME Up and OOME Down costs increased from $54 million in 2006 to $76 million in 2007. In constrast, out of merit commitment cost (OOMC and RPRS) decreased from $106 million in 2006 to $77 million in 2007.

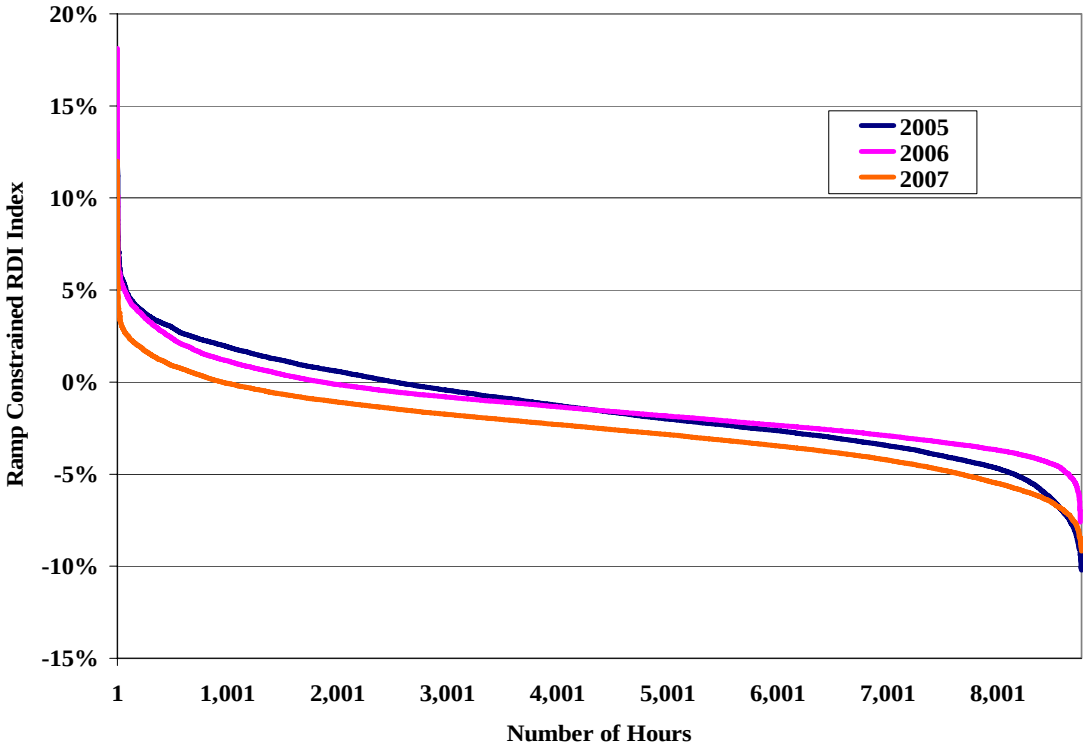
E. Analysis of Competitive Performance

The report evaluates two aspects of market power, structural indicators of market power and behavioral indicators that would signal attempts to exercise market power. The structural analysis in this report focuses on identifying circumstances when a supplier is "pivotal," *i.e.*,

when its generation is needed to serve the ERCOT load and satisfy the ancillary services requirements.

The pivotal supplier analysis indicates that the frequency with which a supplier was pivotal in the balancing energy market decreased significantly in 2007 compared to 2006. The following figure shows the ramp-constrained balancing energy market Residual Demand Index ("RDI") duration curves for 2005 and 2007. When the RDI is greater than zero, the largest supplier's balancing energy offers are necessary to prevent a shortage of offers in the balancing energy market.

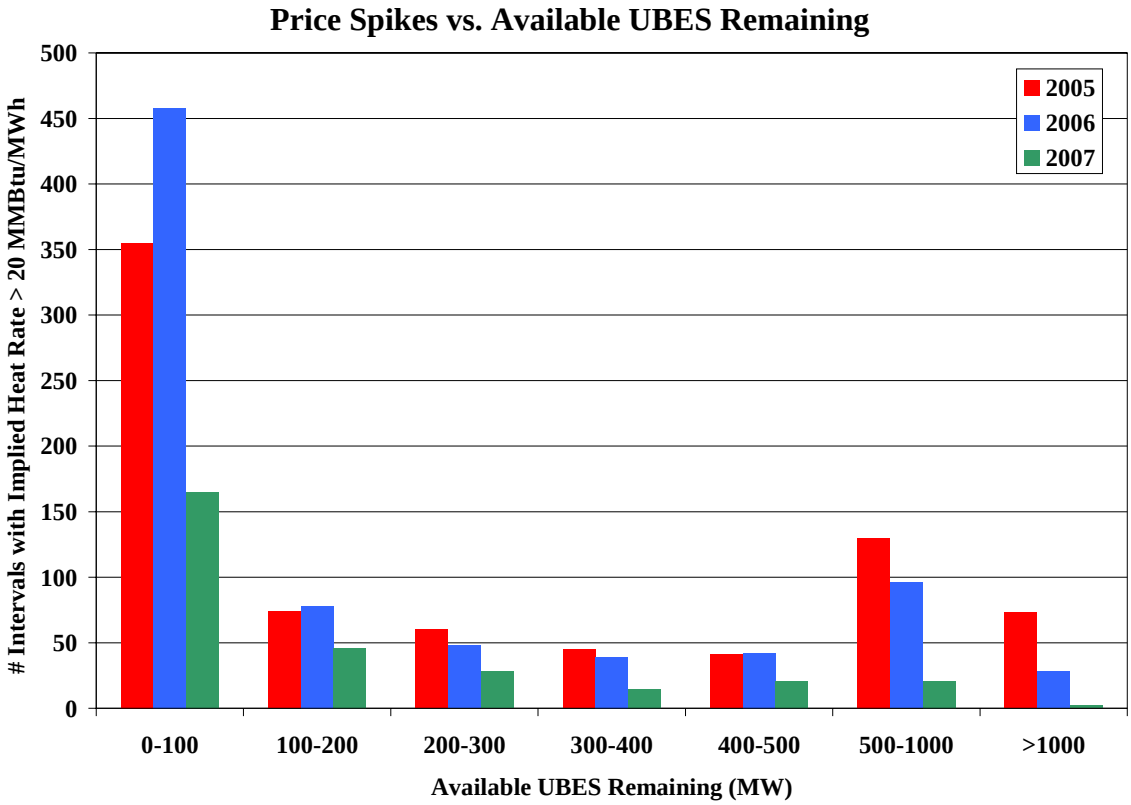
Ramp-Constrained Balancing Energy Market RDI Duration Curve 2005-2007



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

A final measure used to evaluate the potential for economic withholding analyzes the number of balancing energy market price spikes compared to the available Up Balancing Energy Service

("UBES") remaining. If the market is operating competitively, price spikes should occur during shortage and near shortage conditions, and the number of price spikes should reduce significantly as the amount of available surplus energy increases.



The results in the figure above indicate very competitive market outcomes in 2007, with over 92 percent of the price spikes occurring during intervals with less than 500 MW of available UBES remaining. These results show significant improvement over 2005 and 2006 when only 74 and 84 percent, respectively, of the price spikes occurred during intervals with less than 500 MW of available UBES remaining.

While structural market power indicators are very useful in identifying potential market power issues, they do not address the actual conduct of market participants. Accordingly, we analyzed measures of potential physical and economic withholding to further evaluate competitive performance of the ERCOT market. Potential withholding measures were examined relative to the level of demand and the size of each supplier's portfolio. The results of these analyses do not indicate significant concerns related to physical or economic withholding in 2007.

Overall, based upon the analyses in Section V, we find that the ERCOT wholesale market performed competitively in 2007.

F. Summary of Recommendations

As in prior reports, most of the operational issues identified in this report will be significantly improved with the implementation of the nodal market. As such, the following recommendations consist of issues that are either independent of the wholesale market model, or enhancements to the nodal market implementation:

- Real-time co-optimization of energy and reserves: As discussed in Section I.B., future implementation of real-time co-optimization of energy and reserves should be considered as a post-"go live" nodal market enhancement to further improve the efficient operation of the real-time market Real-time co-optimization.
- Operating Reserve Demand Curves: As discussed in Section I.D., relying upon the offers of small participants to ensure scarcity prices during legitimate shortage conditions produced unreliable results in 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. Ideally, operating reserve demand curves would be implemented in conjunction with real-time co-optimization of energy and reserves, although the latter is not an absolute prerequisite.
- Efficient Responsive Reserve Pricing: As discussed in Section III.C., ERCOT manages over-supply of Loads Acting as Resources ("LaaRs") in the responsive reserve market by relying upon administrative rules rather than prices to ration the product. This is inefficient and leads to excessive reliability costs for consumers. To improve the efficiency of responsive reserve pricing and incentives for suppliers, ERCOT should impose two responsive reserves constraints in the ancillary services auction: (i) that the responsive reserves procurement (including bilateral schedules) be greater than or equal to 2,300 MW and (ii) that the responsive reserves procurement from LaaRs (including bilateral schedules) be less than or equal to 1,150 MW. The clearing price paid to generators would be equal to the shadow price of the first constraint only, while the clearing price paid to LaaRs would be equal to the shadow price of the first constraint

minus the shadow price of the second constraint (a single price would result if the LaaR constraint is not binding).

- Day-Ahead Load Forecast Error: As discussed in Section I.D., ERCOT's day-ahead load forecast exhibited a persistent positive bias in 2007 that was particularly high during the summer months, which 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. ERCOT should review the causes of the positive bias in its day-ahead load forecast,
- Assessment of Ancillary Service Products and Quantities: 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 recently. Substantial addition of more unpredictable and uncontrollable resources has significant implications related to efficient and reliable unit commitment and real-time operations.
- Re-evaluation of the Reserve Discount Factor: As discussed in Section I.D., ERCOT implemented a factor that discounts the stated capacity of online generating units for the purpose of calculating available responsive reserves in 2007. To compensate for the application of the discount factor, the quantity of responsive reserves procured was increased by amounts ranging from 200 to 500 MW in 2008. In parallel, Protocol Revision Request ("PRR") No. 750 was implemented in March 2008 related to unannounced unit testing. The objective of this increased testing is increased confidence in the stated capacity of generating resources and the elimination of the discount factor, thereby also eliminating the incremental quantities of responsive reserve procurement. The increased responsive reserve quantities are an interim measure. The more efficient and less costly solution for consumers is to re-establish confidence in the stated capacity values for generating resources. Therefore, ERCOT should obtain sufficient unit testing data to provide for a statistical re-evaluation of the reserve discount factor and the associated increased quantities of responsive reserve in 2008. If possible, ERCOT should

eliminate the discount factor or at least reducing it to two percent or lower (which would eliminate the procurement of additional responsive reserve quanties above 2,300 MW).

- Peaker Net Margin Calculation: As discussed in Section I.D., PUCT rules specify the price that is used to calculate the peaker net margin as the price at an ERCOT-wide hub.[7] Essentially, this is an average price for the ERCOT market. To better account for regional price disparities, we recommend that the price that is used in the peaker net margin calculation in the PUCT's rules be modified to be a set of regional prices, and that the cumulative peaker net 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, we recommend ERCOT transition from the high system offer cap to the low system offer cap for the duration of the scarcity pricing mechanism cycle.

[7] The Peaker Net Margin ("PNM") is designed to measure the annual net revenue for a hypothetical peaking unit. Under PUCT rules, if the PNM reaches a cumulative total of $175,000 per MW in a calendar year, the system-wide offer cap is reduced to the higher of $500 per MWh or 50 times the daily gas price index.

I. REVIEW OF MARKET OUTCOMES

A. Balancing Energy Market

1. Balancing Energy Prices During 2007

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

Unless there are barriers that prevent arbitrage of the prices in the spot and forward markets, the prices in the forward market should be directly related to the prices in the spot market (*i.e.*, the spot prices and forward prices should converge over the long-run).[8] Hence, artificially-low prices in the balancing energy market will translate to artificially-low forward prices. Likewise, price spikes in the balancing energy market will increase prices in the forward markets. The analyses in this section summarize and evaluate the prices that prevailed in the balancing energy market during 2007.

To summarize the price levels during the past two years, Figure 1 shows the load-weighted average balancing energy market prices in each of the ERCOT zones in 2006 and 2007.[9] Balancing energy market prices were 2 percent higher in 2007 than in 2006, with September 2007 showing the largest increase from the same month in 2006.

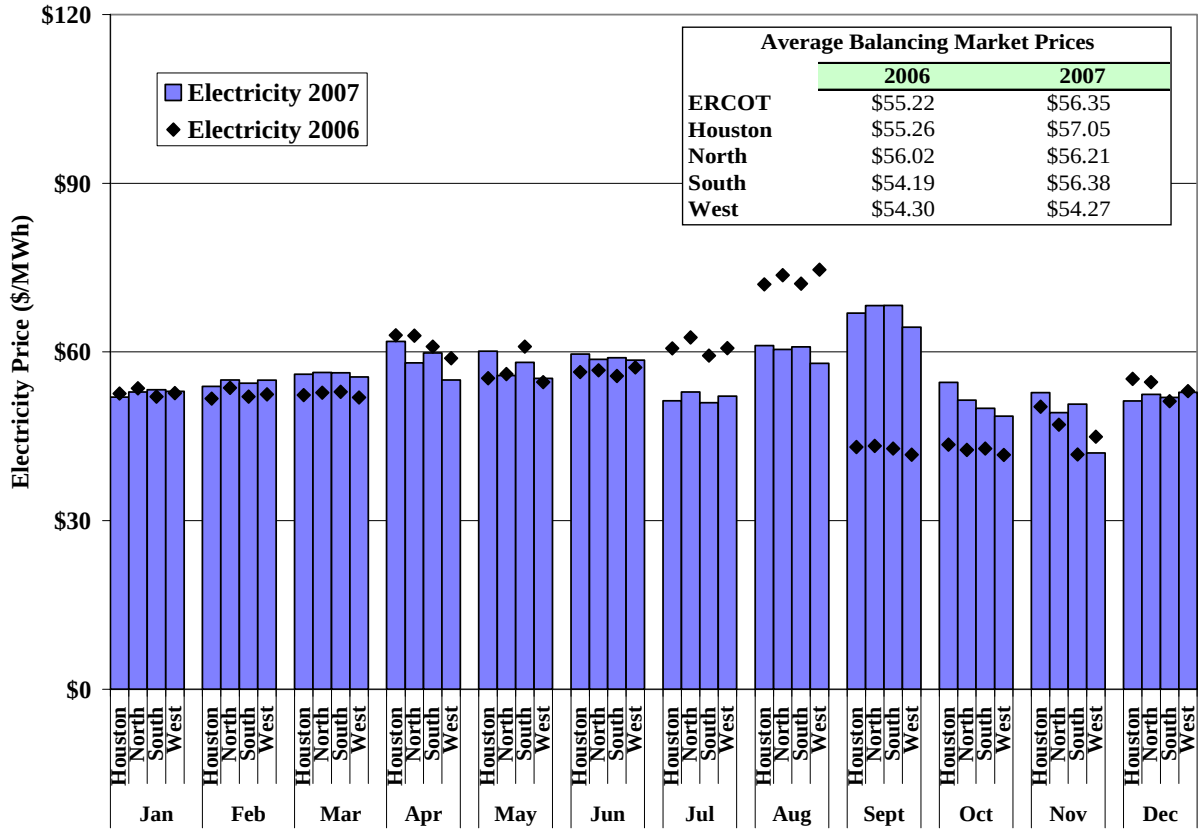
The average natural gas price in 2007 increased 4 percent from 2006, with the largest increase occurring in September at 25 percent. Natural gas is typically the marginal fuel in the ERCOT

8 See Hull, John C. 1993. *Options, Futures, and other Derivative Securities*, second edition. Englewood New Jersey: Prentice Hall, p. 70-72.

9 The load-weighted average prices are calculated by weighting the balancing energy price in each interval and zone by the total zonal loads in that interval. This is not consistent with average prices reported elsewhere that are weighted by the balancing energy procured in the interval, which is a methodology we use to evaluate certain aspects of the balancing energy market. For this evaluation, balancing energy prices are load-weighted since this is the most representative of what loads are likely to pay (assuming that balancing energy prices are generally consistent with bilateral contract prices).

market. Hence, the changes in energy prices from 2006 to 2007 were largely a function of natural gas price movements.

Figure 1: Average Balancing Energy Market Prices 2006 & 2007



Average Balancing Market Prices

	2006	2007
ERCOT	$55.22	$56.35
Houston	$55.26	$57.05
North	$56.02	$56.21
South	$54.19	$56.38
West	$54.30	$54.27

The next analysis evaluates the total cost of serving load in the ERCOT market. In addition to the costs of energy, loads incur costs associated with operating reserves, regulation, and "uplift".[10] We have calculated an average all-in price of electricity for ERCOT that is intended to reflect energy costs as well as these additional costs. Figure 2 shows the monthly average all-in price for all of ERCOT from 2003 to 2007.

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

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

- Energy costs: Balancing energy market prices are used to estimate energy costs, under the assumption that the price of bilateral energy purchases converges with balancing energy market prices over the long-term, as discussed above.
- Ancillary services costs: These are estimated based on the demand and prices in the ERCOT markets for regulation, responsive reserves, and non-spinning reserves.
- Uplift costs: Uplift costs are assigned market-wide on a load-ratio share basis.

Figure 2: Average All-in Price for Electricity in ERCOT 2003 to 2007

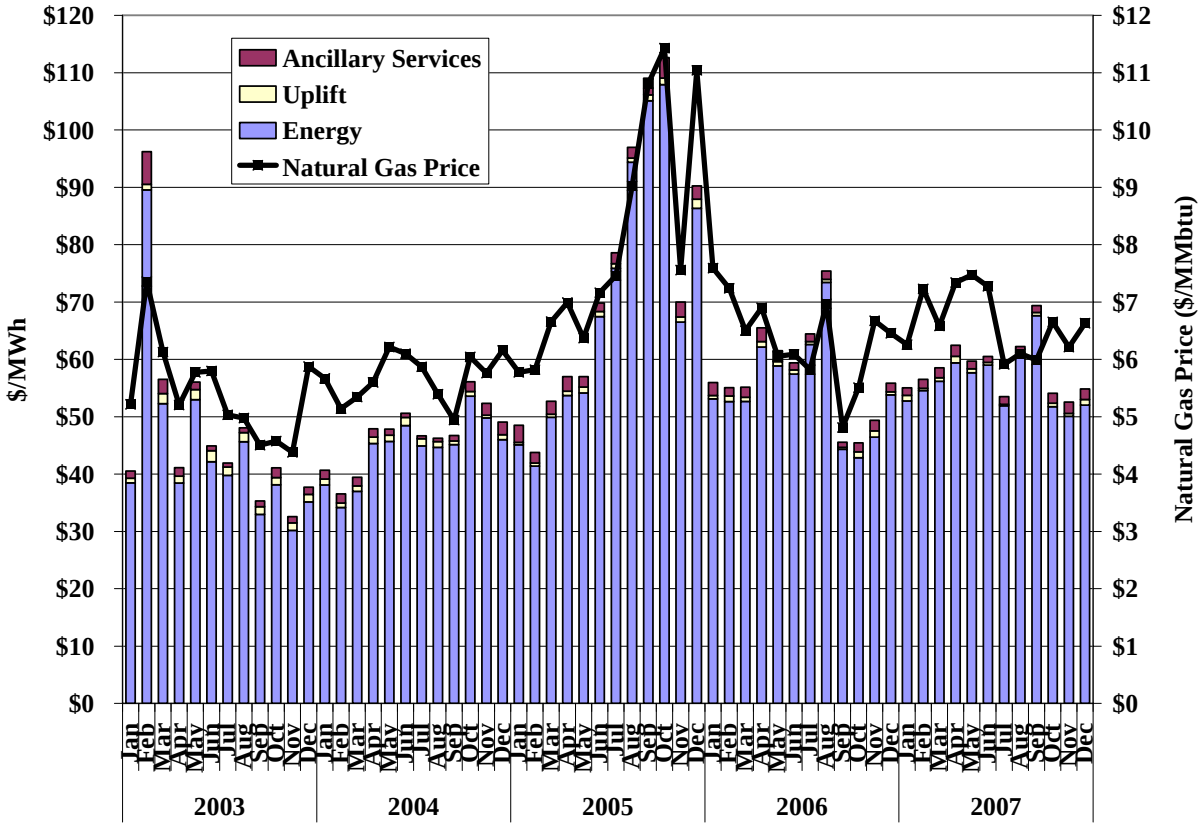


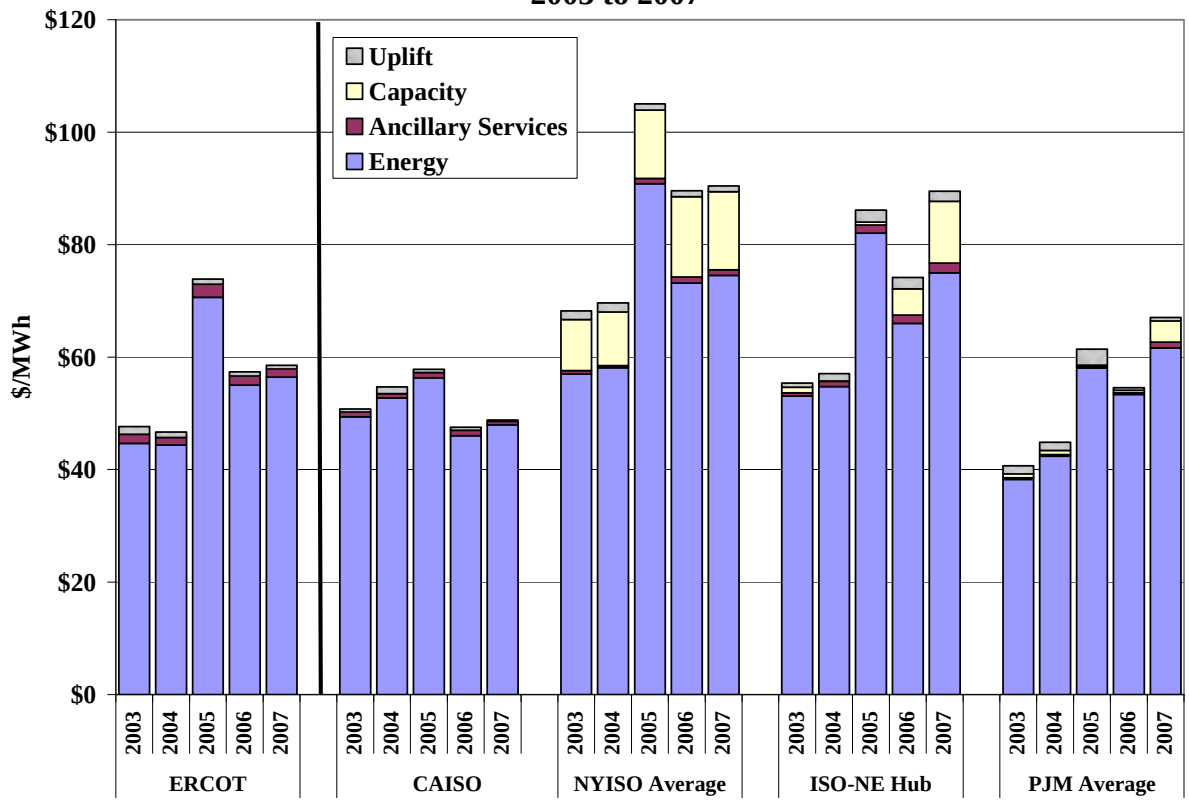
Figure 2 indicates that natural gas prices were a primary driver of the trends in electricity prices from 2003 to 2007. This is not surprising given that natural gas is the predominant fuel in ERCOT, especially among the generating units that most frequently set the balancing energy market prices. In 2007, the average natural gas price increased by 4 percent from 2006 levels and the all-in price for electricity increased by 0.5 percent.

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 three other factors contributed to price outcomes in 2007. First, as discussed in Section III of this report, ERCOT peak demand and installed capacity were relatively flat in 2007, and energy production increased only slightly in 2007 compared to 2006. In contrast to prior years with increasing demand and decreasing supply, the static supply and demand characteristics from 2006 to 2007 contributed to comparable wholesale pricing outcomes over the course of these two years. Second, the balancing energy offer cap was raised to $1,500 in 2007, whereas the offer cap was $1,000 in 2006. The increased offer caps are intended to produce higher prices during system shortage conditions. However, as discussed later in this section, this mechanism was not always effective in achieving this intended outcome. Finally, the overall competitive performance of the market exhibited continued improvement in 2007, which will tend to lower prices and is examined in detail in Section V. Analyses in the next sub-section adjust for natural gas price fluctuations to better highlight variations in electricity prices not related to fuel costs.

From 2006 to 2007, an 8 percent decrease in ancillary services costs result in a 0.2 percent decrease in the all-in price for electricity. Generally, the ancillary service prices coincided with price movements in the balancing energy market, which is to be expected since the energy and ancillary service requirements are satisfied by the same resources.

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

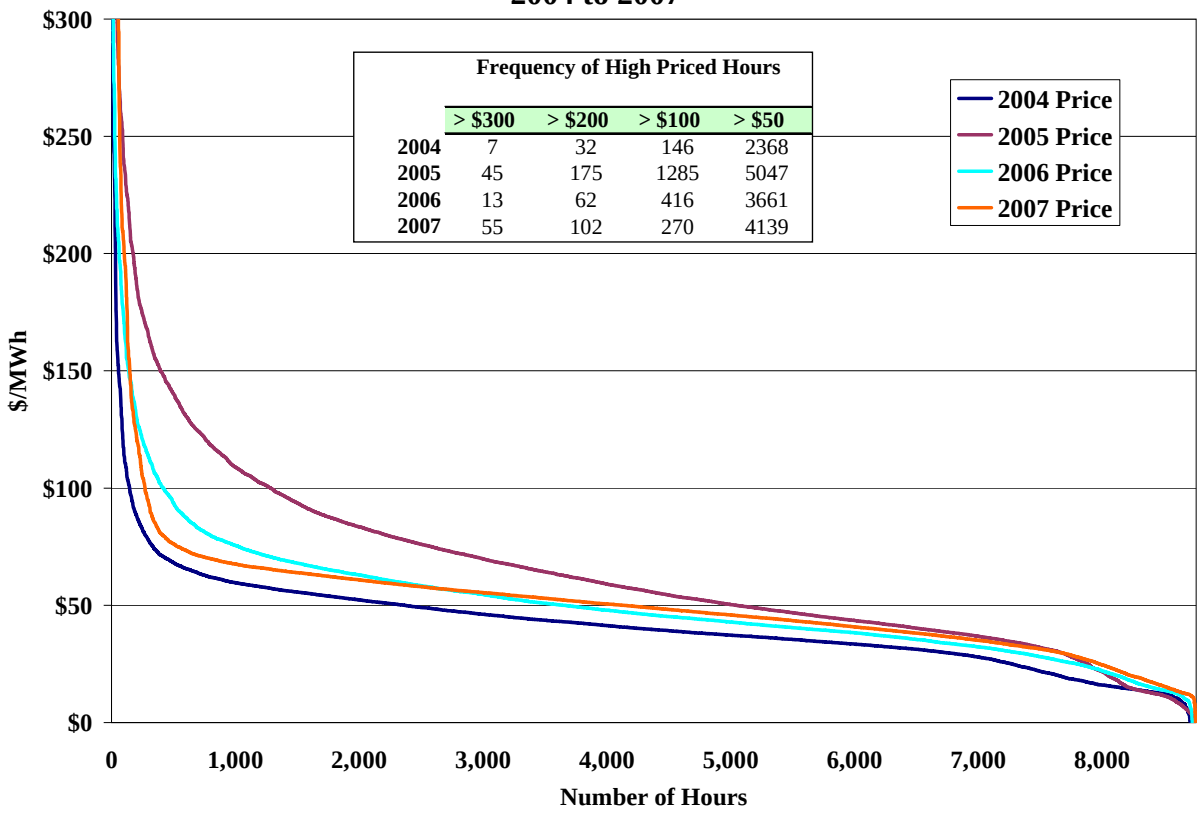
Figure 3: Comparison of All-in Prices Across Markets 2003 to 2007



Wholesale electricity markets in the U.S. experienced substantial increases in energy prices from 2004 to 2005 due to increased fuel costs. In 2006, energy prices in the U.S. dropped in every region due to decreased fuel costs. In 2007, the all-in prices increased in all the above five regions, with relatively small increases in ERCOT, California and New York, and more significant increases in New England and PJM.

Figure 4 presents price duration curves for the ERCOT balancing energy market in each year from 2004 to 2007. A price duration curve indicates the number of hours (shown on the horizontal axis) that the price is at or above a certain level (shown on the vertical axis). The prices in this figure are hourly load-weighted average prices for the ERCOT balancing energy market.

Figure 4: ERCOT Price Duration Curve
2004 to 2007



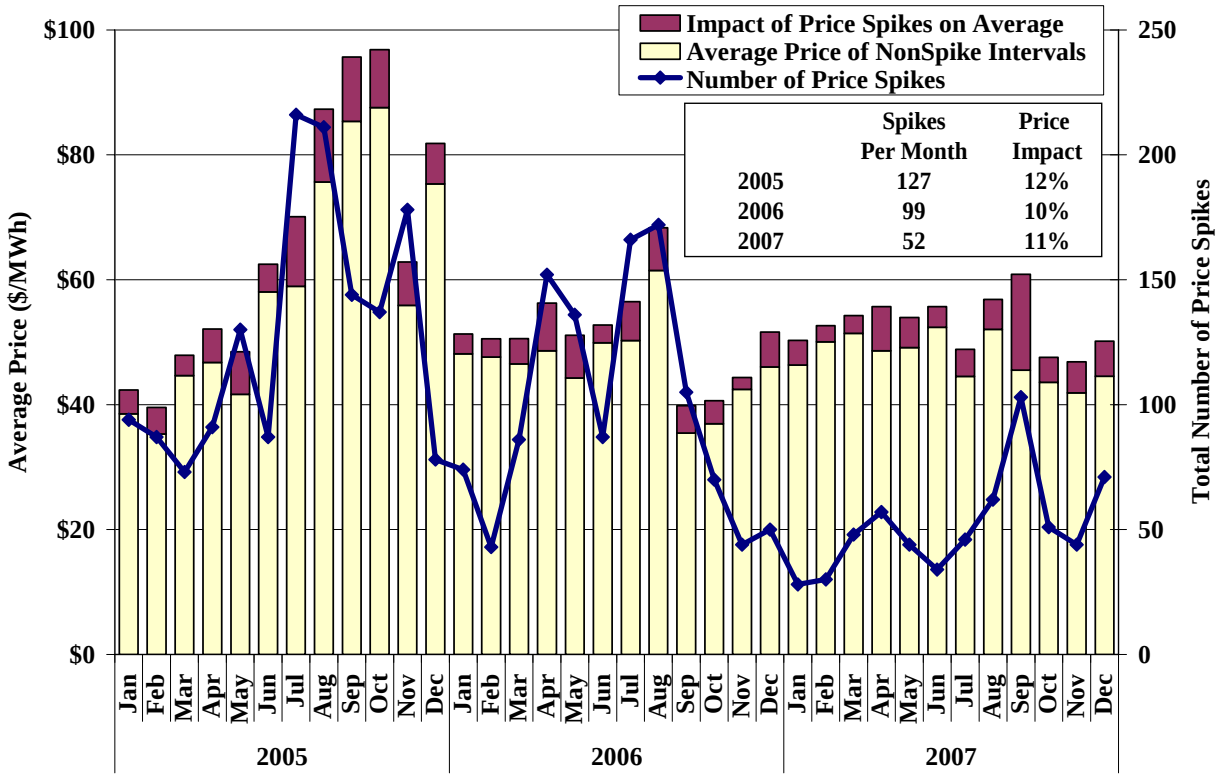
Frequency of High Priced Hours

	> $300	> $200	> $100	> $50
2004	7	32	146	2368
2005	45	175	1285	5047
2006	13	62	416	3661
2007	55	102	270	4139

Balancing energy prices exceeded $50 in more than 4,000 hours in 2007 compared to more than 3,500 hours in 2006. These year-to-year changes reflect the effects of slightly higher fuel prices in 2007, which impact electricity prices in a broad range of hours.

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. Figure 4 shows that there were differences in balancing energy market prices between 2004 and 2007 at the highest price levels. For example, 2007 experienced considerably more price spikes greater than $300 per MWh than 2005 or 2006, even though average prices were comparable to 2006 and lower than in 2005. 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 2007. Figure 4 shows average prices and the number of price spikes in each month of 2005 to 2007. 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 5: Average Balancing Energy Prices and Number of Price Spikes 2005 to 2007



The number of price spike intervals was 127 per month during 2005. The number decreased in 2006 to 99 per month, and further decreased to 52 per month in 2007. 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 per MWh during 2005. In 2006, the impact was $4.68 per MWh in average in 2006 and the impact averaged $5.30 per MWh in 2007. Even though price spikes account for a small portion of the total intervals, they have a significant impact on overall price levels.

Figure 6 through Figure 8 show the frequency of price spikes in the regulation and responsive reserve markets during 2005 through 2007. These figures show that price spikes in the markets for ancillary services have also dropped significantly over this time period.

Figure 6: Average Regulation Up Prices and Number of Price Spikes 2005 to 2007

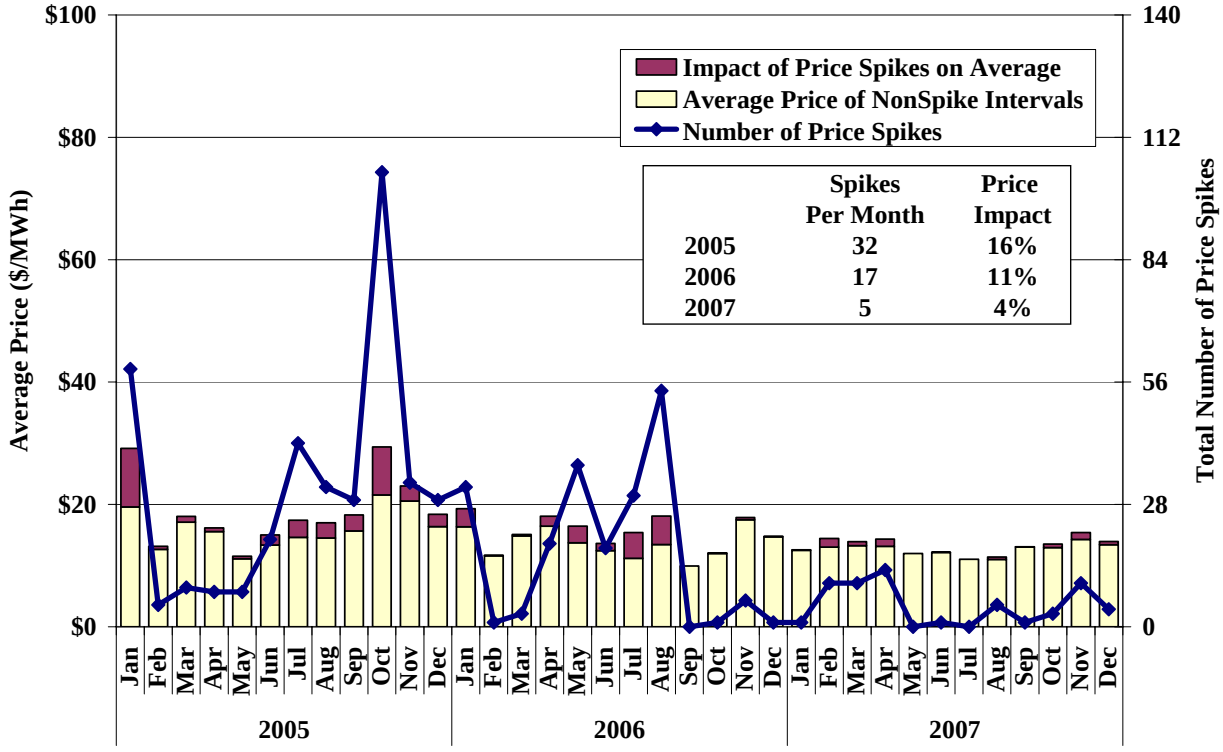


Figure 7: Average Regulation Down Prices and Number of Price Spikes 2005 to 2007

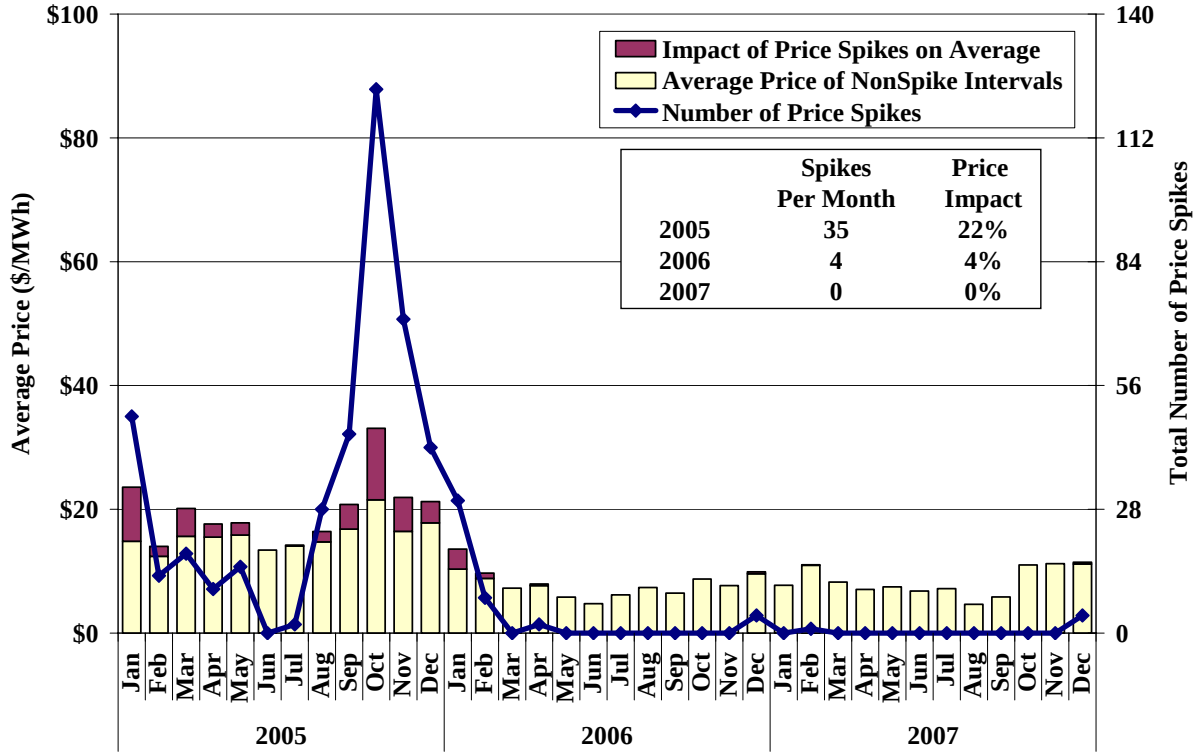
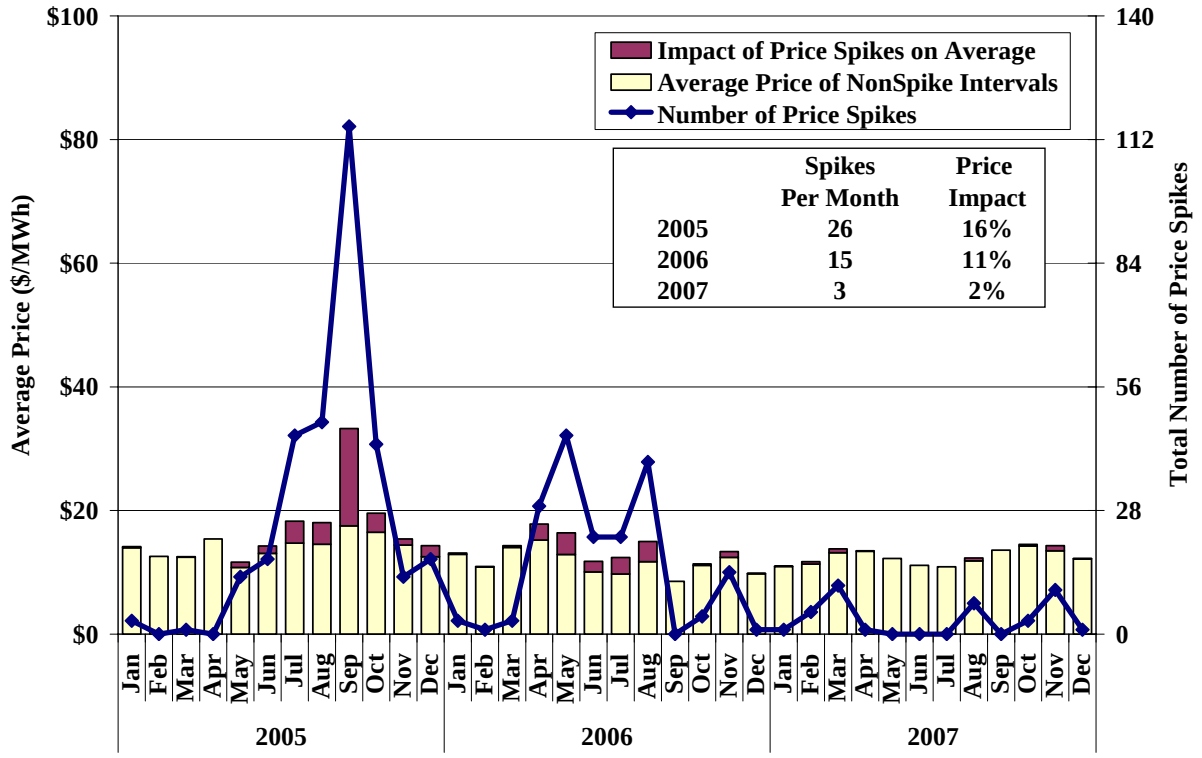


Figure 8: Average Responsive Reserve Prices and Number of Price Spikes 2005 to 2007



During 2005, there were 32 price spike hours per month for regulation up, 35 for regulation down, and 26 for responsive reserves.[11] In 2006, the number of price spike hours decreased, with 17 per month for regulation up, 4 per month for regulation down, and 15 per month for responsive reserves. In 2007, the number of price spike hours further decreased, with 5 per month for regulation up, 0 for regulation down, and 3 for responsive reserves. Because the same resources are used to supply ancillary services and energy, fluctuations in energy prices should lead to corresponding changes in ancillary services prices. The relationship between balancing energy prices and ancillary services prices is discussed in greater detail later in this section.

While the price spikes directly impact a small portion of the total consumption of energy and ancillary services, persistent price spikes will eventually flow through to consumers. Price spikes in the ancillary service markets have decreased over the last three years, as has the frequency of overall price spikes in the balancing energy market. However, the frequency of extreme price spikes (i.e., prices greater than $300 per MWh) was higher in 2007 than in 2005 or

[11] Price spikes are defined as hours where the price exceeds a threshold of $50 per MW for regulation up, regulation down, or responsive reserves.

2006. To the extent that price spikes reflect true scarcity of generation resources, they send efficient economic signals in the short-run for commitment and dispatch, and in the long-run for new investment. However, to the extent that price spikes occur when economic resources are not efficiently utilized, they raise costs to consumers and send inefficient economic signals. This issue is examined in more detail in Section V.

2. Balancing Energy Prices Adjusted for Fuel Price Changes

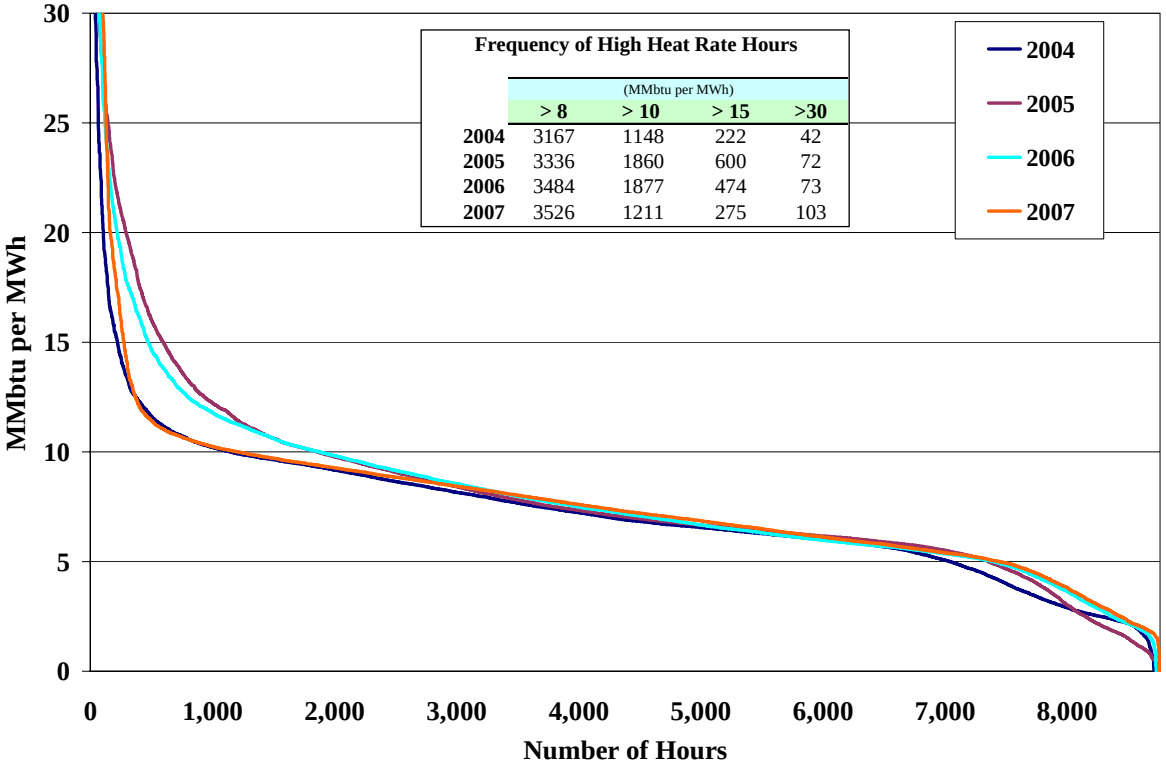
The pricing patterns shown in the prior sub-section are driven to a large extent by changes in fuel prices, natural gas prices in particular. However, prices are influenced by a number of other factors as well. To clearly identify changes in electricity prices that are not driven by changes in natural gas prices, Figure 9 and Figure 10 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*.[12] The second chart shows the same duration curves for the top five percent of hours in each year. The figure shows duration curves for the implied marginal heat rate for 2003 to 2007.

In contrast to Figure 4, Figure 9 shows that the implied marginal heat rates were relatively consistent across the majority of hours from 2003 to 2007. For instance, the table in Figure 9 indicates that the number of hours when the implied heat rate exceeded 8 MMbtu per MWh was relatively consistent across the five years. The rise in energy prices from 2003 to 2007 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 10 was 1,860 in 2005 and 1,877 in 2006, but declined to 1,211 in 2007. This indicates that there are price differences that are due to factors other than changes in natural gas prices.

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

Figure 10 shows the implied marginal heat rates for the top five percent of hours in 2004 through 2007. These data reveal that the frequency of price spikes with an implied marginal heat rate greater than 30 increased significantly in 2007 compared to prior years.

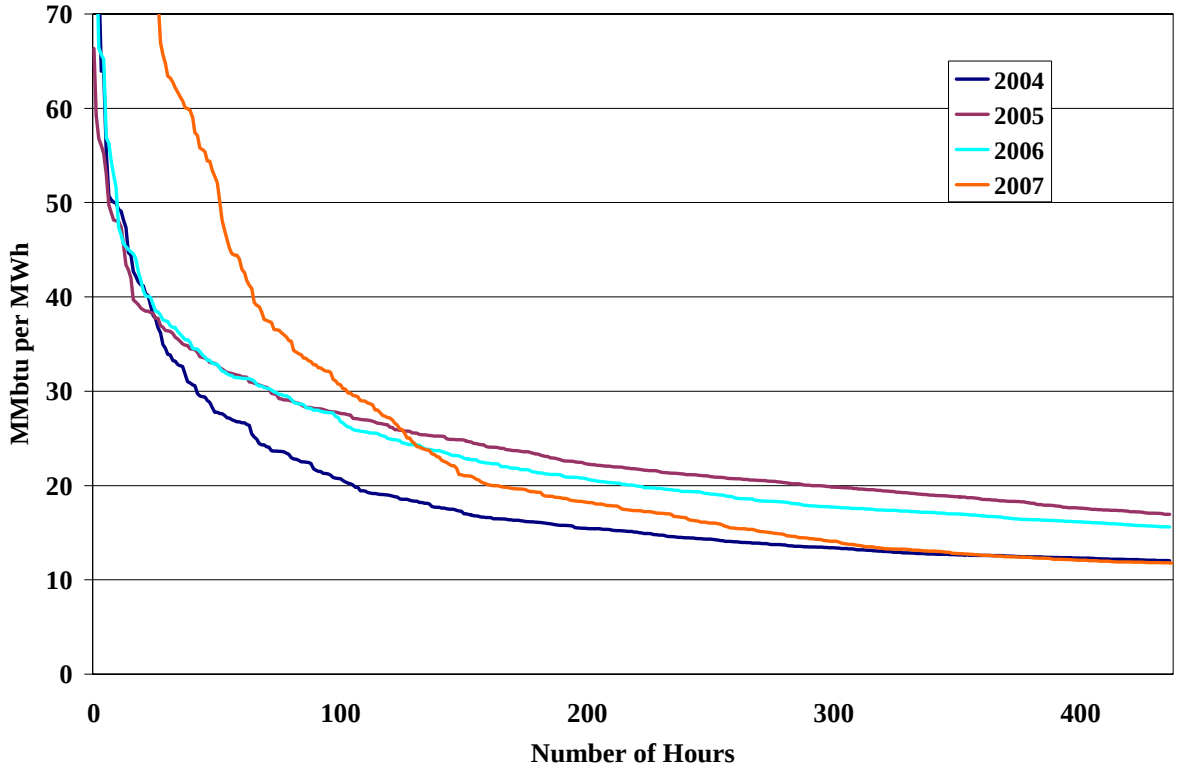
Figure 9: Implied Marginal Heat Rate Duration Curve
All Hours – 2004 to 2007



Frequency of High Heat Rate Hours

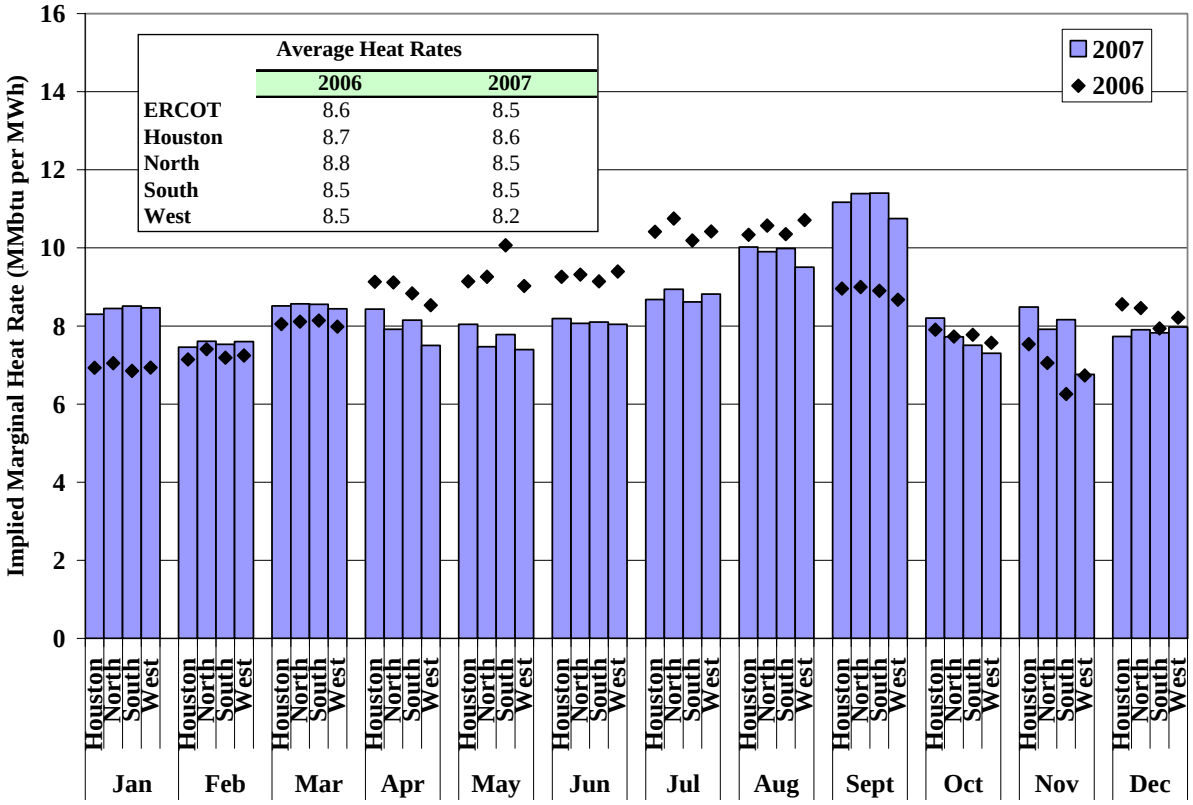
	(MMbtu per MWh)			
	> 8	> 10	> 15	>30
2004	3167	1148	222	42
2005	3336	1860	600	72
2006	3484	1877	474	73
2007	3526	1211	275	103

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

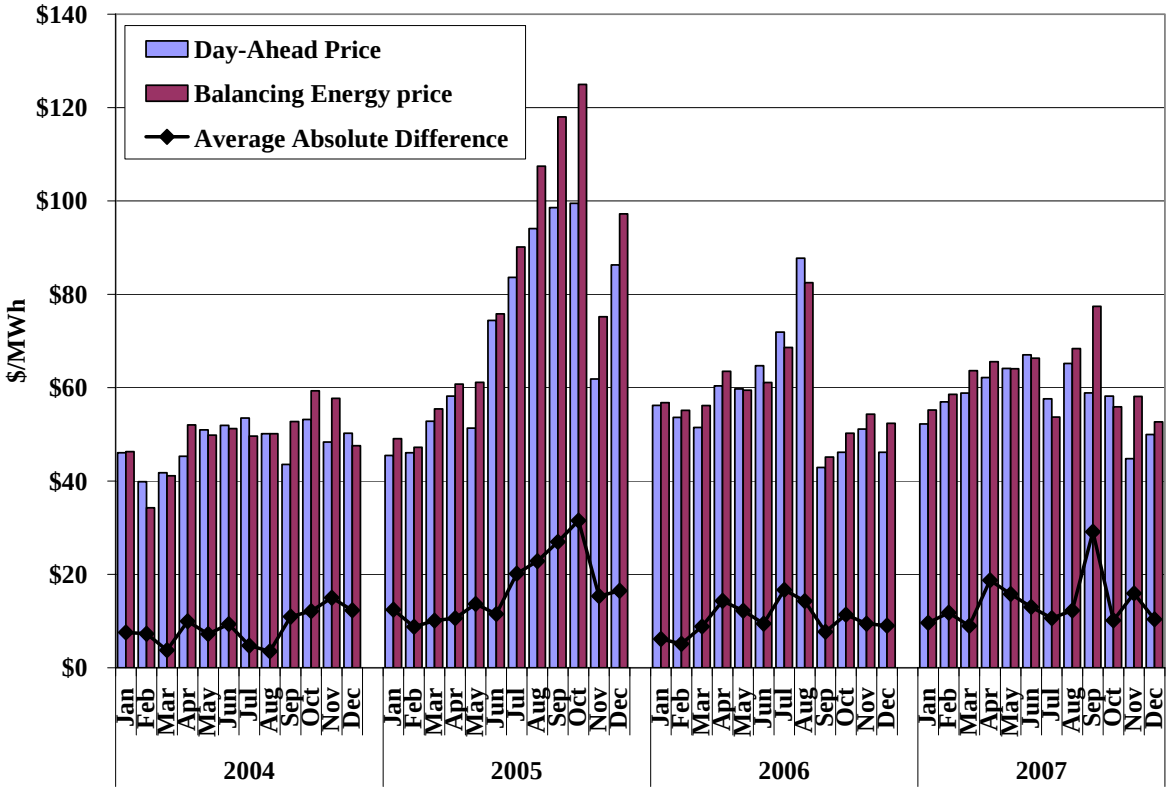


Figure 12 shows price convergence during peak periods (i.e. weekdays between 6 AM and 10 PM). This timeframe matches the definition of peak hours that are commonly traded in the forward market. During most of 2004, the average day-ahead price was consistent with the average balancing energy price. However,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

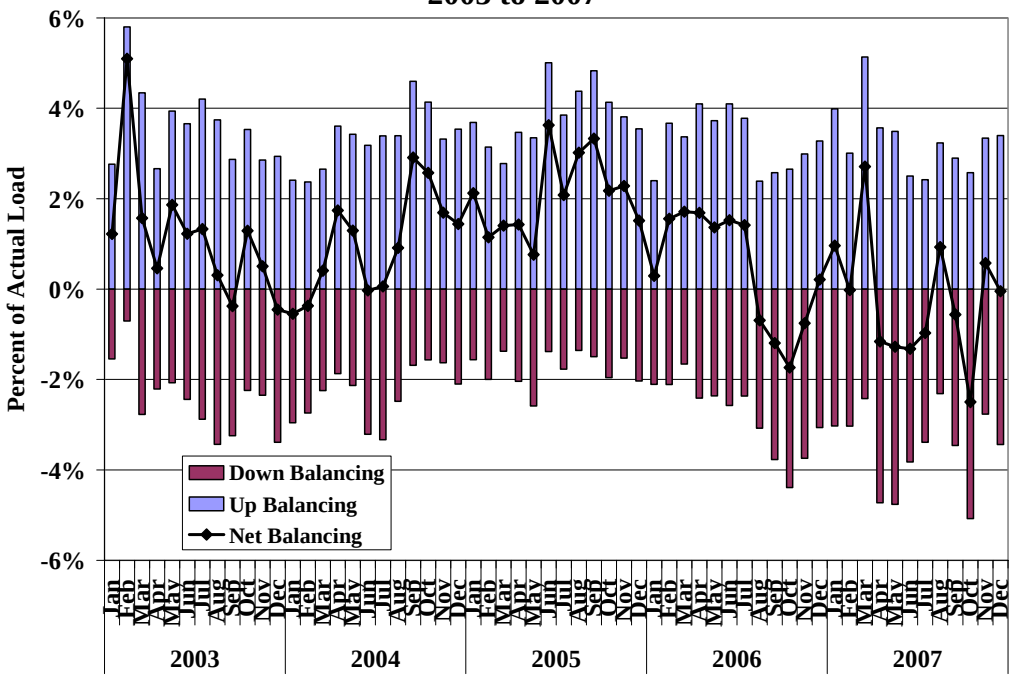


Figure 13 shows that the total volume of balancing up and balancing down energy as a share of actual load increased from an average of 5.6 percent in 2005 to 6.1 percent in 2006 and 6.8 percent in 2007. Starting in August 2006, the average volume of balancing down energy began to increase. In 2007, the 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

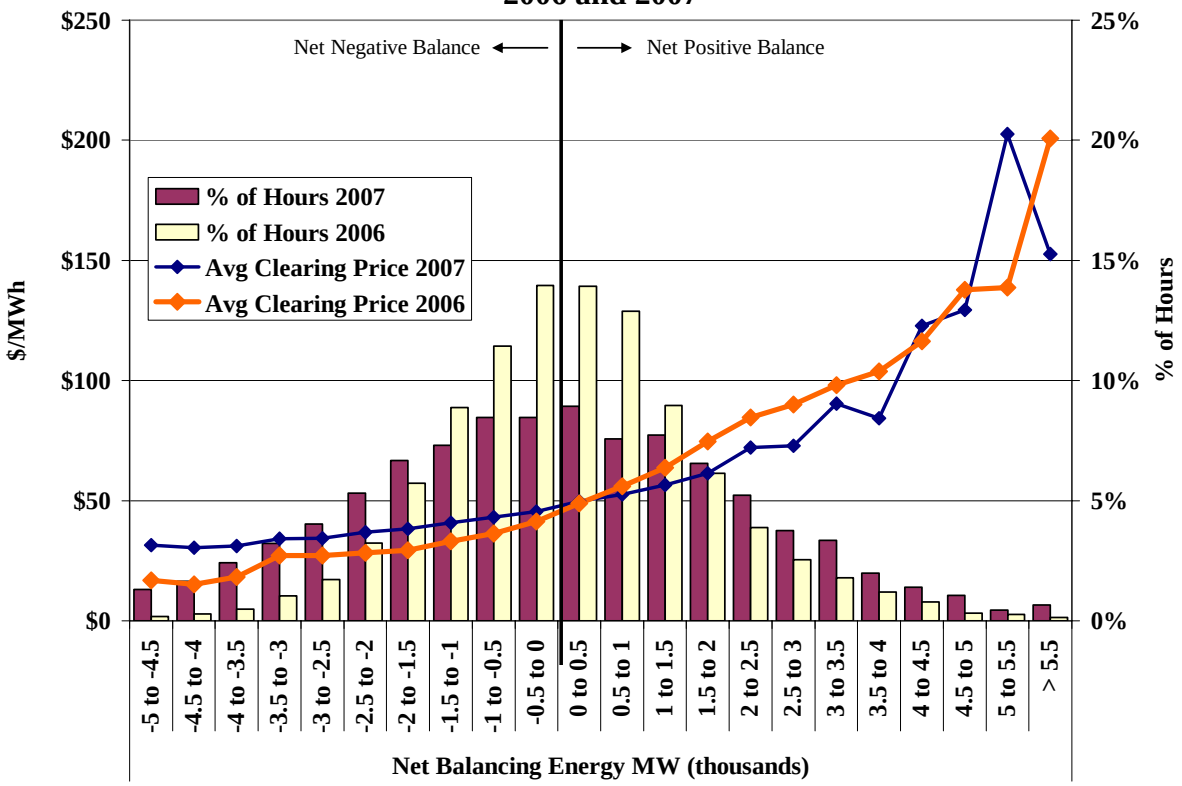


Figure 14 shows a relatively symmetrical distribution of net balancing energy purchases in 2007 centered around zero gigawatts, but the distribution is wider and flatter than 2006. This is consistent with Figure 13 which showed that there were comparable portions of net balancing up and down quantities 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,

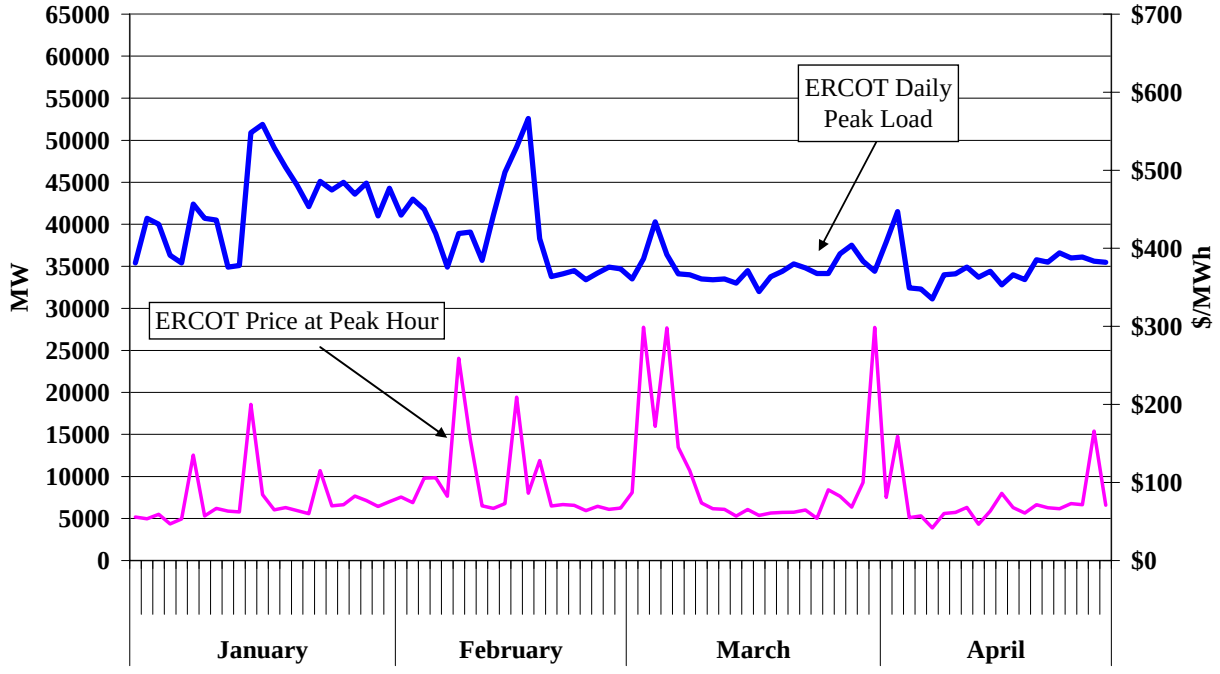
one should expect that prices would be primarily determined by more fundamental factors, such as actual load levels and fuel prices. However, this figure clearly indicates that balancing energy prices increase as net balancing energy volumes increase. This is also consistent with the patterns of prices 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.

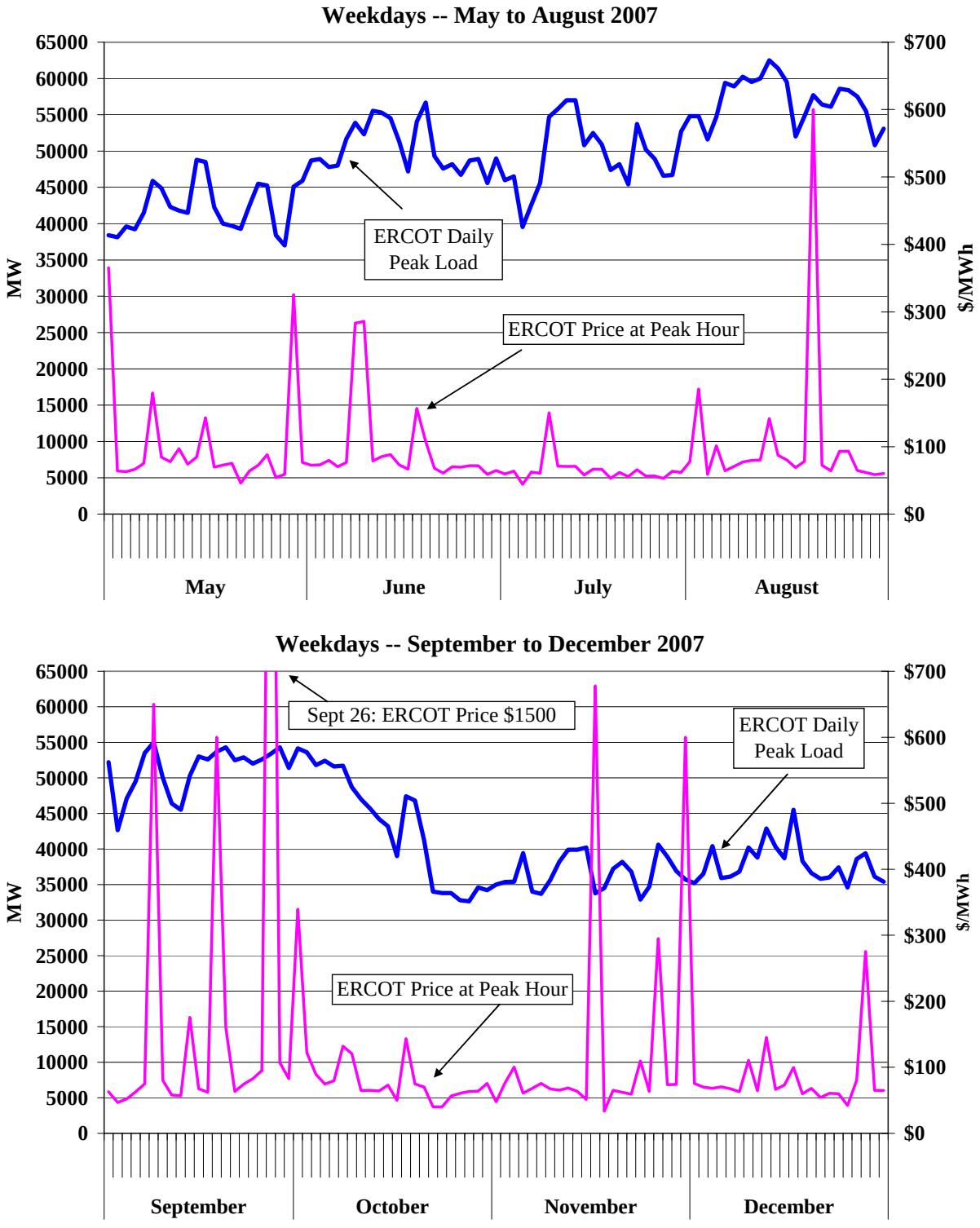
5. Determinants of Balancing Energy Prices

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





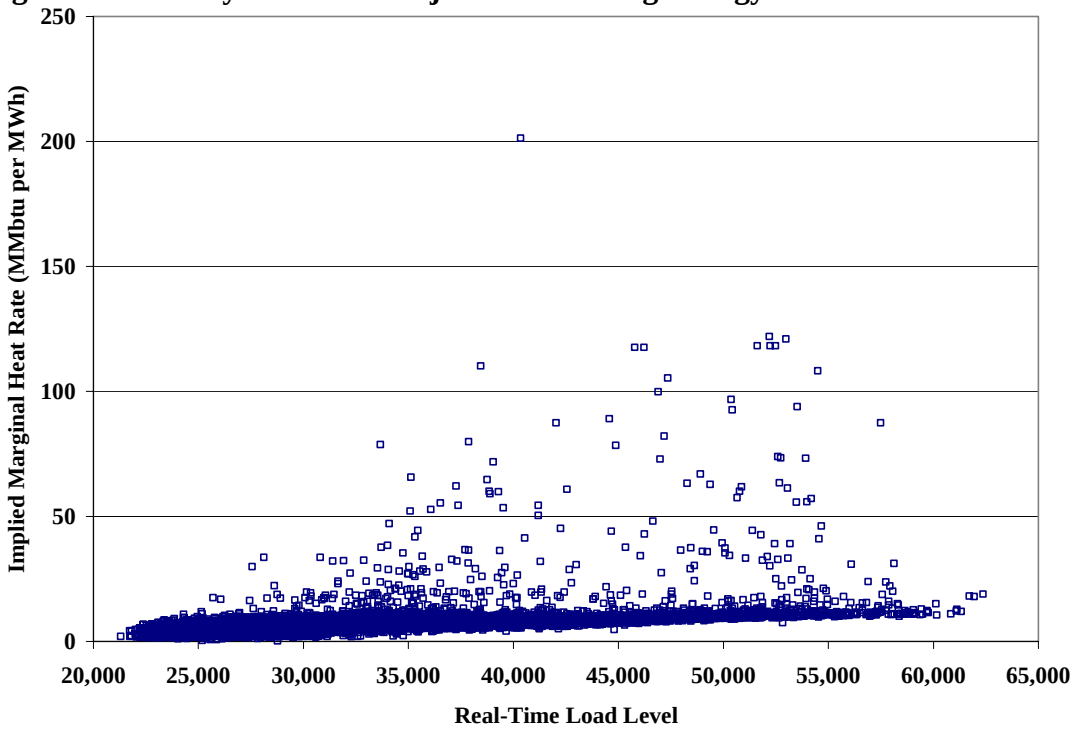
The figures indicates some relationship between high prices (*e.g.*, greater than $200/MWh) and periods when demand is high or rising significantly relatively to the previous days, although the

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

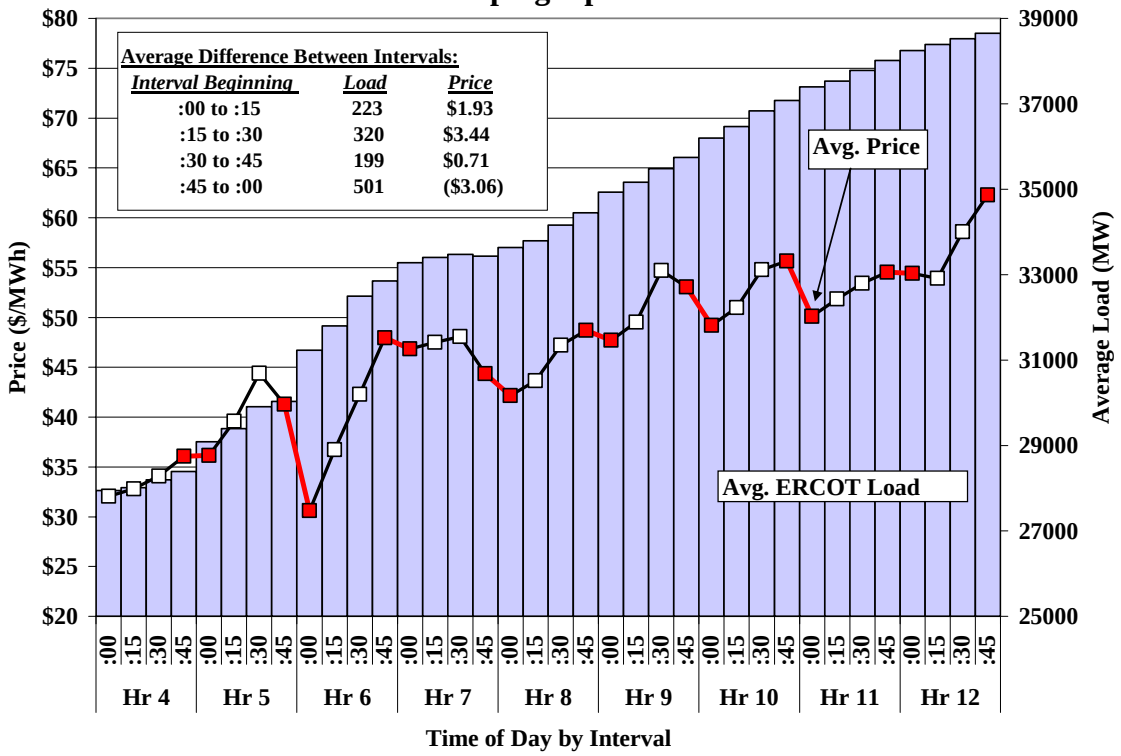


Figure 17 shows that, with the exception of hour 7 and 9, the load steadily increases in every interval and prices generally move upward from about $32 per MWh at 4:00 AM to $62 per MWh at 12:45 PM. If actual load were the primary determinant of energy prices, the balancing energy prices would rise gradually 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

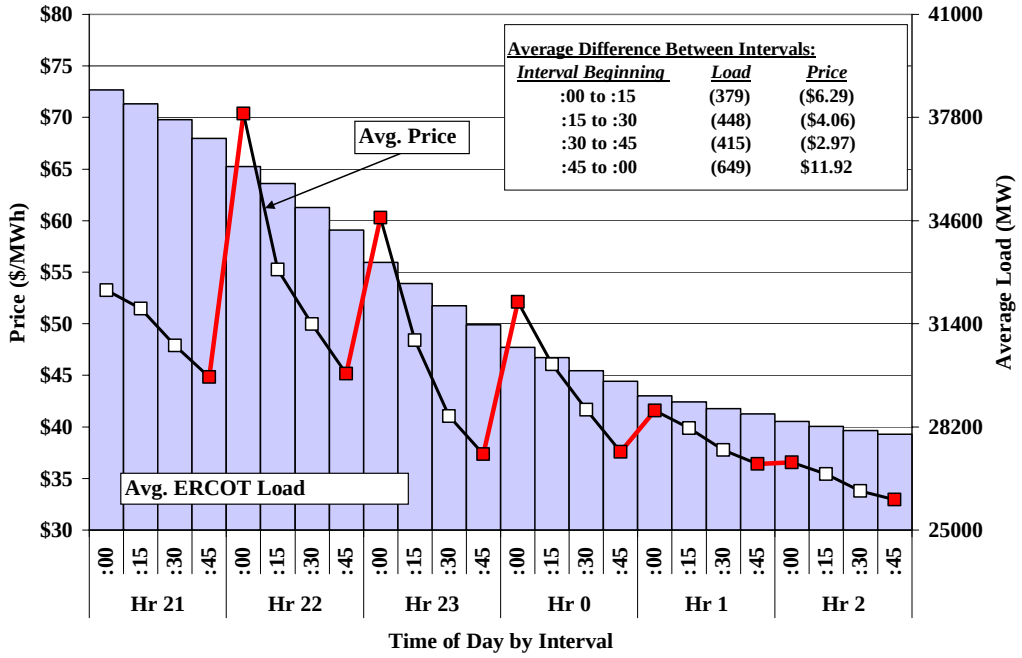


Figure 18 shows that while balancing energy prices decrease over these intervals, they follow a similar pattern as exhibited in the ramping-up hours. The balancing energy price decreases in each interval of the hour before rising substantially in the first interval of the following hour. The balancing 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.

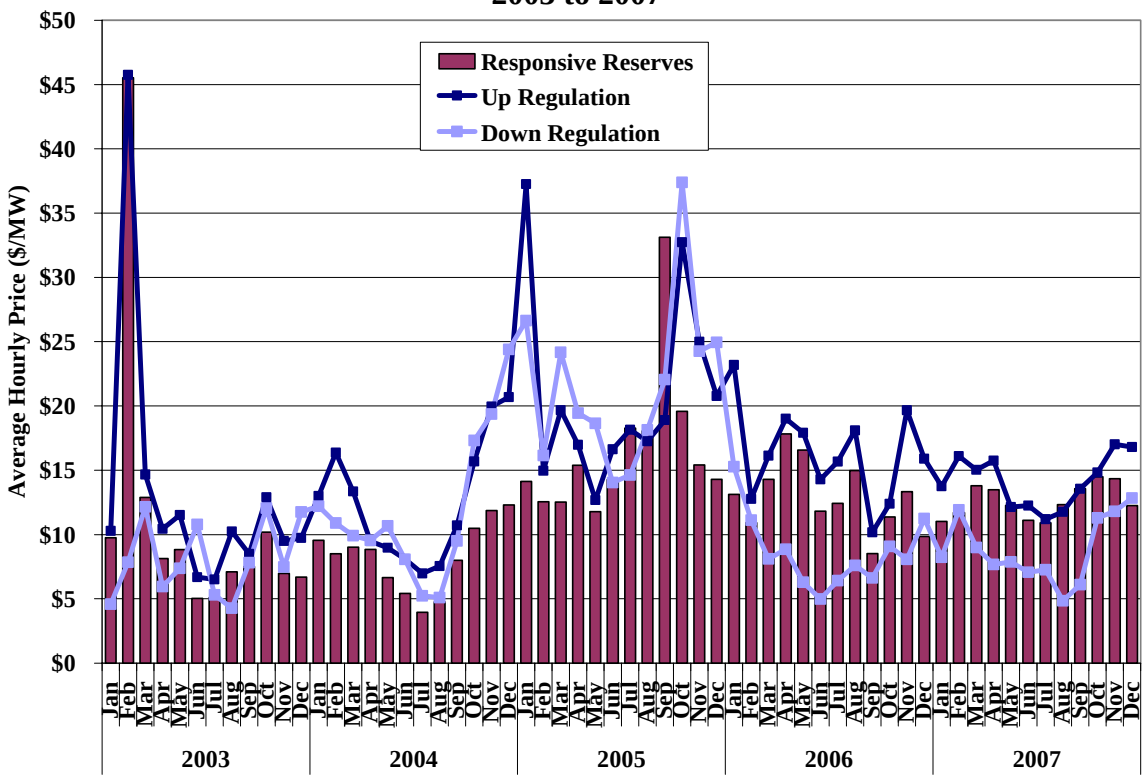
B. Ancillary Services Market Results

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.

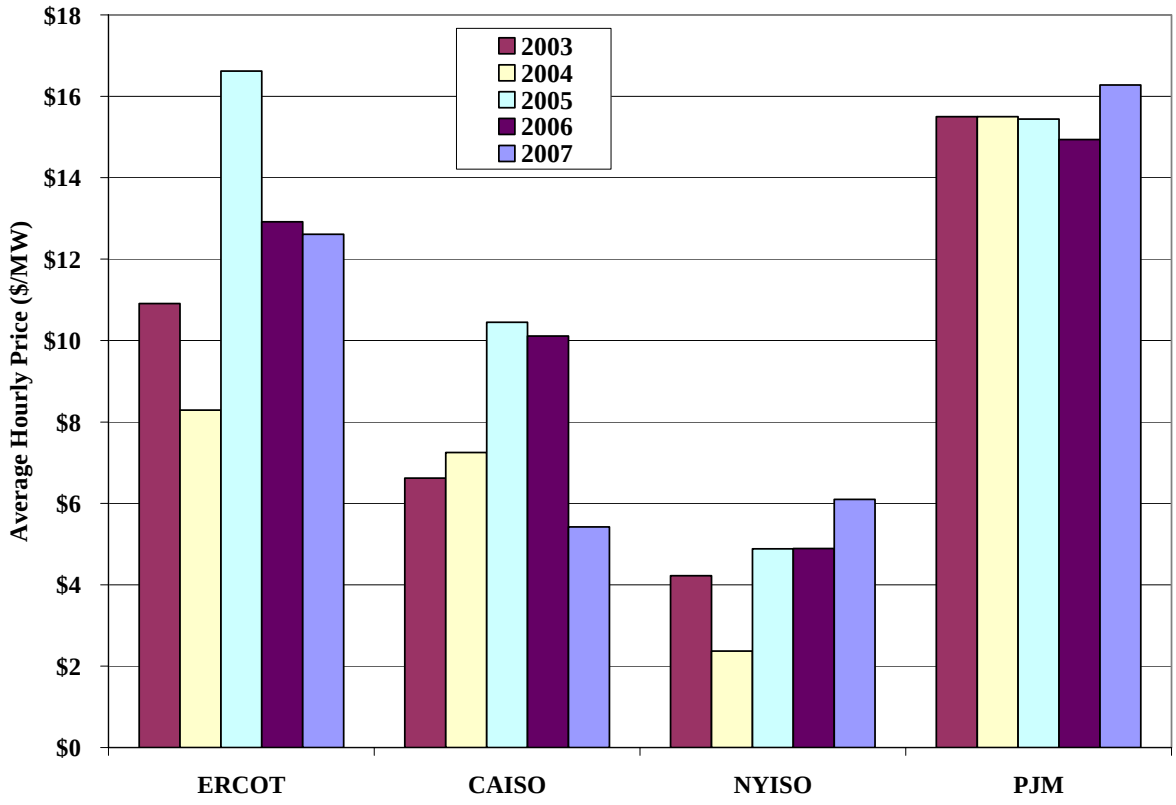
The expected costs of being deployed for reserves are based on the following two factors: (a) the average difference between the resource's production cost and the balancing energy price, and (b) the probability of being deployed. In 2007, about 1.9 percent of the responsive reserves were actually deployed, 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

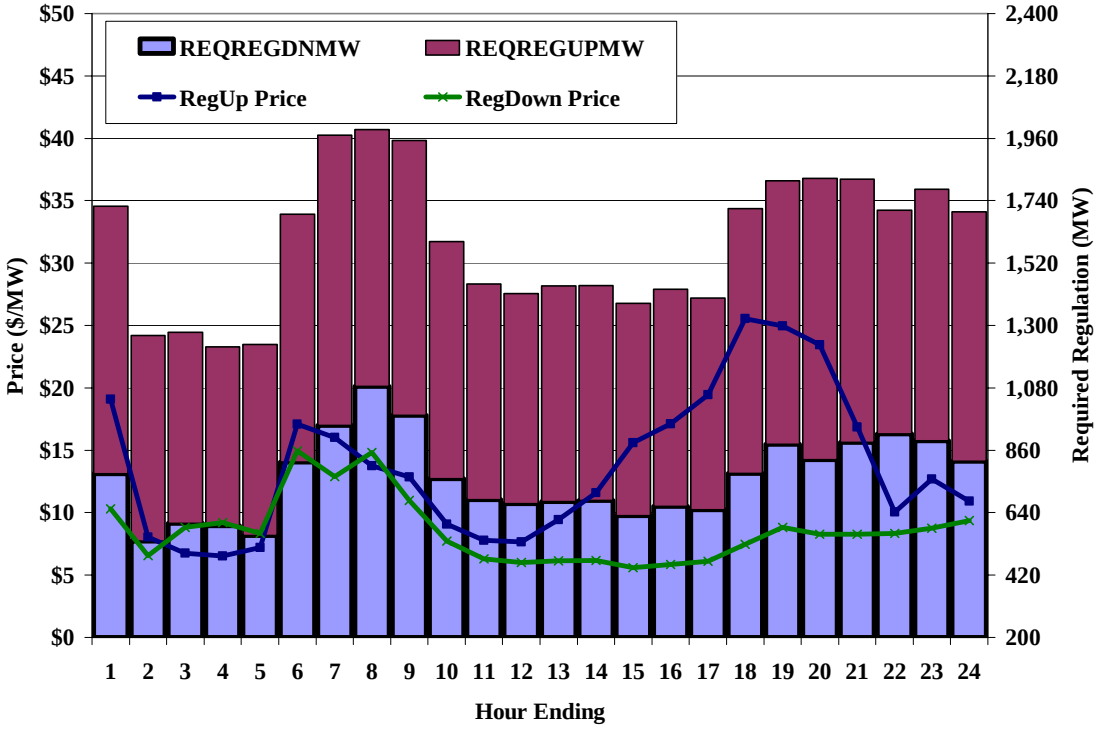


Figure 21 indicates that average regulation prices are generally correlated with the regulation quantity purchased and the typical load pattern in ERCOT. During non-ramping hours, such as overnight and late morning, regulation up and down prices range from $5 to $10 per MW. During the ramping hours in 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

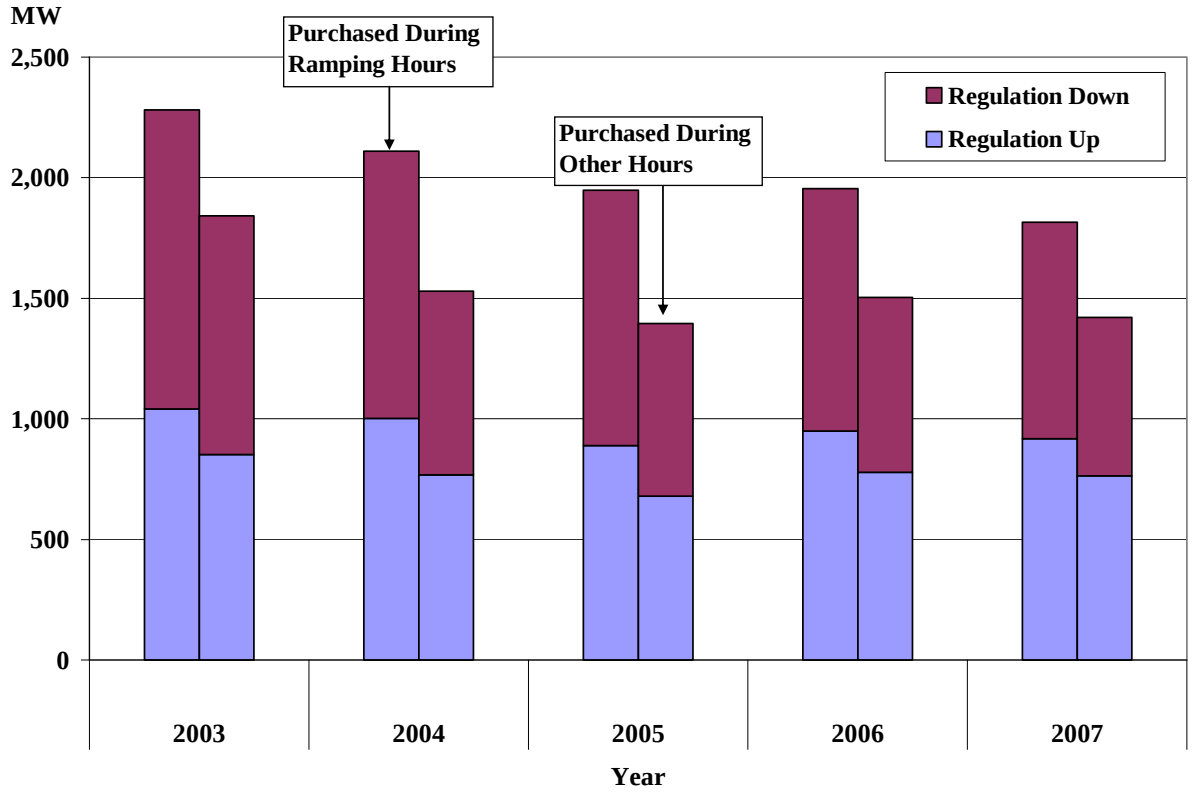


Figure 22 shows that ERCOT has reduced the average regulation quantity scheduled since 2003. The average regulation quantity had steadily declined from 2003-2005, but increased slightly in 2006. In 2007, the average regulation quantities decreased in both the ramping and non-ramping hours compared to 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

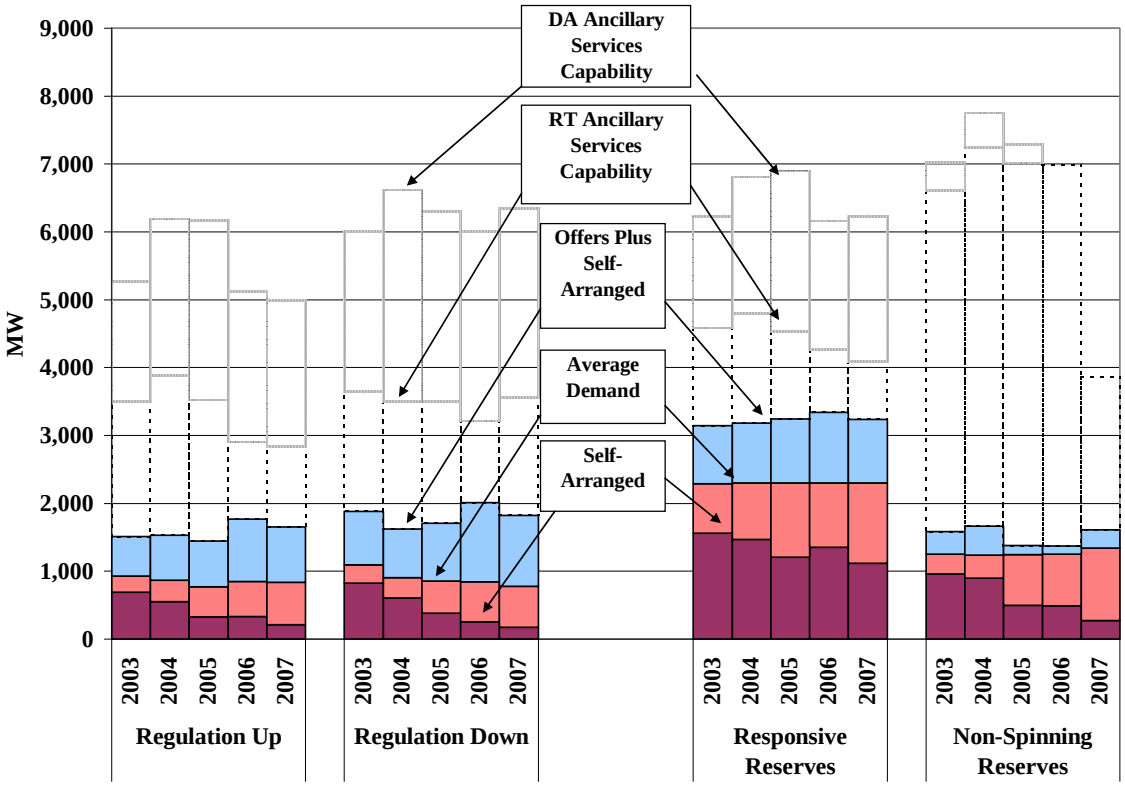
procurement of fixed quantities over 4 to 5 blocks of hours in each day. The result of this change has been a relative decrease in regulation quantities procured in many hours of each day, with an increase in some hours when regulation demand is the highest. Overall change in the procurement methodology 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.

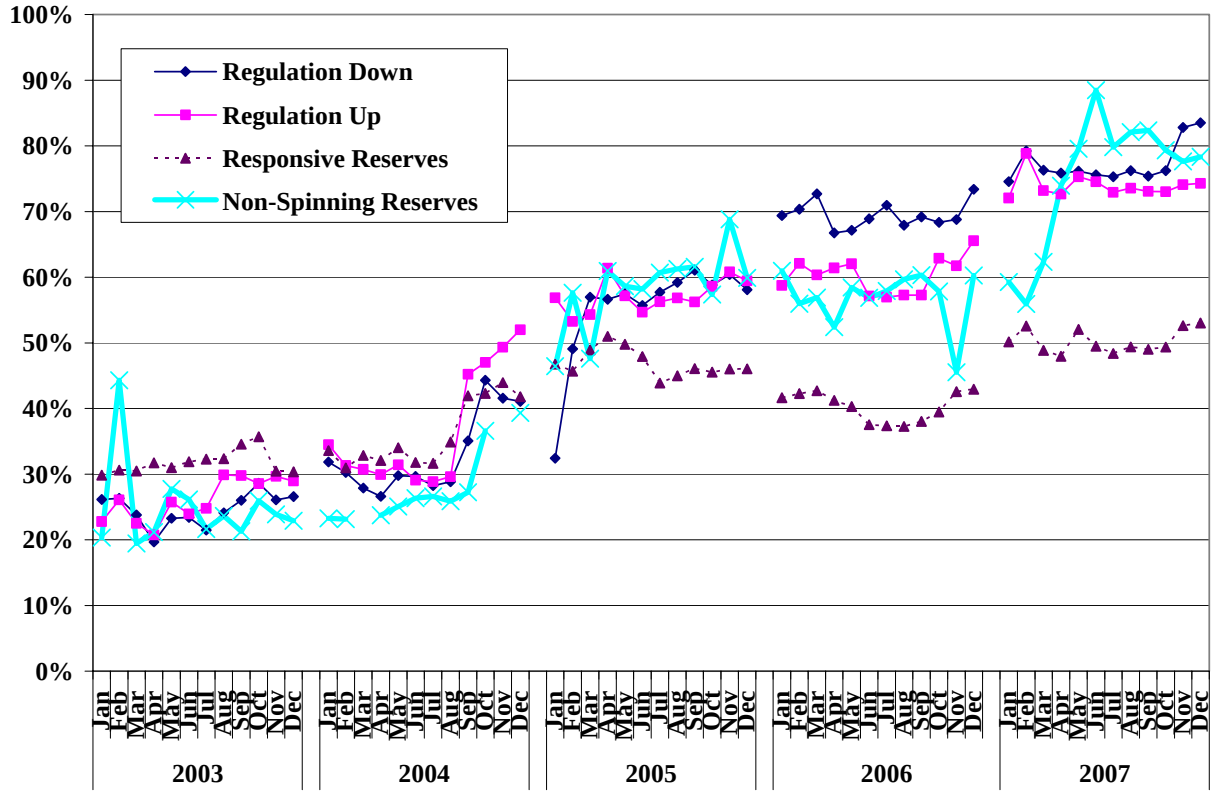
In addition, participants may not offer the capability of resources they do not expect to commit for the following day. Suppliers could submit offer prices high enough to ensure that their costs of committing additional resources to support the ancillary services offers are covered. However, under the 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



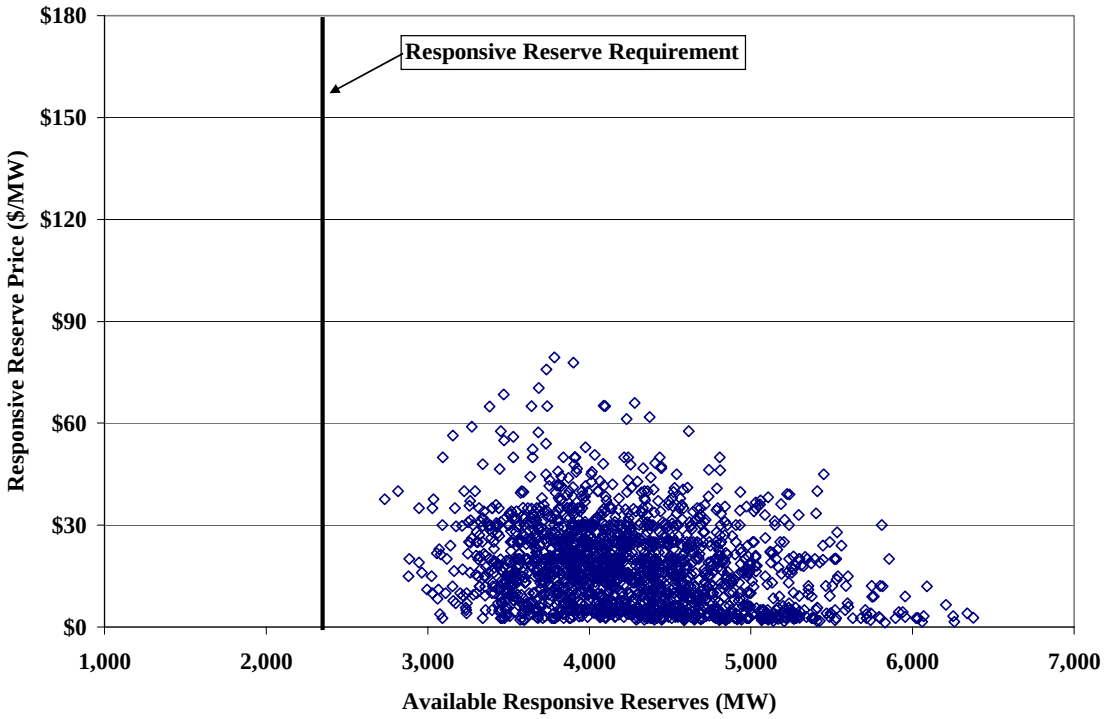
This figure shows that purchases of all ancillary services from the ERCOT markets have generally increased over time, although the purchases of responsive reserve from the ERCOT market have dropped slightly in 2006 (*i.e.*, the quantity of self-arranged responsive reserve has increased slightly). As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

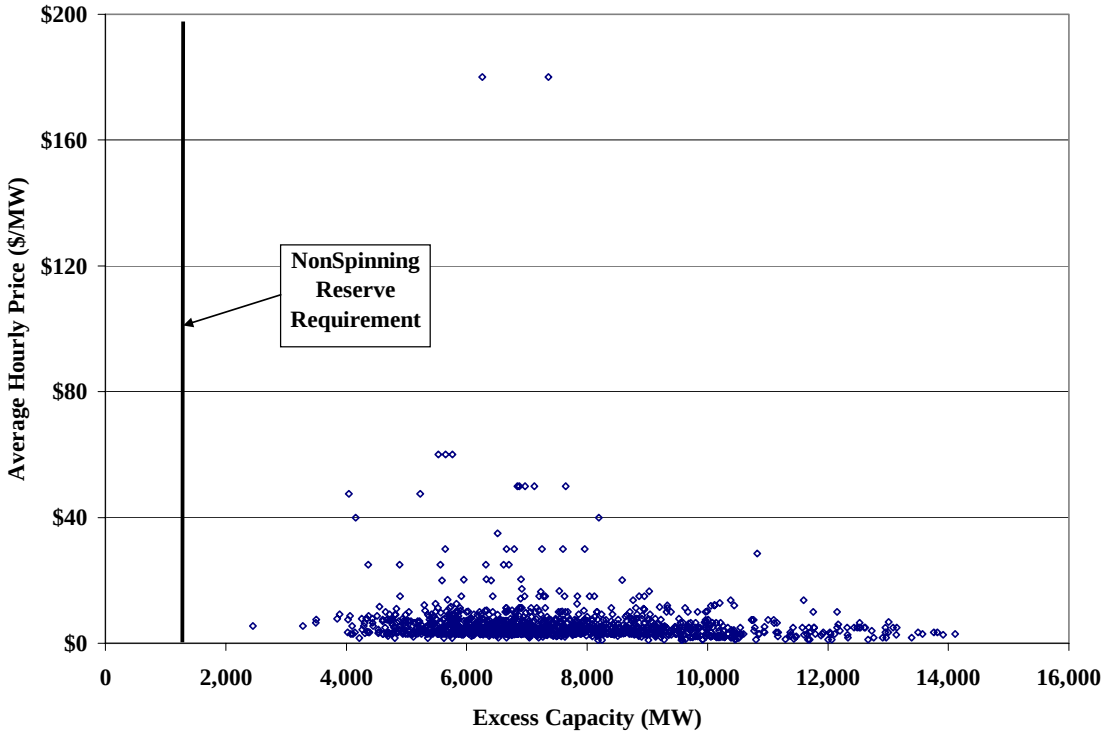


Compared to prior years, this figure indicates a much stronger relationship between the hourly available responsive reserves capability in real time and the responsive reserves prices. In a well

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:

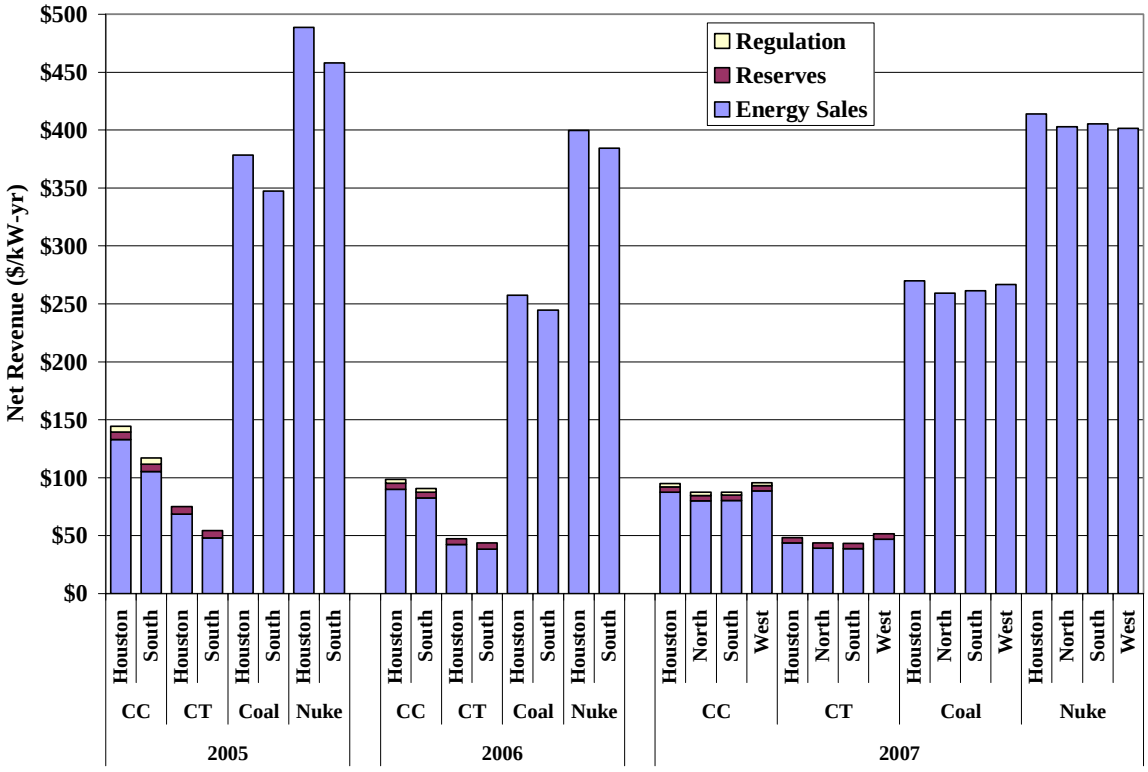
- New capacity is not needed because there is sufficient generation already available;
- Load levels, and thus energy prices, are temporarily low due to mild weather or economic conditions; or
- Market rules are causing revenues to be reduced inefficiently.

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

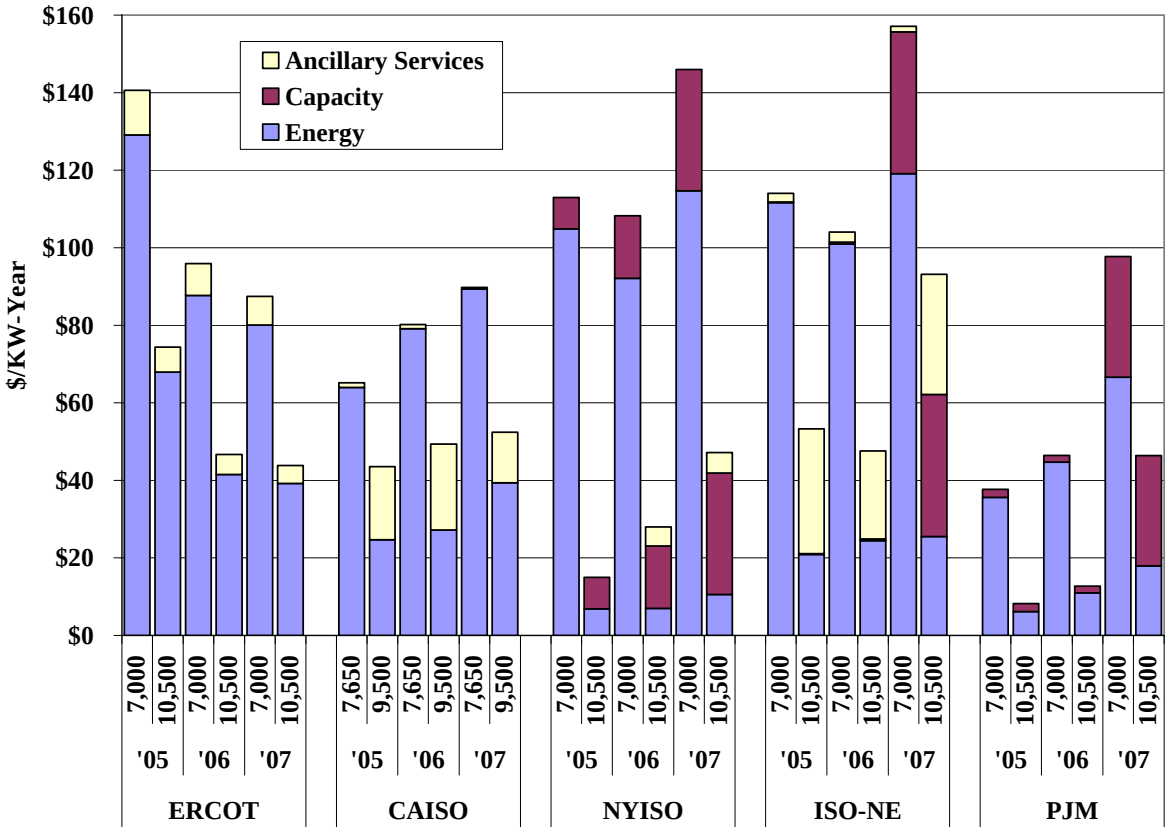


Figure 28 shows that net revenues increased in California, New York, New England and PJM from 2005 to 2007, and decreased in ERCOT. ERCOT is much more dependent on natural gas than the other markets. The decrease in natural gas prices in some of the other regions over this period does not translate as 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.

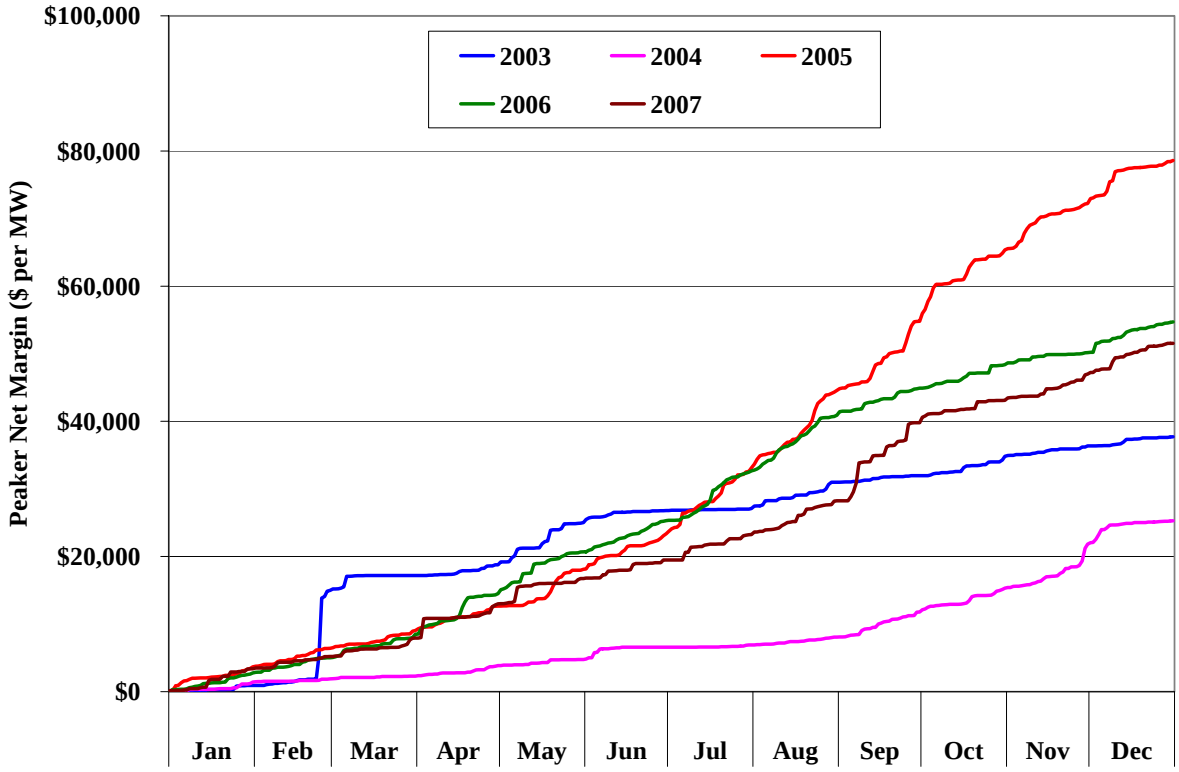
Unlike markets with a long-term capacity market where fixed capacity payments are made to resources across the entire year regardless of the relationship of supply and demand, the objective of the energy-only market design is to allow energy prices to rise significantly higher during legitimate shortage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

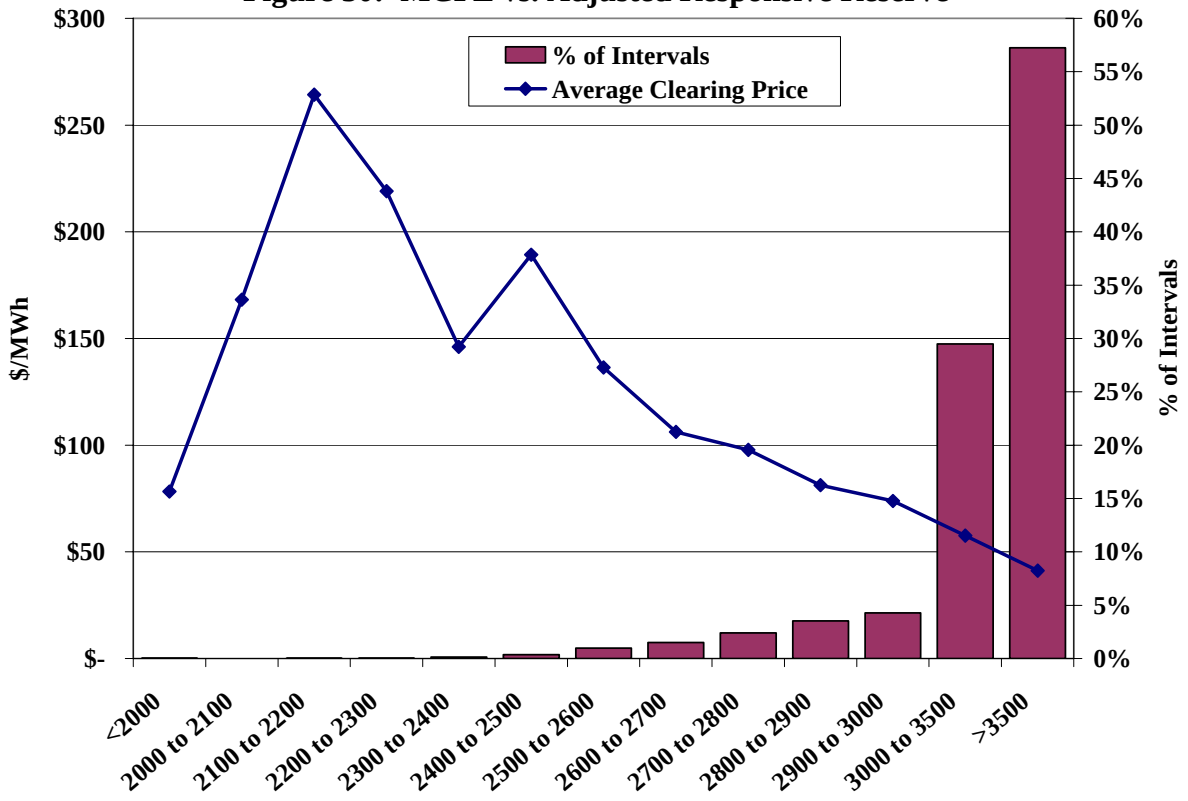
- A strong positive bias in ERCOT's day-ahead load forecast that tended to regularly commit online resources in excess of the quantity required to meet expected demand and operating reserve requirements.

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.

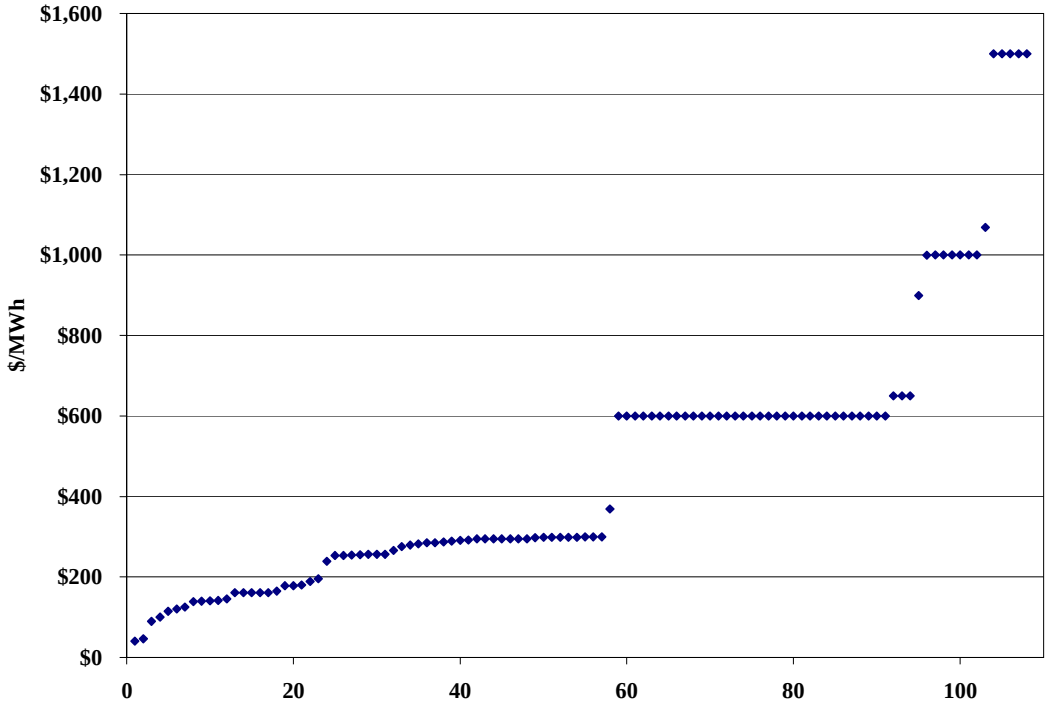
Efforts in 2007 to address these inefficiencies led to an interim measure that was implemented in January 2008 that increased the procurement of responsive reserves to offset the effect of the application of the discount factor, thereby significantly reducing the "overlap" between market and reliability 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

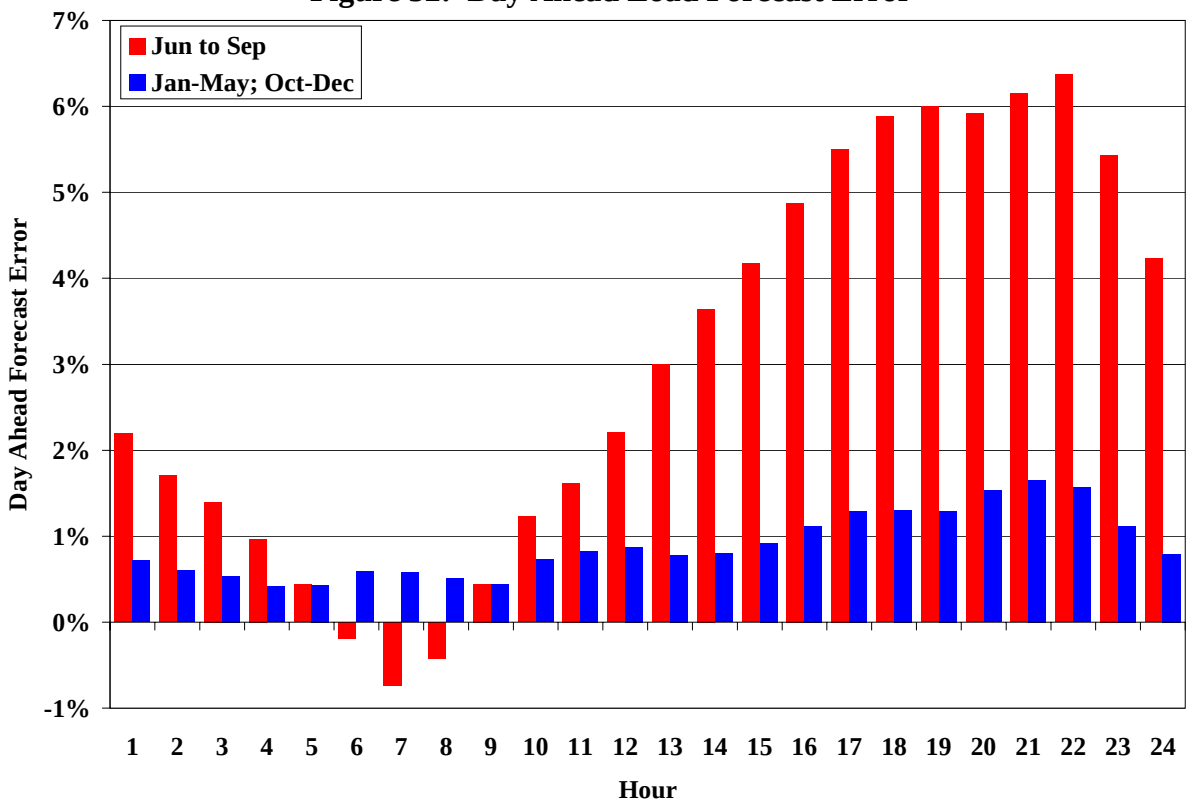


Figure 32 shows the average hourly day-ahead load forecast error for the summer months of June through September, and also for the months of January through May and October through December. In this figure, positive values indicate a day-ahead load forecast that was greater than the actual real-time 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

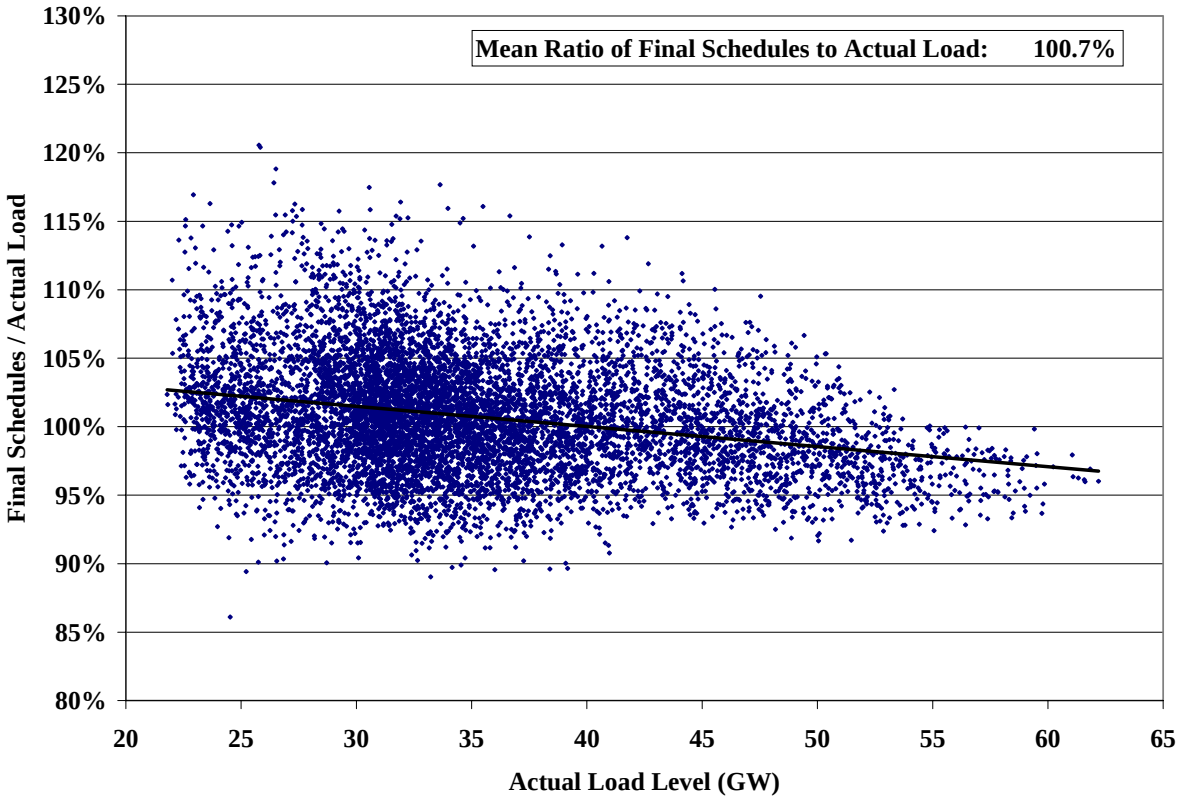


Figure 33 shows that final load schedules generally come very close to actual load in the aggregate, as indicated by an average ratio of the final load schedules to actual load of 100.7 percent. However, the figure also includes a trend line indicating that the ratio of final load schedules to actual 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

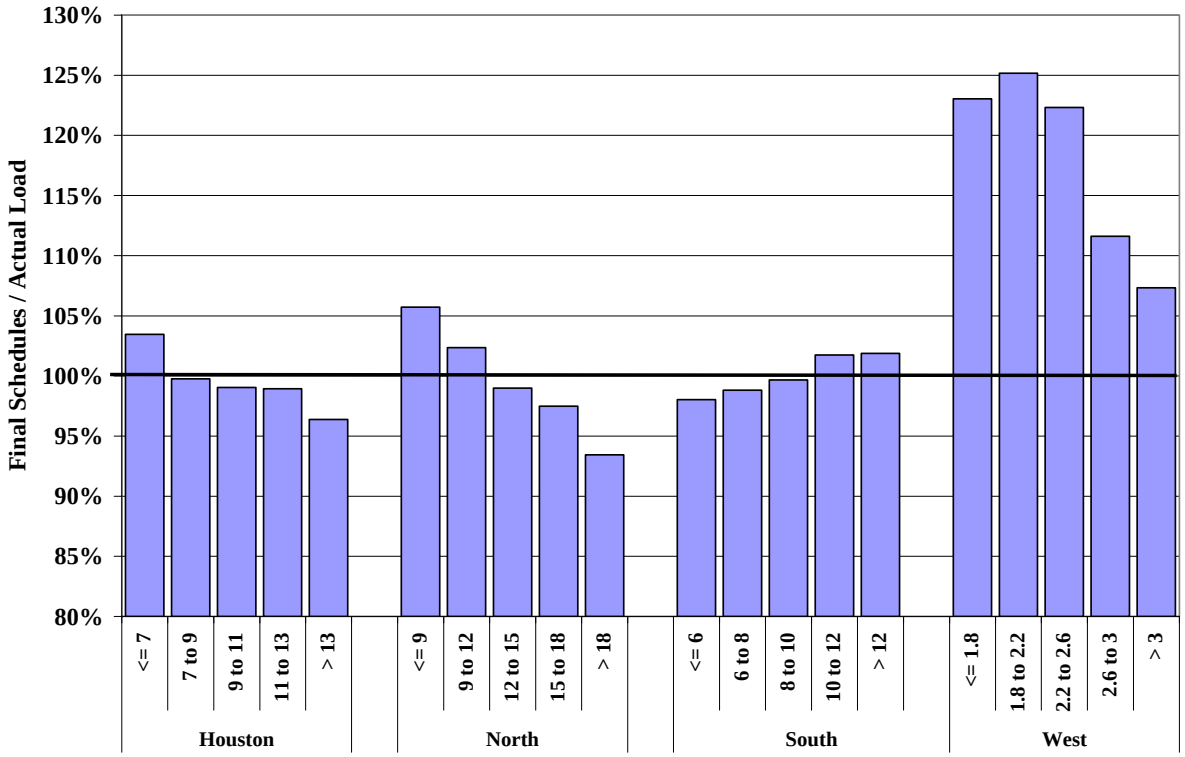


Figure 34 is a further analysis of final load schedules that shows the ratio of final load schedules to actual load evaluated at five different load levels for each of the ERCOT zones.

Figure 34 shows that:

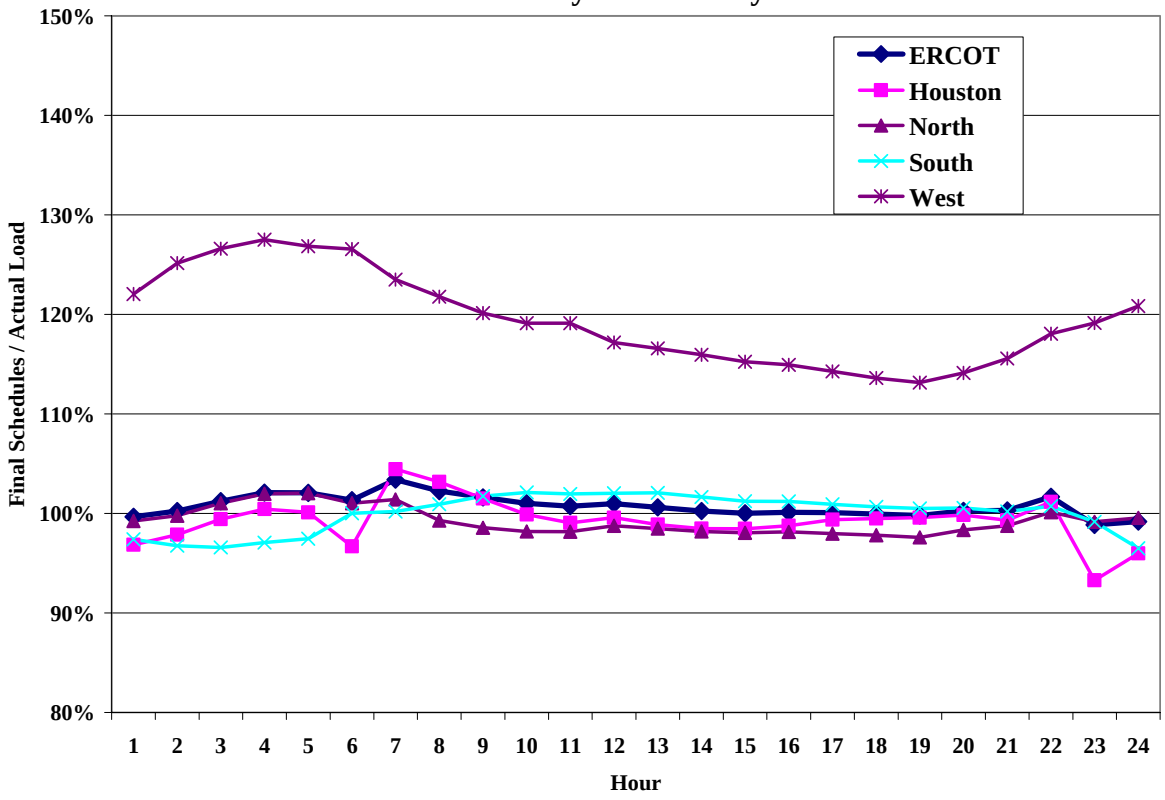
- The final schedule quantity decreases in three of the four zones as actual load increases. In contrast, the schedules in the South zone increase slightly as actual load increases.
- The West Zone is generally over-scheduled, although the ratios decline as load increases.
- Houston is under-scheduled 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.

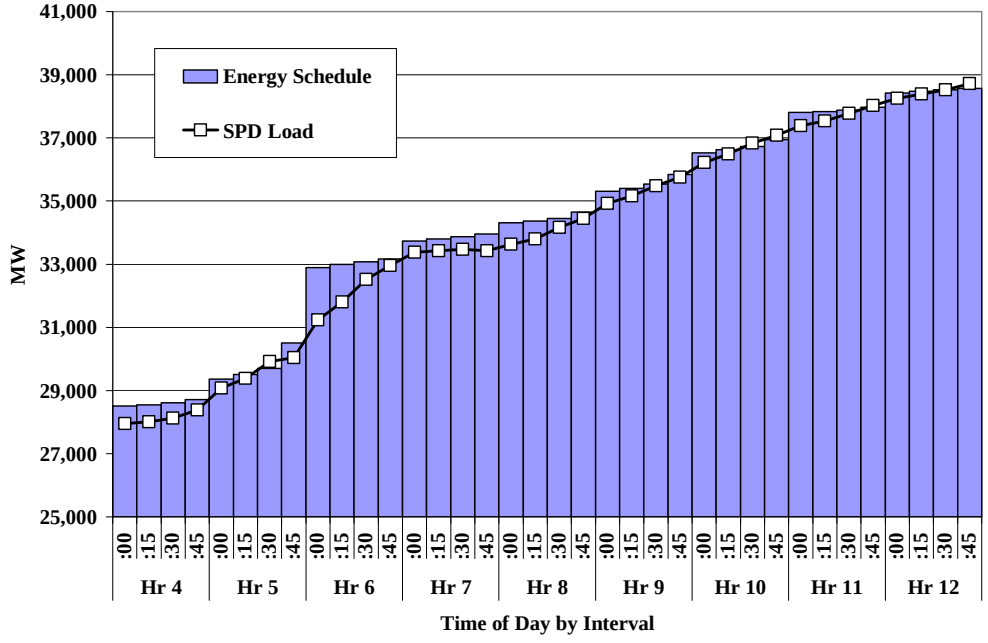
B. Balancing Energy Market Scheduling

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.

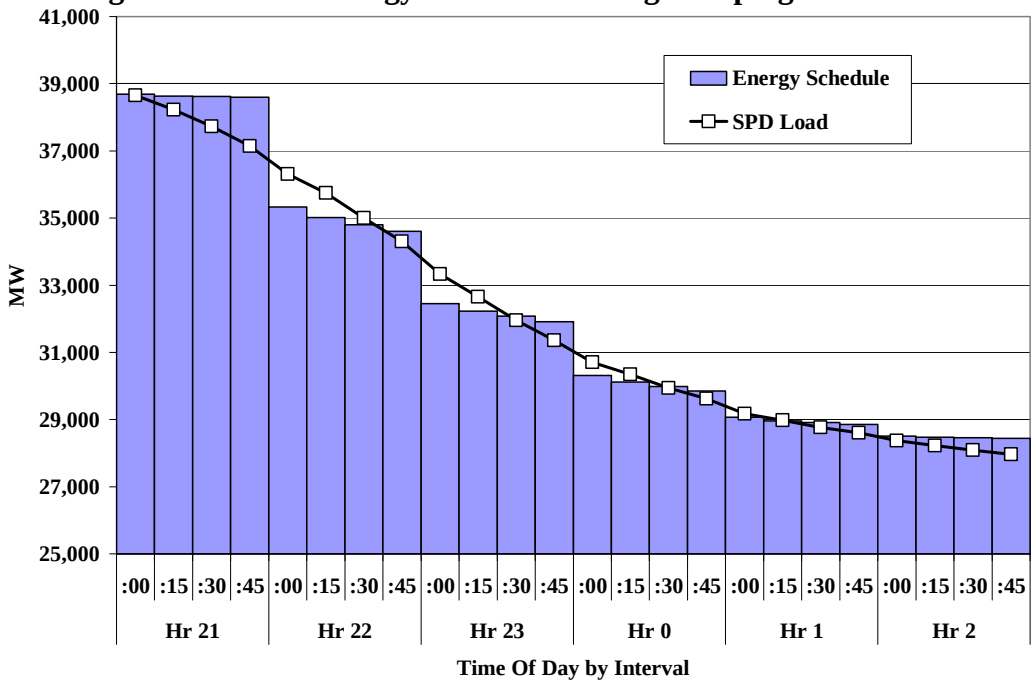
In general for ERCOT as a whole, energy schedules that are less than the actual load result in balancing energy purchases while energy schedules higher than actual load result in balancing energy sales. On average, load increases from approximately 28 GW to almost 39 GW in the nine hours shown in Figure 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.

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. Deviations between the energy schedules and load scheduled by SPD will result in purchases or sales in the balancing energy market. Specifically, net balancing up energy equals SPD load 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 36 and Figure 37). This analysis is similar to that shown in Figure 17 and Figure 18, except instead of showing balancing energy prices relative to load, we show balancing energy prices relative to balancing energy deployments. Figure 38 shows the analysis for the ramping-up hours.

Figure 38: Balancing Energy Prices and Volumes
Ramping-Up Hours

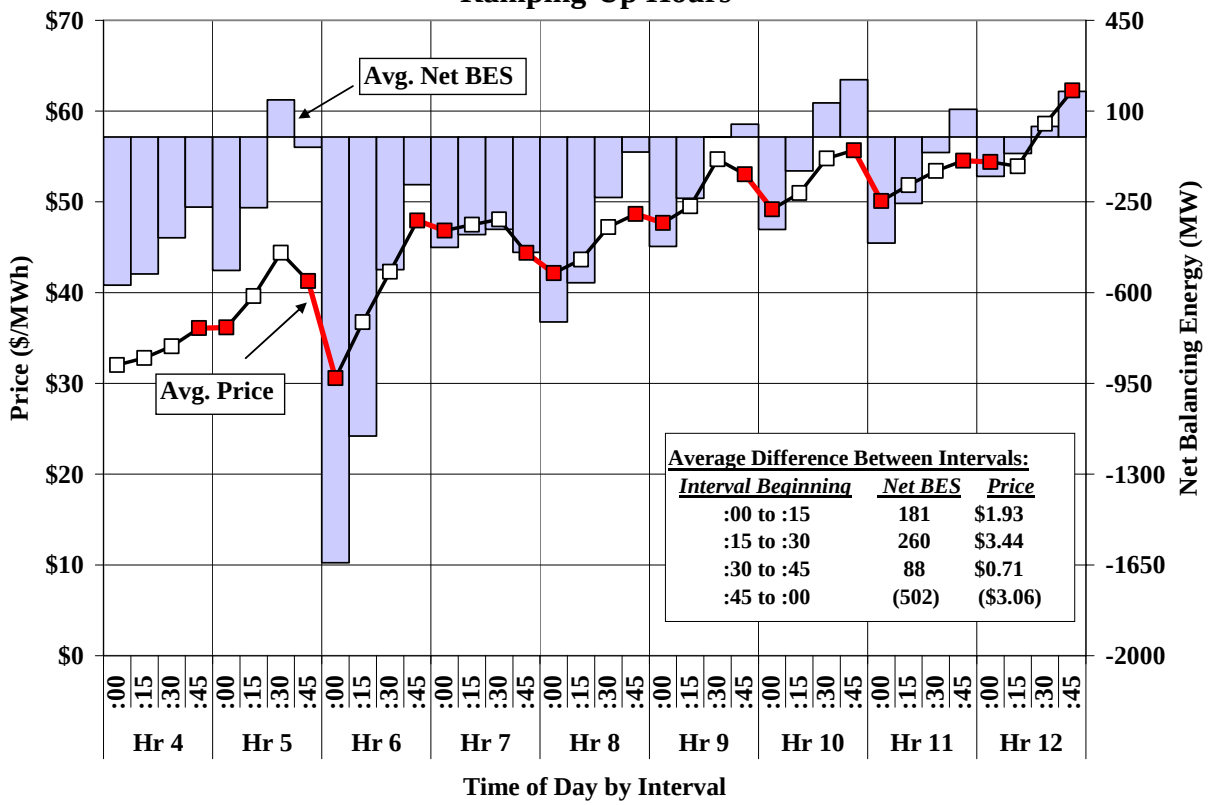
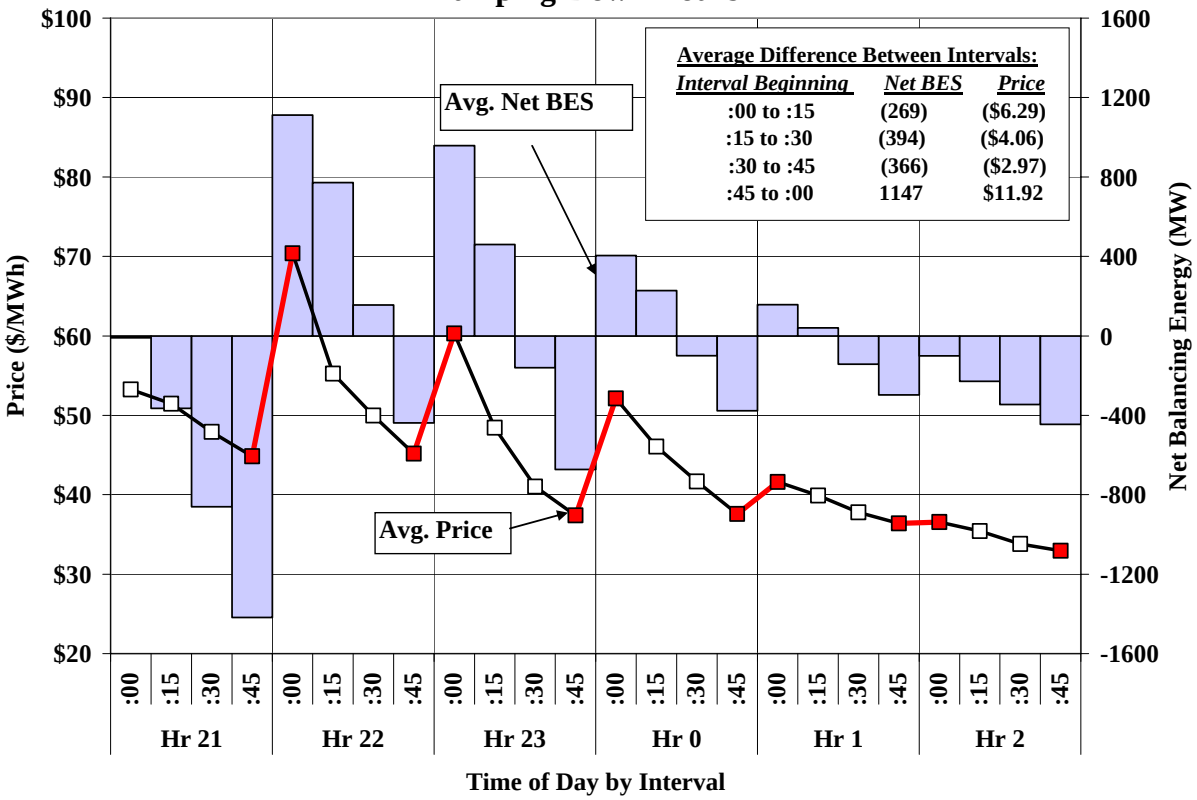


Figure 38 reveals two key aspects of the balancing energy market. First, as discussed above, balancing energy prices are highly correlated with balancing energy deployments. Second, with

the exception of hour 7 and 9, there is a distinct pattern of increasing purchases during the hour. At the beginning of the hour, purchases tend to be smaller than at the end of 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. Whatever the reason for the scheduling patterns that create these balancing deployments, the effect on the ERCOT prices is inefficient. These prices are relatively volatile and could result in erratic dispatch signals to the generators. Figure 39 shows the same analysis for the ramping-down hours. As discussed later in this section, most of these inefficiencies are due to structural issues that are inherent to the zonal market design, and implementation of the nodal market will largely resolve these inefficiencies.

Figure 39: Balancing Energy Prices and Volumes
Ramping-Down Hours



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

While QSEs have the option to submit flexible schedules (i.e., every 15 minutes), 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 38 and Figure 39.

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 has been cited in previous reports, and has continued to be a concern in 2007. To address this issue, we have previously recommended that ERCOT implement an optional capability for QSEs to automatically adjust their hourly balancing energy offers for the changes in their 15-minute schedules. However, because of the resource demands and the timeframe for the nodal transition, such changes will not be accommodated in the zonal market design. 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.

The volatility of the balancing energy prices in each interval is primarily related to the balancing energy deployments. However, as explained in this subsection, this volatility can be exacerbated when the portfolio ramp rates are binding. Portfolio ramp rates are constraints QSEs submit with their balancing energy offers to limit the quantity of balancing up or balancing down 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 unnecessary price spikes (as well as large negative prices). There are three aspects of the current market design that inhibit QSEs from fully utilizing the ramp capability of their portfolio. These are: (1) portfolio ramp rates; (2) portfolio level rather than unit level dispatch; and (3) lack of coordination between energy schedules and ramping. These issues were discussed in detail in the 2005 SOM Report.[21] The operational implications associated with these issues continued in 2007 and will likely continue until the current zonal market design is replaced. However, each of these issues will be significantly ameliorated or eliminated with the implementation of the nodal market.

C. Balancing Energy Market Offer Patterns

In this section, we evaluate balancing energy offer patterns by analyzing the rate at which capacity is offered.[22] Figure 40 shows the average amount of capacity offered to supply balancing up service relative to all available capacity.

[21] 2005 SOM Report at 68-76.

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

Figure 40: Balancing Energy Offers Compared to Total Available Capacity
Daily Peak Load Hours

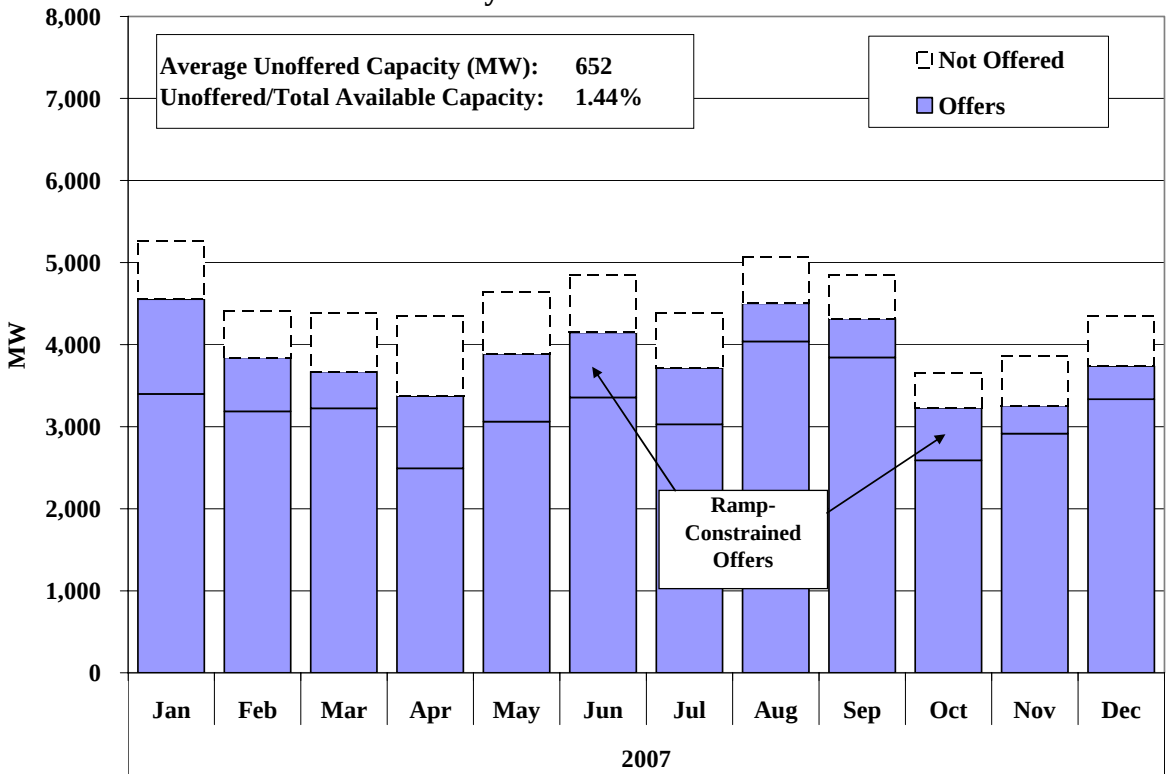
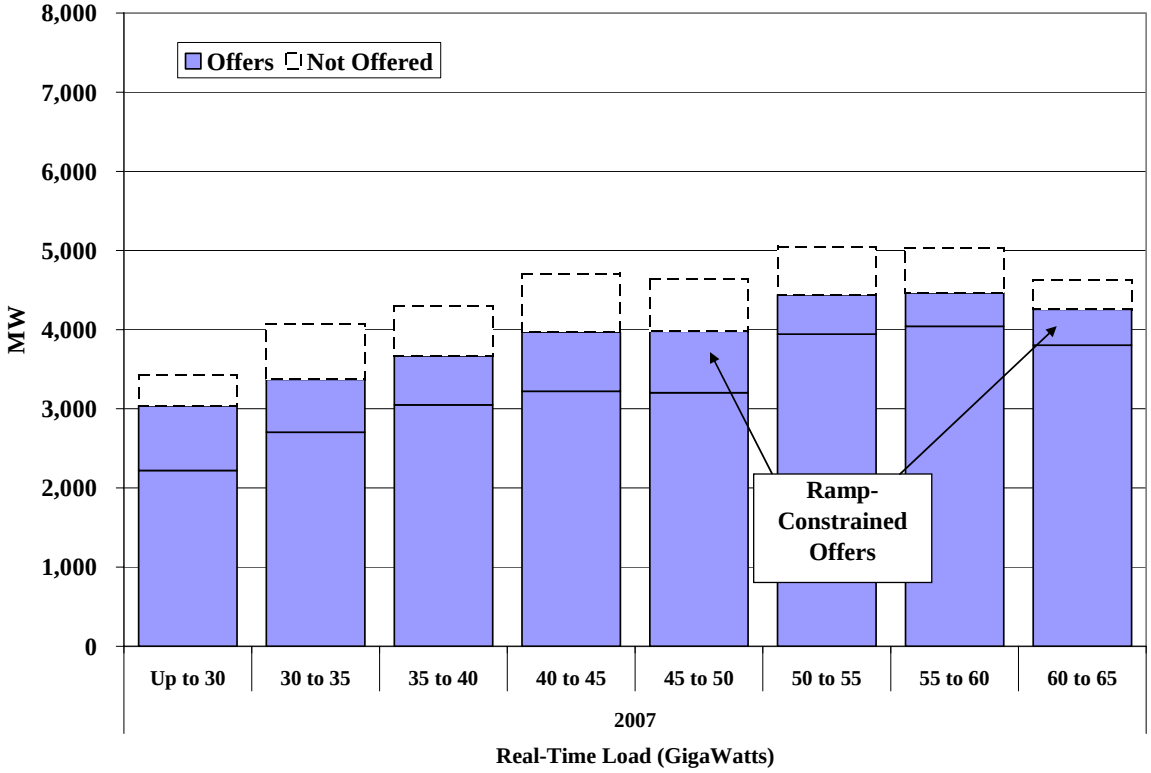


Figure 40 shows only slight variation in 2007 over time in quantities of energy available and offered to the balancing energy market. Up balancing offers are divided into the portion that is capable of being deployed in one interval and the portion which would take longer due to portfolio ramp rate offered by the QSE (*i.e.*, "Ramp-Constrained Offers").

Un-offered energy can raise competitive concerns to the extent that it reflects withholding by a dominant supplier that is attempting to exercise market power. To investigate whether this has occurred, Figure 41 shows the same data as the previous figure, but arranged by load level for daily peak hours in 2007. Because prices are most sensitive to withholding under the tight conditions that occur when load is relatively high, increases in the un-offered capacity at high load levels would raise competitive concerns.

Figure 41: Balancing Energy Offers Compared to Total Available Capacity
Daily Peak Load Hours



The figure indicates that in 2007 the average amount of capacity available to the balancing market increased gradually up to 60 GW of load and then declined at higher levels. The decline in balancing energy available at higher load levels is associated with the fact that scheduled generation increases at higher load levels, thereby leaving less residual capacity available to be offered as balancing energy. As indicated in the figure, the quantity of un-offered capacity does not change significantly as load levels increase.

The pattern of un-offered capacity shown in Figure 41 does not raise significant competitive concerns. If the capacity were being strategically withheld from the market, we would expect it to occur under market conditions most susceptible to the exercise of market power. Thus, we would expect more un-offered capacity under higher load conditions. However, the figure shows that portions of the available capacity that are un-offered do not change significantly as load levels increase. Based on this analysis and other analyses in the report at the supplier level, we do not find that the un-offered capacity raises potential competitive concerns.[23]

[23] *See* 2006 SOM Report at 67 for a discussion of the residual un-offered capacity.

III. DEMAND AND RESOURCE ADEQUACY

The prior sections 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 2007 and the existing generating capacity available to satisfy the load and operating reserve requirements.

A. ERCOT Loads in 2007

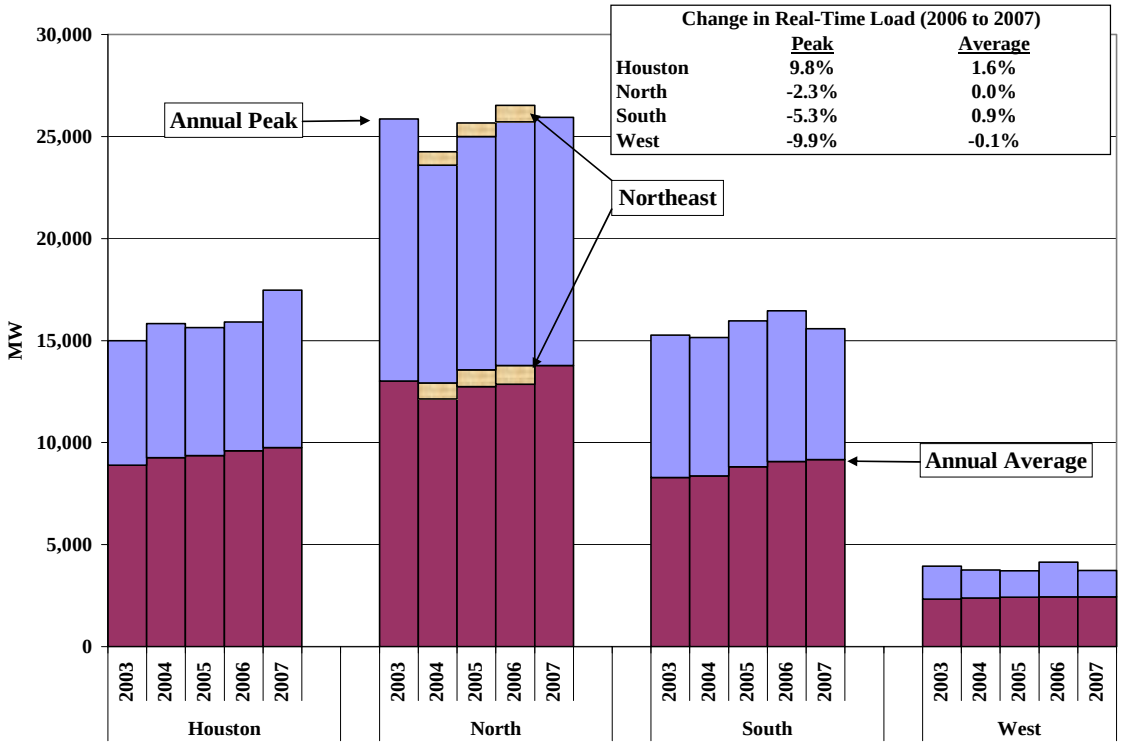
There are two important dimensions of load that should be evaluated separately. First, the changes in overall load levels from year to year can be shown by tracking the changes in average load levels. This metric will tend to capture changes in load over a large portion of the hours during the year. 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 the maintenance of required operating reserves is challenged). Hence, both of these dimensions of load during 2007 are examined in this subsection and summarized in Figure 42.[24]

This figure shows peak load and average load in each of the ERCOT zones from 2003 to 2007. 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); 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 42 shows the annual non-coincident peak load for each zone. This is the highest

[24] The load values in this Section are from ERCOT settlement data. In previous State of the Market Reports, the load values were from ERCOT's Scheduling, Pricing and Dispatch software (including transmission and distribution losses). Data from 2003 to 2006 have also been adjusted.

load that occurred in a particular zone for one hour during the year; however, the peak can occur in different hours for different zones. As a result, the sum of the non-coincident peaks for the zones was greater than the annual ERCOT peak load.

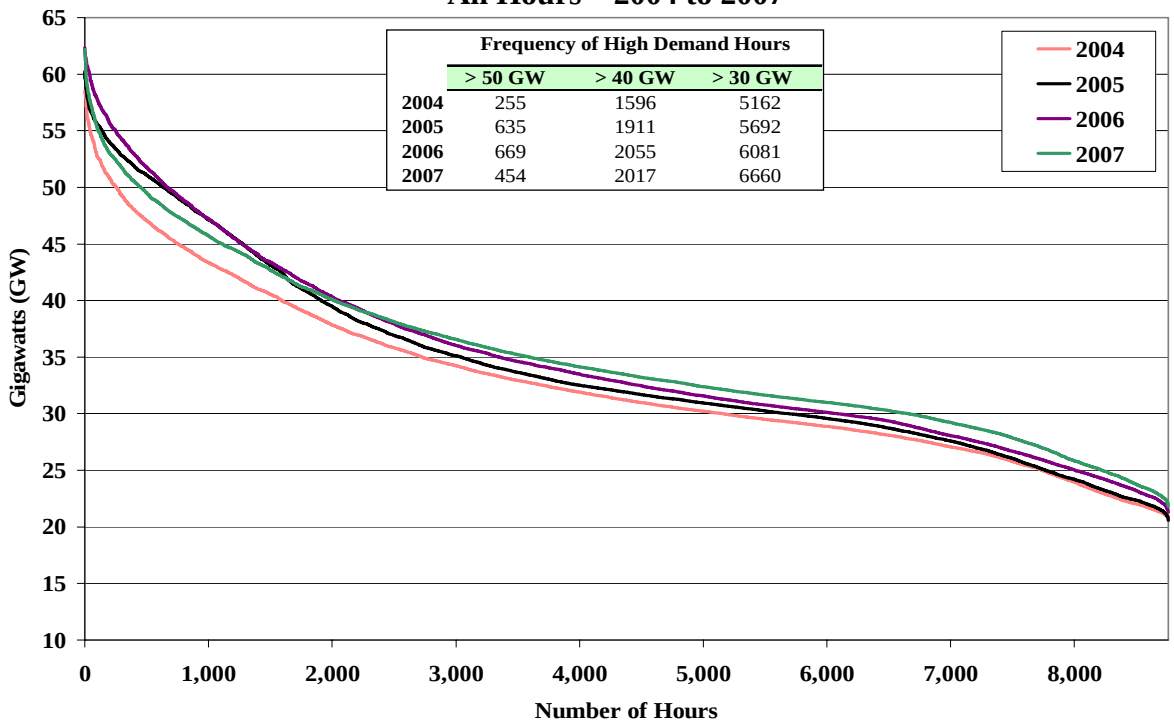
Figure 42: Annual Load Statistics by Zone
2003 to 2007



No load statistics are shown for the Northeast Zone before 2004 because it was separated from the North Zone at the beginning of 2004. For comparison purposes, the Northeast Zone is also shown stacked with the North Zone from 2004 to 2006.

To provide a more detailed analysis of load at the hourly level, Figure 43 compares load duration curves for each year from 2003 to 2007. 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.

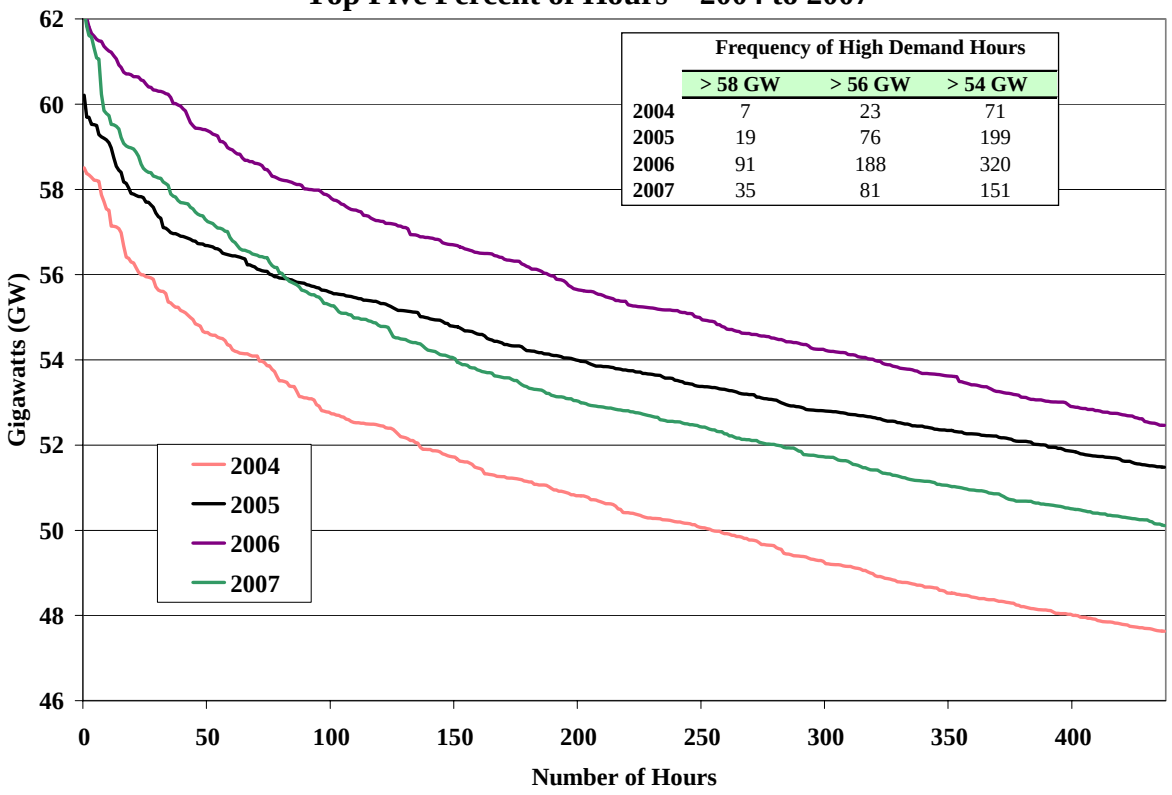
Figure 43: ERCOT Load Duration Curve
All Hours – 2004 to 2007



As shown in Figure 43 , the load duration curve for 2007 lies above the curves for the previous four years at load levels less than 40 GW. Load increased about 0.7 percent from 2006 to 2007. In 2007, there were 10 percent more hours when load exceeded 30 GW than in 2006.

To better show the differences in the highest-demand periods between years, Figure 44 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 2003 to 2006, the frequency of high demand hours in 2007 dropped compared with year 2006. Load exceeded 58 GW in 35 hours in 2007, 91 hours in 2006, 19 hours in 2005, 7 hours in 2003 and 8 hours in 2004. The same pattern prevailed at lower load levels.

Figure 44: ERCOT Load Duration Curve
Top Five Percent of Hours – 2004 to 2007

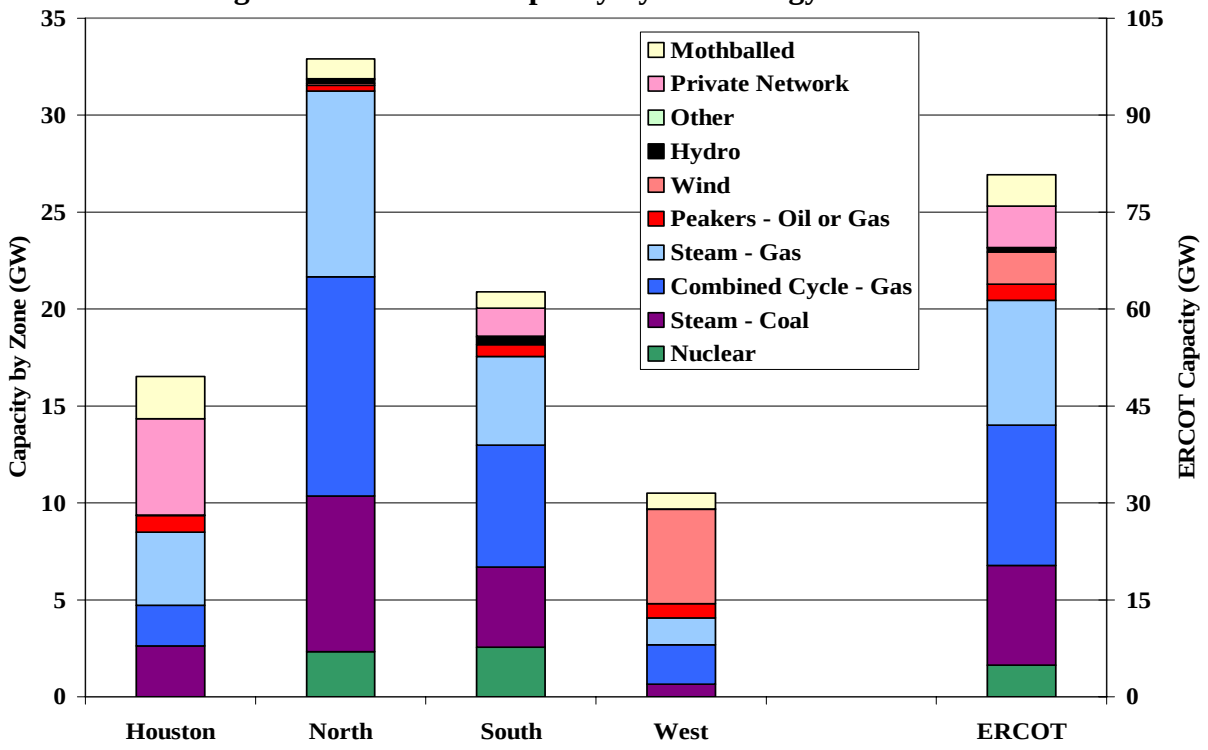


This figure also shows that the peak load in each year was roughly 15 to 25 percent greater than the load at the 95th percentile of hourly load. For instance, in 2006, the peak load value was over 62 GW while the 95th percentile was about 52 GW. This is typical of, and even somewhat flatter than, the load patterns in most electricity markets. This implies that a substantial amount of capacity, more than 10 GW, is needed to supply energy in less than 5 percent of the hours. This 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. 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 45 shows the installed generating capacity by type in each of the ERCOT zones.

Figure 45: Installed Capacity by Technology for each Zone



The nuclear capacity is located in both the North and South Zones, and lignite and coal generation is also a significant contributor in ERCOT. However, the primary fuel in all five zones is natural gas (or sometimes oil) -- accounting for 70 percent of generation capacity in ERCOT as a whole, and 85 percent in the Houston Zone. Much 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 20,000 MW of coal and nuclear generation in ERCOT. Because there are very few hours when ERCOT load drops as low as 20,000 MW, 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 play a much less significant role in setting spot electricity prices.

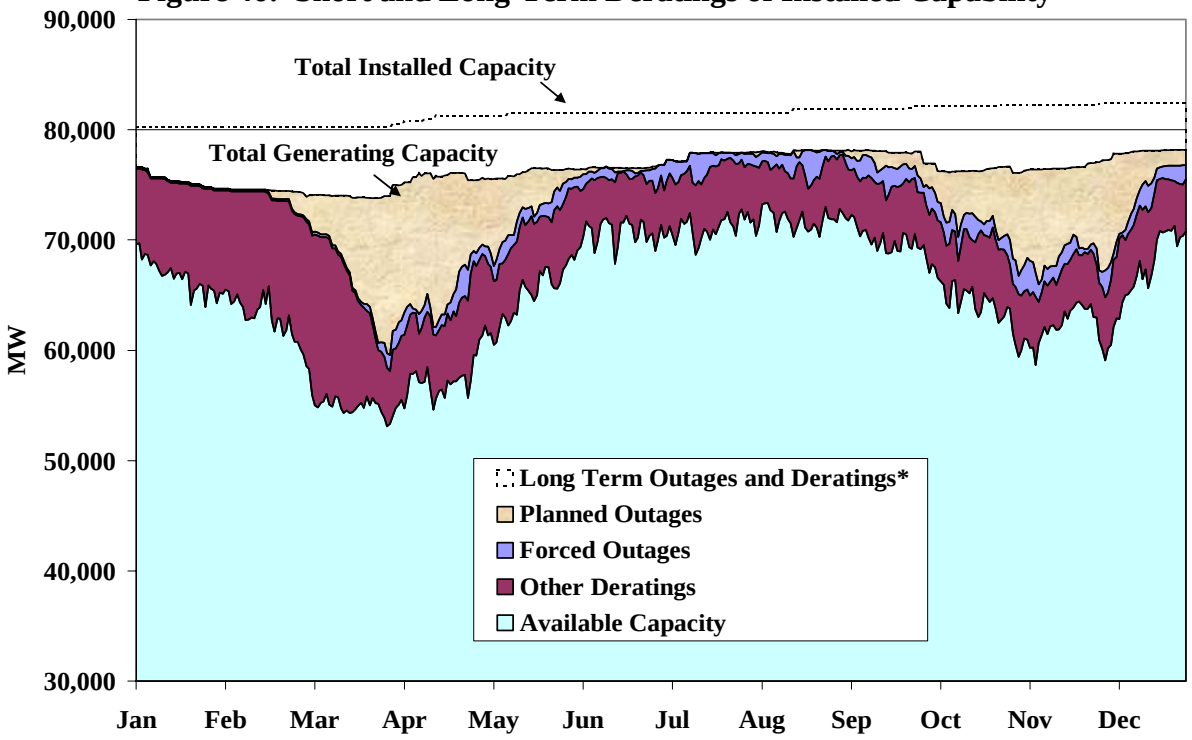
The distribution of capacity among the ERCOT zones is similar to the distribution of demand. This is consistent with the legacy of investment under the regulated vertically integrated utilities when load and resources were largely integrated within separate control areas. The North Zone accounts for 38 percent of capacity, the South Zone 28 percent, the Houston Zone 22 percent, and the West Zone 11 percent. The Houston is typically an importer of power, while the North and South Zones typically export power. Because large amounts of power flow out of the South and the North Zones into the Houston Zone, the South-to-Houston CSC and the North-to-Houston CSC experienced the greatest amounts of congestion during 2007, although transmission lines on the South-to-Houston interface were upgraded in mid-2007 which greatly reduced the congestion on this interface.

1. Generation Outages and Deratings

Figure 45 in the prior subsection shows that installed capacity far exceeds the annual peak load plus ancillary services requirements in ERCOT. This might suggest that the adequacy of resources is not a concern in ERCOT in the near-term, although 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 can be fully derated (rating equals 0) due to a forced or planned outage. However, it is very common for generators to be partially derated (*e.g.*, by 5 to 10 percent) because the resource cannot achieve its installed capability level due to technical factors or environmental factors (*e.g.*, ambient temperature conditions).

In this subsection, we evaluate long-term and short-term deratings to inform our evaluation of ERCOT capacity levels. Figure 46 below shows a breakdown of total installed capability for ERCOT on a daily basis during 2007. 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 46: 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 46 shows that long-term outages and other deratings fluctuated between 7 and 22 GW. These outages and deratings reduce the effective resource margins in ERCOT from the levels reported above. Most of these deratings reflect:

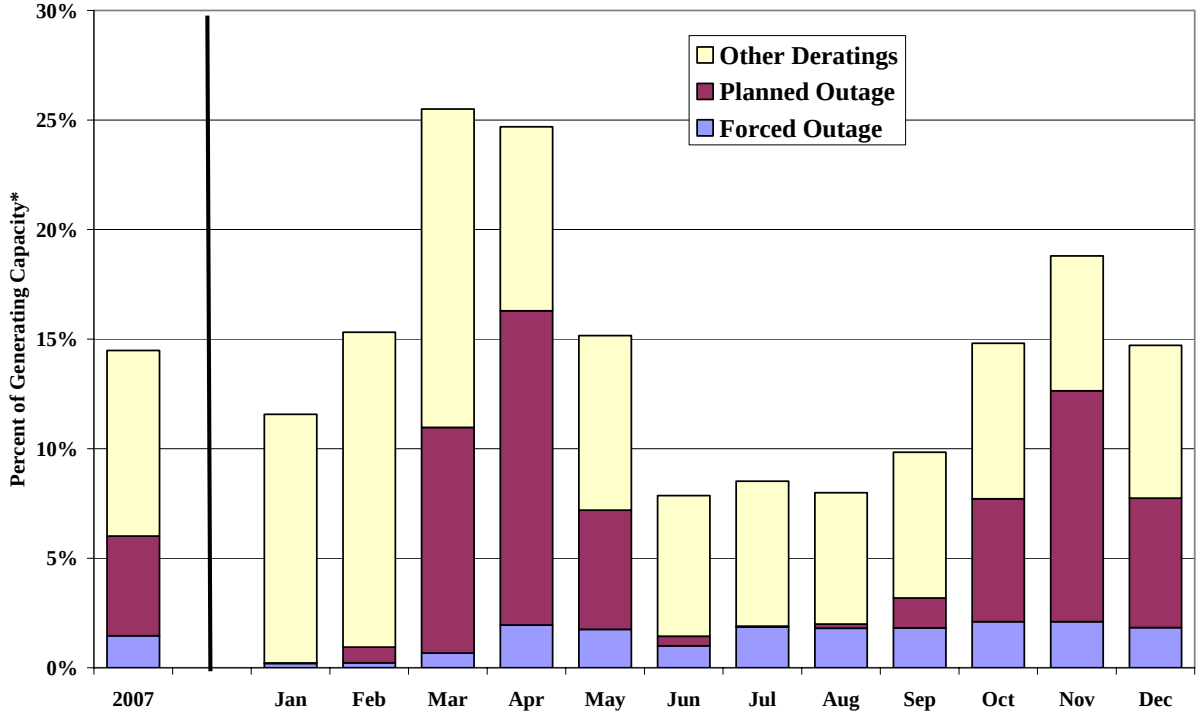
- Cogeneration resources unavailable to serve market load because they are being used to serve self-serve load;
- Resources out-of-service for economic reasons (*e.g.*, mothballed units);
- Output ranges on available generating resources that are not capable of producing up to the full installed capability 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, as expected.

The next analysis focuses specifically on the short-term forced outages and other short-term deratings. Figure 47 shows the average magnitude of the outages and deratings lasting less than 60 days for the year and for each month during 2007.

Figure 47: Short-Term Outages and Deratings*



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

Figure 47 shows that total short-term deratings and outages were as large as 25 percent of installed capacity in the spring and fall, and dropped below 8 percent for the summer. Most of this fluctuation was due to anticipated planned outages, which ranged as high as 5 to 14 percent of installed capacity during March, April, October, and November. Short-term forced outages occurred more randomly, as would be expected, ranging between 0.2 percent and 2 percent of total capacity on a monthly average basis during 2007. These rates are relatively low in comparison to other operating markets, which can be attributed to a number of factors mentioned 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 47 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 6 percent on average during the summer in 2007 and as high as 14 percent in other months. In conclusion, the patterns of outages do not indicate physical withholding or raise other competitive concerns. However, this issue is analyzed in more detail in Section V 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[25] units minus the demand for energy, responsive reserve, up regulation and non-spinning reserve provided from online capacity or quick-start units. If the goal were to have no excess capacity, ERCOT would have to dispatch quick-start resources each day to meet its energy demand. Normally, however, because it is uneconomic to dispatch quick-start units for energy on most days, additional slow-starting resources with lower production costs are committed.

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

To evaluate the commitment of resources in ERCOT, Figure 48 plots the excess capacity in ERCOT during 2007.The figure shows the excess capacity in only the peak hour of each weekday because largest amount of additional generation commitment usually occurs at the peak hour. Hence, one would expect larger quantities of excess capacity in other hours.

Figure 48: Excess On-Line and Quick Start Capacity During Daily Peaks on Weekdays

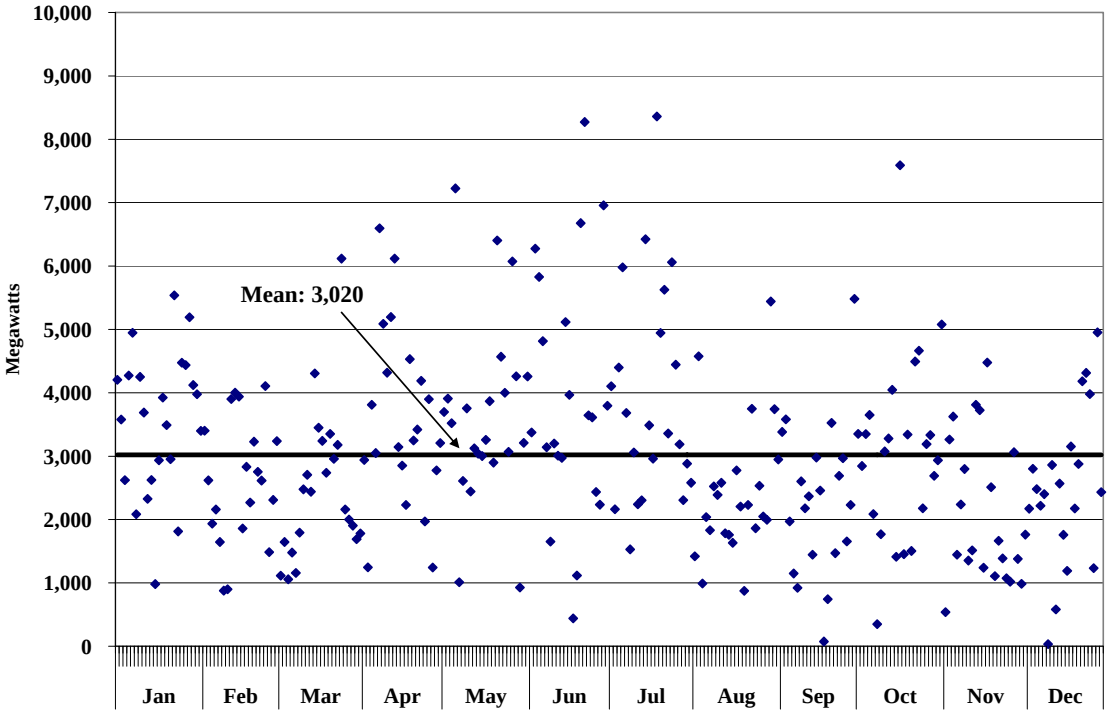


Figure 48 shows that the excess on-line capacity during daily peak hours on weekdays averaged 3,020 MW in 2007, which is approximately 8 percent of the average load in ERCOT. This is at comparable levels as in 2006, with the average daily peak excess on-line capacity being 2,927 MW.

The overall trend in excess on-line capacity also indicates a movement toward more efficient unit commitment across the ERCOT market than 2004 and 2005; however, the current market structure is still based primarily upon a decentralized unit commitment process whereby each participant makes independent generator commitment decisions that are not likely to be optimal. Further contributing to the suboptimal results of the current unit commitment process is that the decentralized unit commitment is comprised of non-binding resource plans that form the basis for ERCOT's day-ahead planning decisions. However, these non-binding plans can be modified

by market participants after ERCOT's day ahead planning process has concluded causing ERCOT to take additional actions that may be more costly and less efficient. Hence, the introduction of a day-ahead energy market with centralized Security Constrained Unit Commitment ("SCUC") that is financially binding under the nodal market design promises substantial efficiency improvements in the commitment of generating resources.

C. Demand Response Capability

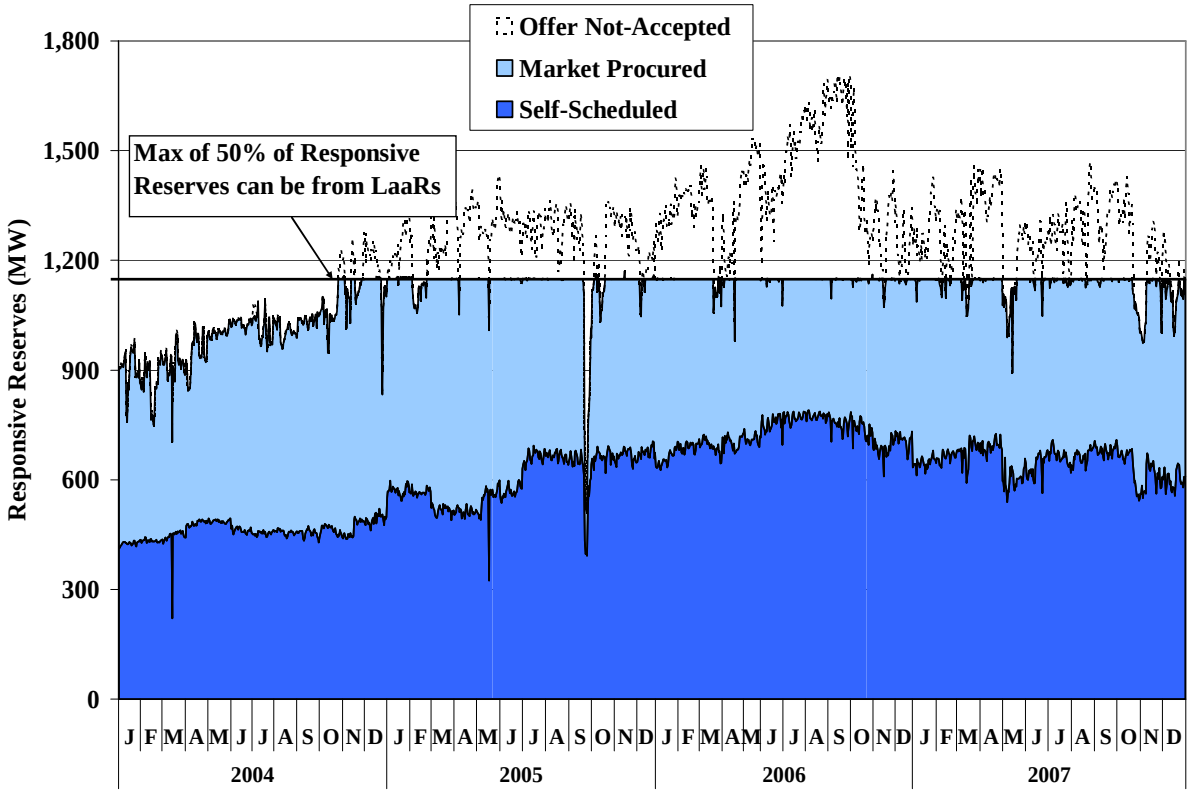
Demand response is a term that broadly refers to actions that can be taken by end users of electricity to reduce load in response to instructions from ERCOT or in response to certain market or system conditions. The ERCOT market allows participants with demand-response capability to provide energy and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

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

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

As of December 2007, around 2,050 MW of capability were qualified as LaaRs. These resources regularly provided reserves in the responsive reserves market, but never participated in the balancing energy market and only a very small portion participated in the non-spinning reserves market. Figure 49 shows the amount of responsive reserves provided from LaaRs on a daily basis in 2007.

Figure 49: Provision of Responsive Reserves by LaaRs
Daily Average



The high level of participation by demand response sets ERCOT apart from other operating electricity markets. Figure 49 shows that the amount of responsive reserves provided by LaaRs gradually increased from about 900 MW at the beginning of 2004 to an average of 1,256 MW in 2007. The majority of this increase was procured through self-provision and bilateral agreements rather than the ERCOT administered auction. In 2007, LaaRs are permitted to supply up to 1,150 MW of the responsive reserves requirement. In 2005 and 2006, it became commonplace for the 1,150 MW restriction to limit the set of demand resources that could provide responsive reserves. This has highlighted a flaw with the way that the ancillary services auction selects demand resources to provide responsive reserves.

The auction ranks responsive reserves providers according to their offer price from lowest to highest.[26] The auction goes up the offer stack until it reaches the 2,300 MW required quantity of

[26] In October 2005, ERCOT began to use a simultaneous clearing model for regulation up, regulation down, responsive reserves, and non-spinning reserves. This selection mechanism is conceptually similar since resources are selected in merit order. However, a resource with a low-priced responsive reserves offer may

reserves. However, if the auction reaches the 1,150 MW limit before meeting the 2,300 MW requirement, the offers of any additional LaaRs cannot be used and are discarded. In such cases, the marginal generator resource sets the clearing price for responsive reserves at a level that exceeds the offer prices of some of the unaccepted offers from LaaRs.

This mechanism for selecting providers and determining clearing prices for responsive reserves is inefficient and leads to excessive reliability costs for consumers. Routinely, the quantity of LaaRs willing to supply responsive reserves at the clearing price exceeds the demand for this service (*i.e.*, 1,150 MW). When supply exceeds demand for a product at the prevailing price, it should cause the price of the product to decrease until the market reaches a level where the supply equals demand. Under the current market design, there is no mechanism for this to happen since there is only one price for all responsive reserves. Since ERCOT limits the amount of responsive reserves that can be provided by LaaRs, the price of reserves provided by LaaRs should clear below the price of reserves provided by synchronized generators.

The design of this market encourages inefficient behavior by QSEs that want to sell responsive reserves from their demand resources. Under current market conditions, the clearing price for responsive reserves is usually set by a generator. To be selected, it is not sufficient for LaaRs to submit an offer price that is below the clearing price. The LaaR's offer must also be included among the lowest priced 1,150 MW of LaaRs. This gives QSEs an incentive to offer LaaRs at arbitrarily low (even negative) prices. Under these incentives, competition does not lead to having the most efficient resources provide responsive reserves. This also raises the concern that a negative LaaR offer could set the responsive reserves clearing price in the event that 1,150 MW of generators are bilaterally scheduled for reserves. In this unlikely event, LaaRs might receive large invoices to provide reserves, raising potential credit issues.

To improve the efficiency of responsive reserve pricing and incentives for suppliers, we recommend that ERCOT determine potentially separate prices for responsive reserves by

---

be selected to provide another product, such as regulation up, if the reduced cost of the other product exceeds the added cost of not using the resource to provide responsive reserves. In this case, the clearing price for responsive reserves is the marginal cost to the system of meeting the reserves requirement. This is always equal to the marginal reserves provider's offer price plus the opportunity cost of not providing an alternate product in the auction.

imposing all supply constraints in the procurement algorithm. The best way to accomplish this would be by having two responsive reserves constraints in the ancillary services auction: (i) that the responsive reserves procurement (including bilateral schedules) be greater than or equal to 2,300 MW and (ii) that the responsive reserves procurement from LaaRs (including bilateral schedules) be less than or equal to 1,150 MW. The clearing price paid to generators would be equal to the shadow price of the first constraint only, while the clearing price paid to LaaRs would be equal to the shadow price of the first constraint minus the shadow price of the second constraint.

Under this proposal, whenever the 1,150 MW limit on LaaRs providing responsive reserves was binding, the clearing price for responsive reserves from LaaRs would be determined by the offer of the marginal LaaR. Whenever the 1,150 MW limit did not affect the selection of resources (*i.e.*, the shadow price of the second constraint equals $0), the clearing prices would be identical for both types of responsive reserves providers. This recommendation would likely require some slight changes to the ancillary services market clearing engine software.

ERCOT stakeholders considered this change in 2006 and, due to resource constraints, decided not to implement it in the current market and instead drafted a protocol revision to implement it in the nodal market. However, this protocol revision failed to receive the necessary two-thirds vote at the ERCOT Technical Advisory Committee in 2007; thus, there is currently no plan to implement any of the changes described above for the RRS market. As previously discussed, the current mechanism for selecting providers and determining clearing prices for responsive reserves is inefficient and leads to excessive reliability costs for consumers. Therefore, we recommend that these changes be reconsidered for implementation in the nodal market design.

Although LaaRs are active participants in the responsive reserves market, they did not offer into the balancing energy or regulation services markets and their participation in the non-spinning reserves market was negligible in 2007. This is not surprising because the value of curtailed load tends to be very high, and providing responsive reserves offers substantial revenue with very little probability of being deployed. In contrast, providing non-spinning reserves introduces a much higher probability of being curtailed. Participation in the regulation services market requires technical abilities that most LaaRs cannot meet at this point. Hence, most LaaRs will

have a strong preference for providing responsive reserves over regulation services, non-spinning reserves, or balancing energy.

IV. TRANSMISSION AND CONGESTION

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

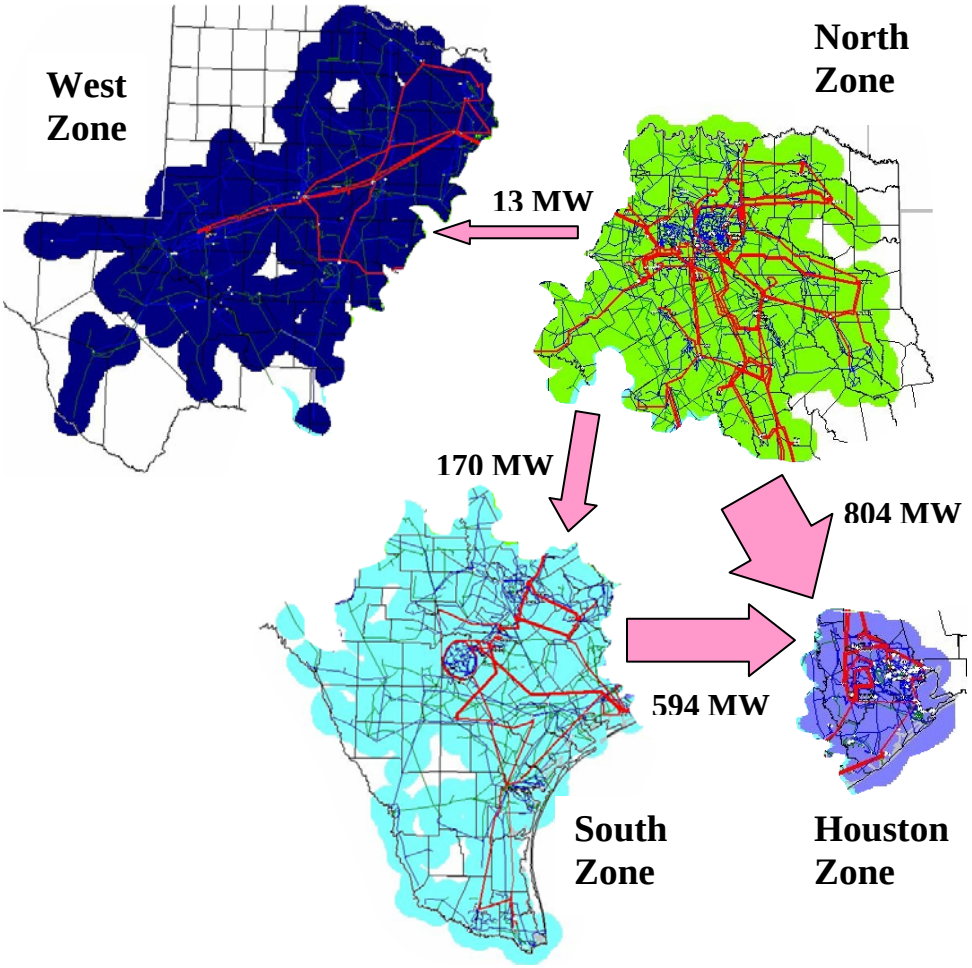
A. Electricity Flows between Zones

In 2007, there were four commercial pricing zones in ERCOT: (a) the North Zone, (b) the West Zone, (c) the South Zone, and (d) the Houston Zone. From year-to-year, slight adjustments are sometimes made to the boundaries of the commercial pricing zones, but the vast majority of customers remained in the same zone from 2006 to 2007. ERCOT operators use the SPD software to economically dispatch balancing energy in each zone to serve load and manage congestion between zones. The SPD model embodies the market rules and requirements documented in the ERCOT protocols.

To manage interzonal congestion, SPD uses a simplified network model with four zone-based locations and five transmission interfaces. These five transmission interfaces, referred to as Commercially Significant Constraints ("CSCs"), are simplified representations of groups of transmission elements. ERCOT operators use planning studies and real-time information to set limits for each CSC that are intended to utilize the total transfer capability of the CSC. In this subsection of the report, we describe the SPD model's simplified representations of flows between zones and analyze actual flows in 2007.

The SPD uses zonal approximations to represent complex interactions between generators, loads, and transmission elements. Because the model flows are based on zonal approximations, the estimated flows can depart significantly from real-time physical flows. Estimated flows that diverge significantly from actual flows are an indication of inaccurate congestion modeling leading to inefficient energy prices and other market costs. This subsection analyzes the impact of SPD transmission flows and constraints on market outcomes. Figure 50 shows the average SPD-modeled flows over CSCs between zones during 2007. A single arrow is shown for the modeled flows of both the North to West and West to North CSCs.

Figure 50: Average SPD-Modeled Flows on Commercially Significant Constraints During All Intervals in 2007



Note: In the figure above, CSC flows are averaged taking the direction into account. So one arrow shows the average flow for the North-to-West CSC was 13 MW, which is equivalent to saying that the average for the West-to-North CSC was *negative* 13 MW.

Figure 50 shows the four ERCOT geographic zones as well as the five CSCs that interconnect the zones: (a) the West to North interface, (b) the South to North interface, (c) the South to Houston interface, (d) the North to Houston interface, and (e) the North to West interface. Based on SPD modeled flows, Houston is a significant importer while the North and South Zones export significant amounts of power.

The most important simplifying assumption underlying the zonal model is that all generators in a zone have the same effect on the flows over the CSC, or the same generation shift factor ("GSF")[27] in relation to the CSC. In reality, the generators within each zone can have widely varying effects on the flows over a CSC. To illustrate this, we calculated flows that would occur over the CSC using actual generation and actual generation shift factors and compared this to flows calculated using actual generation and zonal average shift factors. The flows over the North to West CSC are not shown separately in the table below since they are equal and opposite the flows for the West to North CSC.

Table 2: Average Calculated Flows on Commercially Significant Constraints Zonal-Average vs. Unit-Specific GSFs

CSC 2007	Flows Modeled by SPD (1)	Flows Calculated Using Actual Generation (2)	*Difference = (2) - (1)*	Flows Calculated Using Actual Generation and Unit-specific GSFs (3)	*Difference = (3) - (2)*
West-North	-13	-29	*-15*	-133	*-104*
South-North	-170	-154	*17*	-75	*78*
South-Houston	594	592	*-2*	834	*242*
North-Houston	804	794	*-10*	650	*-144*

The first column in Table 2 shows the average flows over each CSC calculated by SPD. The second column shows the average flows over each CSC calculated using zonal-average GSFs and actual real-time generation in each zone instead of the scheduled energy and balancing energy deployments used as an input in SPD. Although these flows are both calculated using the same zonal-average GSFs, they can differ when the actual generation varies from the SPD

[27] A GSF indicates the portion of the incremental output of a unit that will flow over a particular transmission facility. For example, a GSF of 0.5 would indicate that half of any incremental increase in output from a generator would flow over the interface. Likewise, a GSF of -0.5 would indicate that an incremental increase of 1 MW would reduce the flow over the interface by 0.5 MW.

generation. This difference is shown in the third column (in italics). These differences indicate that the actual generation levels result in higher calculated flows on each CSC except the West to North and North to Houston, where calculated flows are lower.

The fourth column in Table 2 reports the average flows over each CSC calculated using unit-specific GSFs and actual real-time generation. Since the actual generation data used to calculate the flows in this column are identical to those used in column (2), the difference in flows between the two columns can be attributed to using zonal GSFs versus resource-specific GSFs. These differences in flows are shown in the fifth column (in italics). The differences in the last column measure the inaccuracy caused by treating each unit within a particular zone as having identical impact on the CSCs.

These results show that the heterogeneous effects of generators in a zone on the CSC flows can cause the actual flows to differ substantially from the SPD-calculated flows. Table 2 shows that the unit-specific GSFs increased the calculated flows on the South-Houston interface by 242 MW and reduced the calculated flows on the North to Houston CSC by 144 MW. These differences are sizable and are generally larger than the differences that can be attributed to variations in actual generation.

We note that the GSF simplification embedded in the SPD model is important for loads as well. Loads tend to be concentrated within a zone, but the SPD model assumes a generation-weighted average shift factor for all loads in the zone. Using generation-weighted shift factors for load rather than load-weighted shift factors can cause significant differences between SPD flows and actual flows. However, the impact of this assumption is diminished by the fact that loads are not used to manage transmission constraints in real-time. The use of simplified generation-weighted shift factors prevents the SPD model from efficiently assigning the costs of interzonal congestion. In the long run, the use of generation-weighted shift factors for loads systematically biases prices, so that buyers in some zones pay too much, and others pay too little.

To effectively manage interzonal congestion, it is important for SPD to accurately model the major constrained transmission interfaces between zones. In 2007, the five CSCs modeled by SPD did not include all significant interfaces between zones. Sizeable quantities of power were transported on transmission facilities not modeled by SPD as flows on CSCs. Table 3

summarizes the actual net imports into each zone compared to SPD modeled flows from 2003 to 2007.

Table 3: Actual Net Imports vs. SPD-Calculated Flows on CSCs 2003 to 2007

Year	Zone	Actual Net Imports	SPD Flows on CSCs
2003	**Houston**	1,796	565
	North	-507	191
	South	-1,213	-702
	West	-76	-54
2004	**Houston**	2,479	1,265
	North	867	264
	NorthEast	-2,116	-858
	South	-1,531	-800
	West	304	129
2005	**Houston**	2,596	1,247
	North	660	164
	NorthEast	-2,138	-845
	South	-1,501	-728
	West	386	162
2006	**Houston**	3,434	1,744
	North	462	20
	NorthEast	-2,334	-974
	South	-1,741	-870
	West	180	79
2007	**Houston**	3,264	1,398
	North	-2,019	-1,001
	South	-1,319	-764
	West	74	13

Table 3 summarizes the differences between average SPD-calculated flows and average actual flows into each zone. These differences can be attributed to three factors. First, the use of zonal average GSFs, rather than resource-specific GSFs, by SPD to model generators can cause the SPD-calculated flows on a particular CSC to be substantially different from the actual flows.

Second, the use of generation-weighted shift factors to model load causes systematic differences between SPD flows and actual flows. For instance, SPD generally underestimated flows on the South to North CSC because of the difference between load-weighted and generation-weighted shift factors, accounting for a significant portion of the difference between SPD flows and net exports from the South Zone.

Third, significant quantities of power may flow over other transmission facilities that are not defined as part of the CSC. This will tend to cause the actual imports to exceed the SPD-calculated flows over the CSCs. For instance, the South-North interface is made up of the two 345 kV lines connecting the South and North zones, however, ERCOT has defined 19 CREs ("Closely Related Elements") which can also constrain flows from the South Zone to the North Zone. While ERCOT has the discretion to take CREs into account when managing interzonal congestion, they do not have the flexibility to do this efficiently. SPD always uses the CSC shift factors, although shift factors for CREs between the South Zone and North Zone may differ significantly from shift factors for the CSC. This leads to inefficient re-dispatch to manage constrained CREs.

Table 3 shows significant changes in the levels of net imports into each zone between 2003 and 2007. Imports to the Houston zone rose substantially from 2003 to 2004 and remained about the same from 2004 to 2005, followed by a steep increase again in 2006 and then stayed about the same level in 2007.[28] The West Zone shifted from being a net exporter in 2003 to importing substantial quantities from 2004 to 2007, with the average import levels dropping by about 58 percent in 2007 compared to 2006. From 2003 to 2007, net exports increased from the North zone compared with the combined area of the North and Northeast zones from 2004 to 2006. Net exports from the South zone increased from 2003 to 2006, and dropped about 24 percent in 2007. In every case, the SPD-calculated flows on CSCs were significantly less than the actual interchange.

[28] The North to Houston CSC was added in 2004.

B. Interzonal Congestion

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

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

Figure 51: Average SPD-Modeled Flows on Commercially Significant Constraints During Transmission Constrained Intervals in 2007

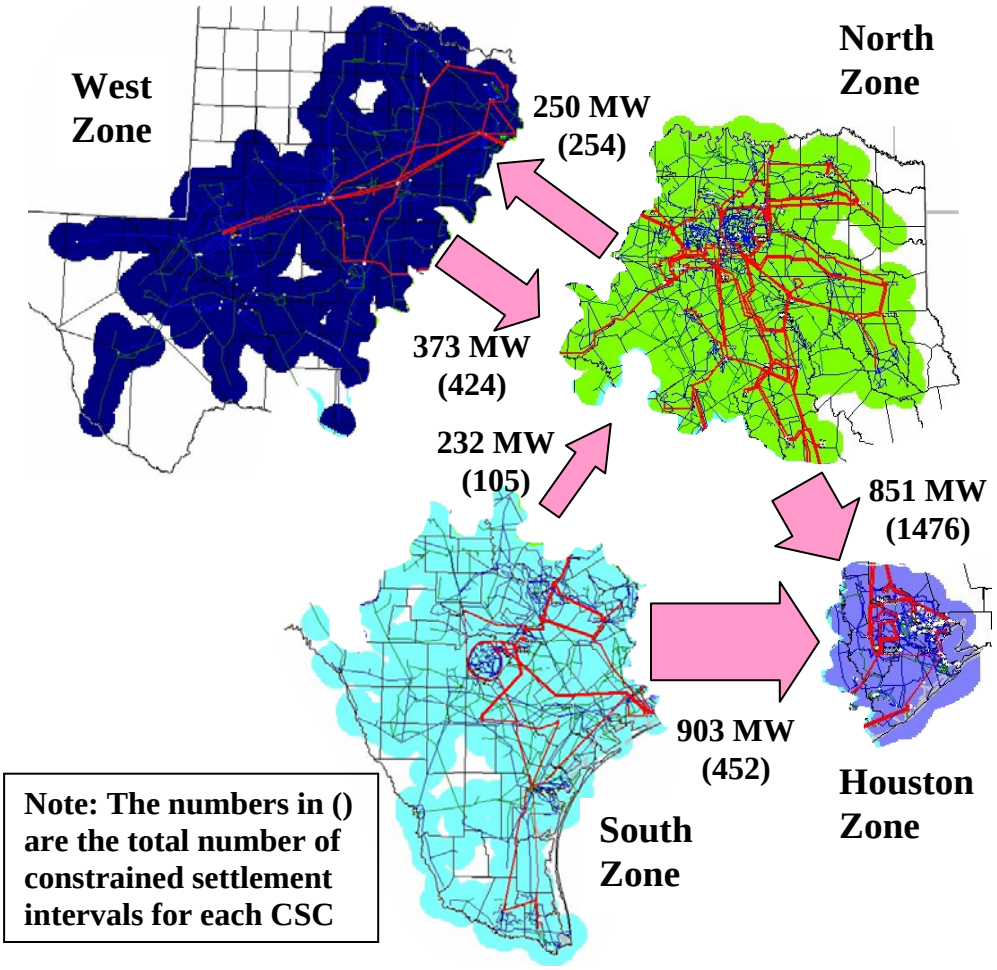


Figure 51 shows that inter-zonal congestion was most significant on the North to Houston CSC which exhibited SPD-calculated flows averaging 851 MW during 1,476 constrained intervals in 2007. Congestion was also significant on the South to Houston and West to North CSCs.

1. Congestion Rights in 2007

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

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

To analyze the congestion rights in ERCOT, we first review the TCRs and PCRs that were allocated for each CSC in 2007. Figure 52 shows the average number of TCRs and PCRs that were allocated for each of the CSCs in 2007, as well as the average SPD-modeled flows during the constrained intervals.

Figure 52: Transmission Rights vs. Real-Time SPD-Calculated Flows Constrained Intervals

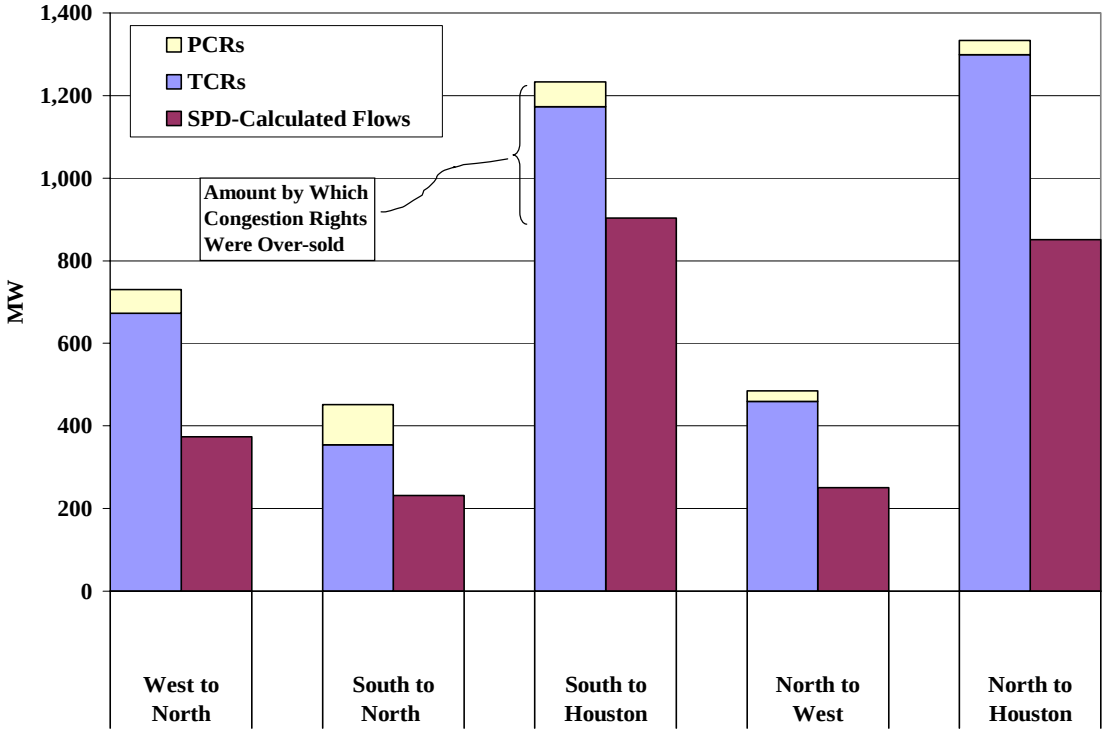


Figure 52 shows that total congestion rights (the sum of PCRs and TCRs) on all the interfaces exceeded the average real-time SPD-calculated flows during constrained intervals. These results indicate that the congestion rights were oversold in relation to the SPD-calculated limits for some CSCs. For instance, congestion rights for the North to Houston CSC were oversold by an average of 482 MW.

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

For instance, suppose the SPD-calculated flow limit is 300 MW for a particular CSC during a constrained interval. Also suppose that the holders of congestion rights own a total of 800 MW

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

To better understand the nature and causes of the shortfall implied by the results of Figure 52, we compare the SPD-calculated flows and congestion rights quantities for each of the constrained intervals by CSC. In addition to the observation of SPD flow versus the congestion rights, we also present the relationship between actual flows versus the actual CSC limits. Over-constraining the CSC limit will cause unnecessary congestion costs and will distort balancing energy market prices, which are undesirable for an efficient market. Under-constraining the CSC limit will cause reliability issues. Although exact matching of the actual flow with the actual physical limit is ideal, fluctuations of the actual flows around the physical limit are expected due to the simplifying assumptions of the zonal market model. However, significant divergence is not desirable.

2. Congestion on South to North CSC

Figure 53 shows the total quantity of congestion rights allocated by ERCOT for the South to North interface relative to the real-time SPD-calculated flows over the interface when the constraint was binding during 2007. Because only congested intervals are shown, some months will have significantly more observations than other months. Although some congestion occurred in every month, the month of November accounted for 28 percent of all constrained intervals during 2007 due to a number of planned transmission outages.

As explained in more detail below, the projected quantity of congestion rights changes from month to month as ERCOT reassesses the capability of each interface. ERCOT then adjusts the quantity of TCRs accordingly in the monthly auctions. Figure 53 shows these changes in the congestion rights relative to the SPD-calculated flows, which fluctuate considerably in the congested intervals. In the figure, Total Congestion Rights include both TCRs and PCRs.

Figure 53: Congestion Rights Allocated vs. SPD Flows during Constrained Intervals South to North

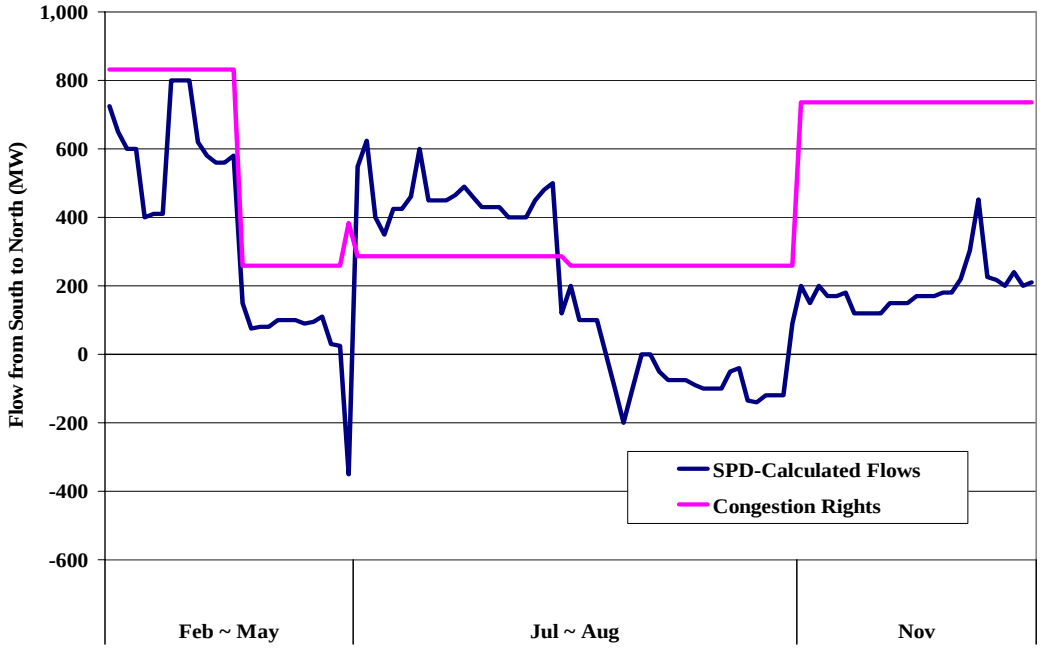


Figure 53 indicates that the quantity of outstanding congestion rights fluctuated considerably during 2007. In November, more than 700 MW of rights for the South to North CSC were available, whereas for March, July and August, less than 300 MW of congestion rights were allocated for the South to North CSC in 2007. This variation has to do with the complex nature of the South to North interface which results in it being constrained under a variety of circumstances.

Prior to each month, ERCOT estimates the transmission capability of the South to North interface based on transmission planning cases which use seasonal peak conditions. While two major lines make up the South to North interface, nearly 20 other transmission elements are defined as Closely Related Elements ("CREs"). Transmission constraints on the CREs can reduce the amount that can be transferred across the two major lines. The pattern of flows can vary considerably, partly because of changes in the particular outages that are anticipated. Also, there is no guarantee that flows across the two main lines and all of the CREs will be in the same direction in every planning case. These issues highlight some of the problems that arise in the simplified zonal congestion management system. The nodal framework is better able to manage individual pieces of the transmission system, allowing more efficient utilization of the grid.

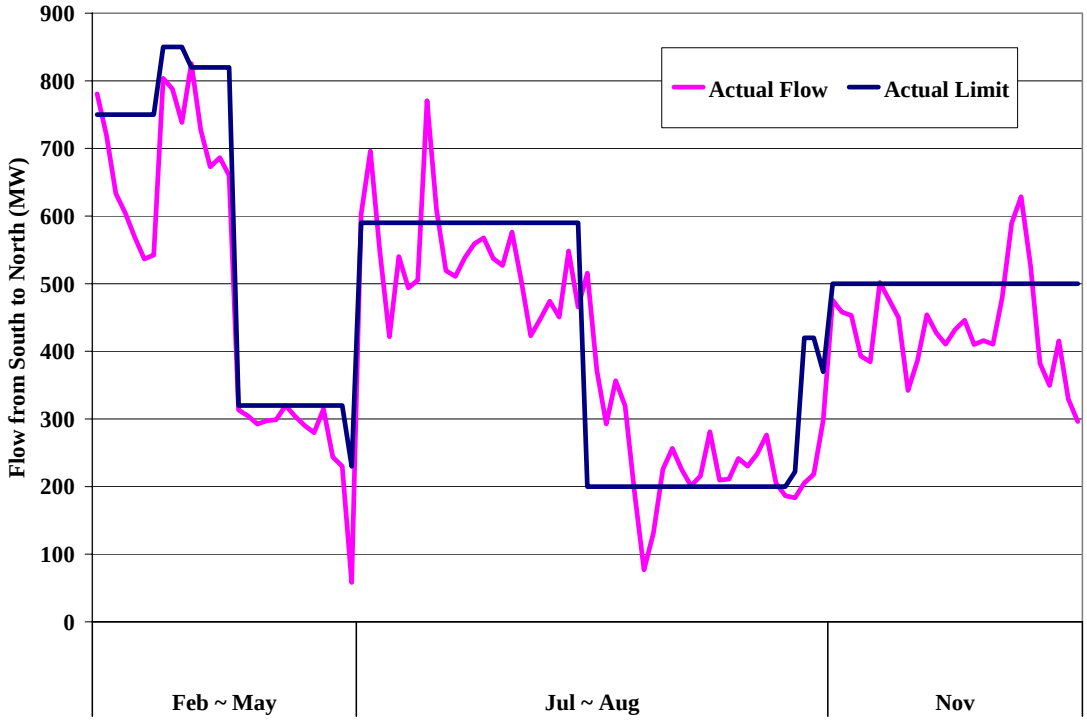
For the South to North CSC, the congestion rights were above and below SPD flows during different months for the congested intervals in 2007. The figure shows ten constrained intervals when the SPD-calculated flows were *negative* at times during May and August.

These very low SPD-calculated flows generally do not reflect the actual physical flows in real time, *i.e.*, when the actual system conditions result in more flows over the South to North constraint than the simplified zonal model would predict. To prevent physical flows from exceeding the physical limits of the CSC, the ERCOT operators manually reduce the limit on the South to North interface in SPD. This causes SPD to redispatch generation in the various zones to reduce flows over the interface. Hence, because the SPD-calculated flows can be substantially different than actual flows, the ERCOT operators manage congestion by lowering the SPD limit when a constraint is physically binding to prevent additional flow over the CSC. Under extreme conditions, the operators must reduce the SPD limit into the negative range.

In 2006, the South-to-North CSC congested 583 times with an average flow of 582 MW, while in 2007, it congested only 105 times with an average flow of 232 MW. Along with the reduction in South-to-North congestion, there was an increase in congestion from North-to-South in 2007. Because there was not a North-to-South CSC defined for 2007, this congestion was managed with local congestion management. However, in response to these changing congestion patterns, a new CSC was added for the North-to-South interface for 2008.

As discussed above, the simplified modeling assumptions specified in the ERCOT protocols for the current zonal market causes the interzonal power flows calculated by SPD to frequently diverge significantly from the actual flows, which would cause unnecessary congestion in some extreme cases. The following figure presents the South to North actual flow versus the actual South to North limit.

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



The South to North CSC experienced 105 intervals of congestion during 2007. During the congestion intervals, the actual flow amount over the CSC was less than the actual physical limit by an average of 80 MW. Because of the long times between the dispatch decisions and the operating interval, as well as the simplifying assumptions of the zonal model, the tendency of ERCOT operations in the zonal model is to operate more conservatively and over-constrain CSCs. As also shown in the figures in the following subsections, this was true for all of the CSCs in 2007. The implementation of the nodal market will improve the efficiency of the management of these constraints by providing more frequent re-dispatch that utilizes data that is more reflective of current operating conditions, and by relying upon a commercial model that is consistent with the operational reality.

3. Congestion on South to Houston CSC

This interface experienced 452 constrained intervals, reduced significantly from 2006, when it congested 1,001 times. The most congestion occurred in May and June due to transmission outages associated with the construction of new transmission lines that effectively relieved the congestion on this interface for the remainder of the year, with the exception of October when

transmission maintenance outages occurred. In the months with significant congestion, SPD flows averaged between 1,020 and 1,156 MW and the congestion rights were about the average level of SPD flows. However, congestion rights were above the SPD flow levels in the months of February through April as well as in October. Figure 55 shows the comparison between actual flow and the congestion rights quantities.

Figure 55: Congestion Rights Allocated vs. SPD Flows during Constrained Intervals South to Houston

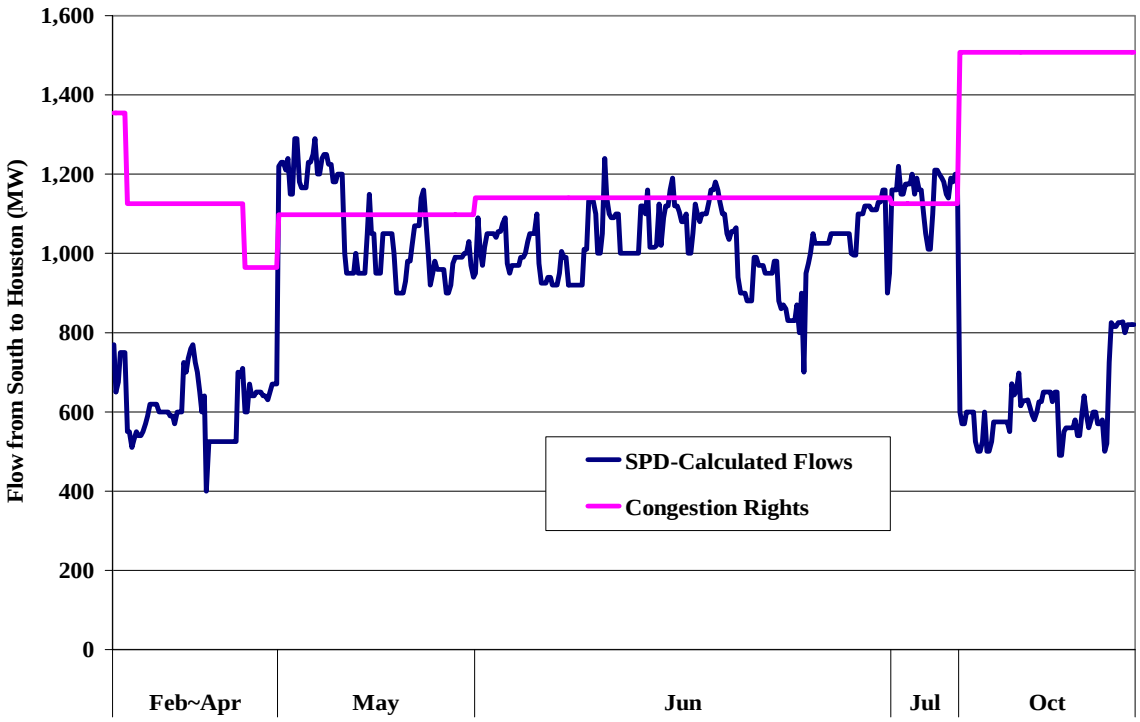


Figure 56: Actual Flows versus Physical Limits During Congestion Intervals South to Houston

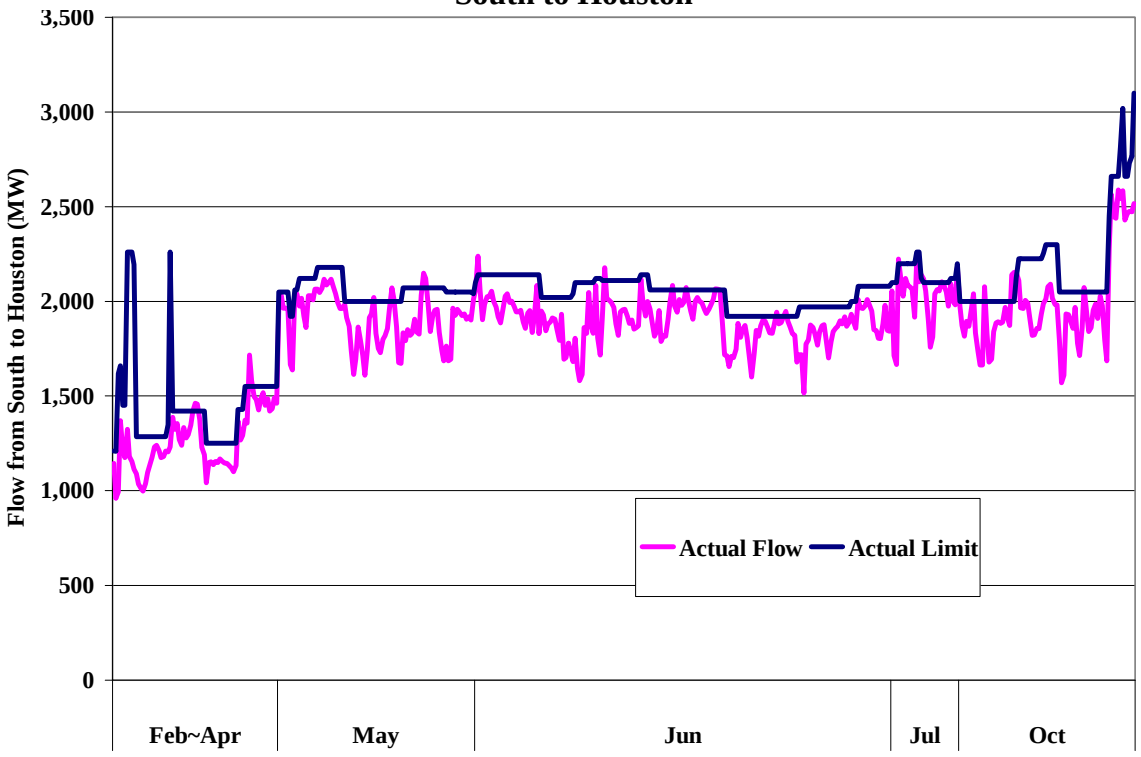


Figure 56 compares the actual flow with the actual limit for the South to Houston CSC. During the congestion intervals, the actual flow over the CSC was less than the physical limit by an average of 169 MW.

4. Congestion on North to Houston CSC

This CSC was created in 2004 to manage congestion on a path into Houston that is usually able to physically transfer more than 2,000 MW. The congestion rights were almost in line with the average SPD flows during the months of June to October. However, the congestion rights were above the SPD flow levels during the months of April and November and below the SPD flow levels during the month of May. In November, the number of congestion rights allocated were above the average SPD flow levels during congestion periods by 1,003 MW due to planned transmission outages that were not accounted for at the time of the TCR auction. The frequency of transmission constraints rose dramatically in November in conjunction with the increase of rights allocated. In 2007, this interface became the most congested interface with congestion occurring in 1,476 intervals, with a significant portion of the congestion occurring in November.

Figure 57: Congestion Rights Allocated vs. SPD Flows during Constrained Intervals North to Houston

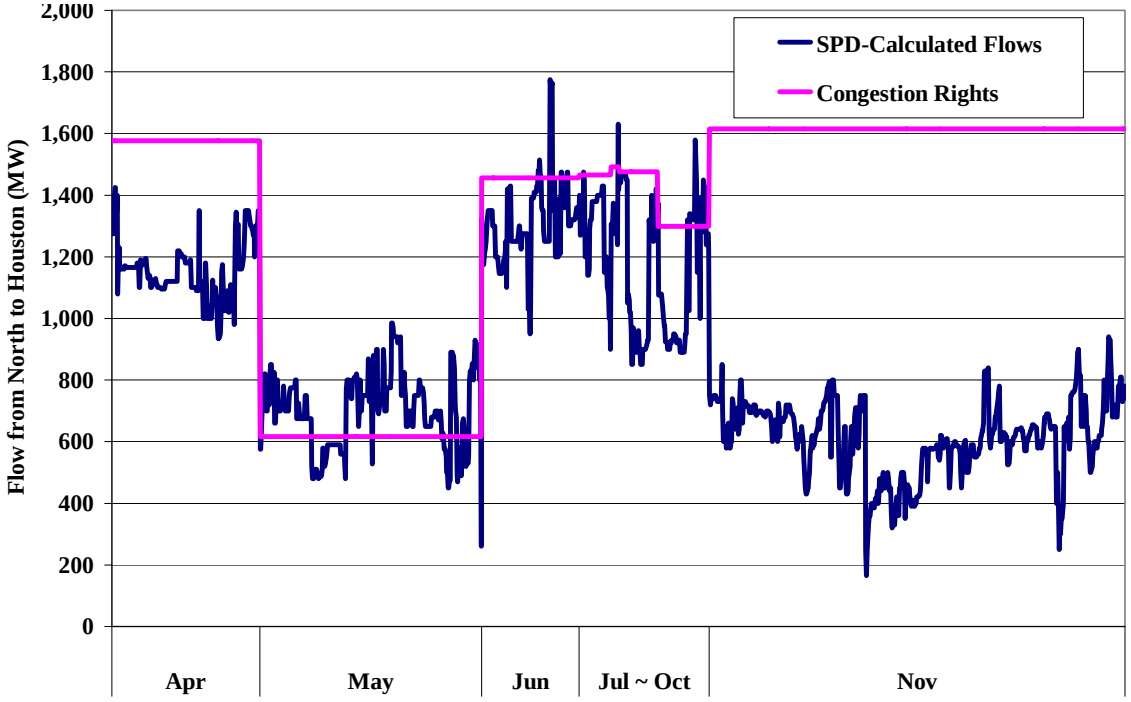


Figure 58: Actual Flows versus Physical Limits during Congestion Intervals North to Houston

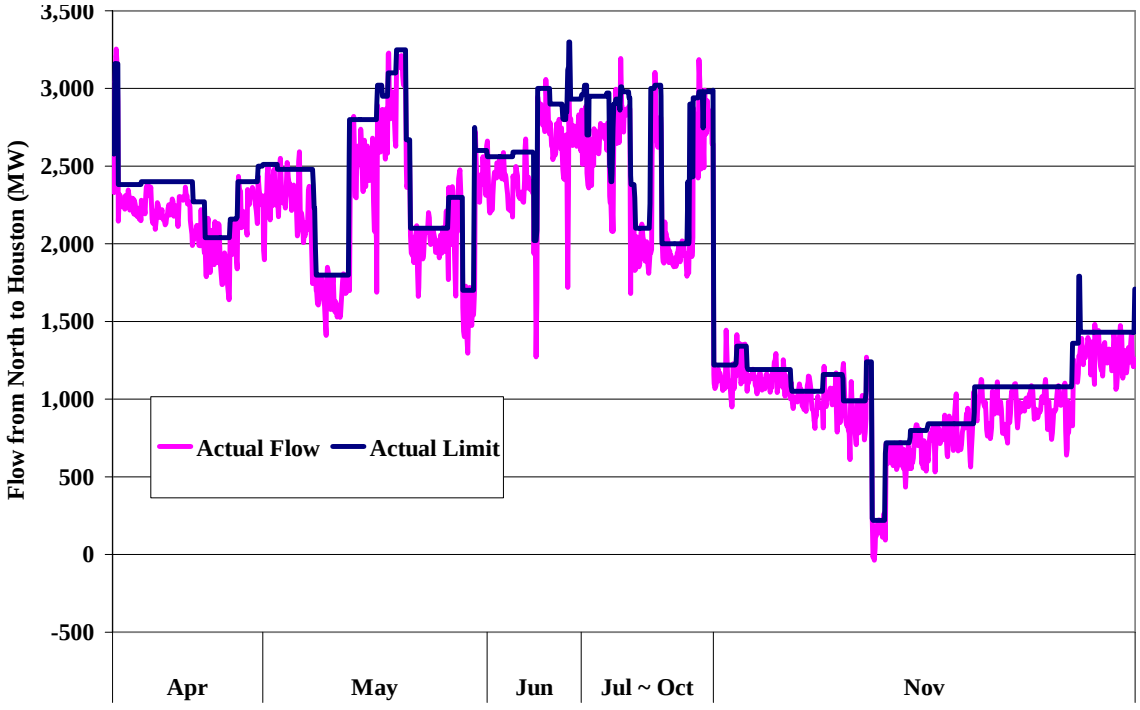


Figure 58 compares the actual flow with the actual limit for the North to Houston CSC. During the congestion intervals, the actual flow over the CSC was less than the physical limit by an average of 167 MW.

5. Congestion on North to West CSC

This CSC was congested most frequently during the winter months with approximately 39 percent of constrained intervals in November to December. Congestion rights were above the SPD flows in the months of January through May and also August through December. Although the number of congestion rights allocated for this interface varied from 178 to 823 MW over the year, the SPD flows averaged just 250 MW during constrained intervals.

Figure 59: Congestion Rights Allocated vs. SPD Flows during Constrained Intervals North to West

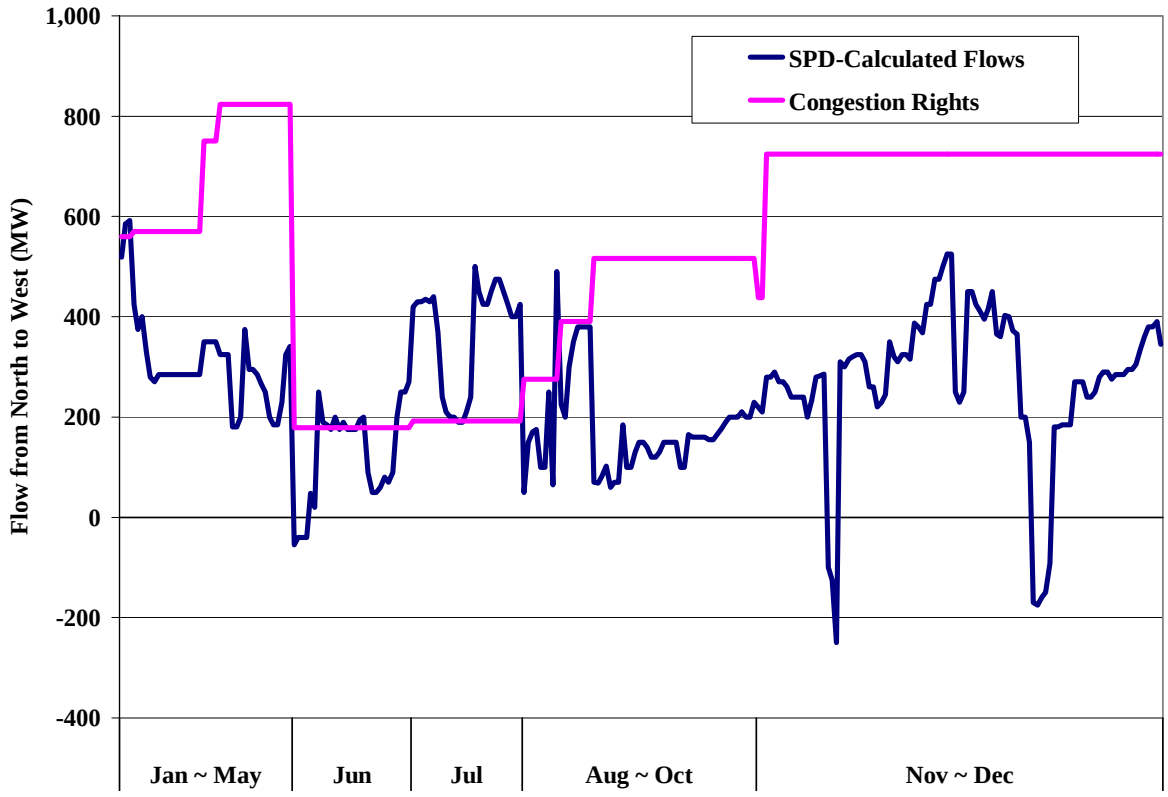


Figure 60: Actual Flows versus Physical Limits during Congestion Intervals North to West

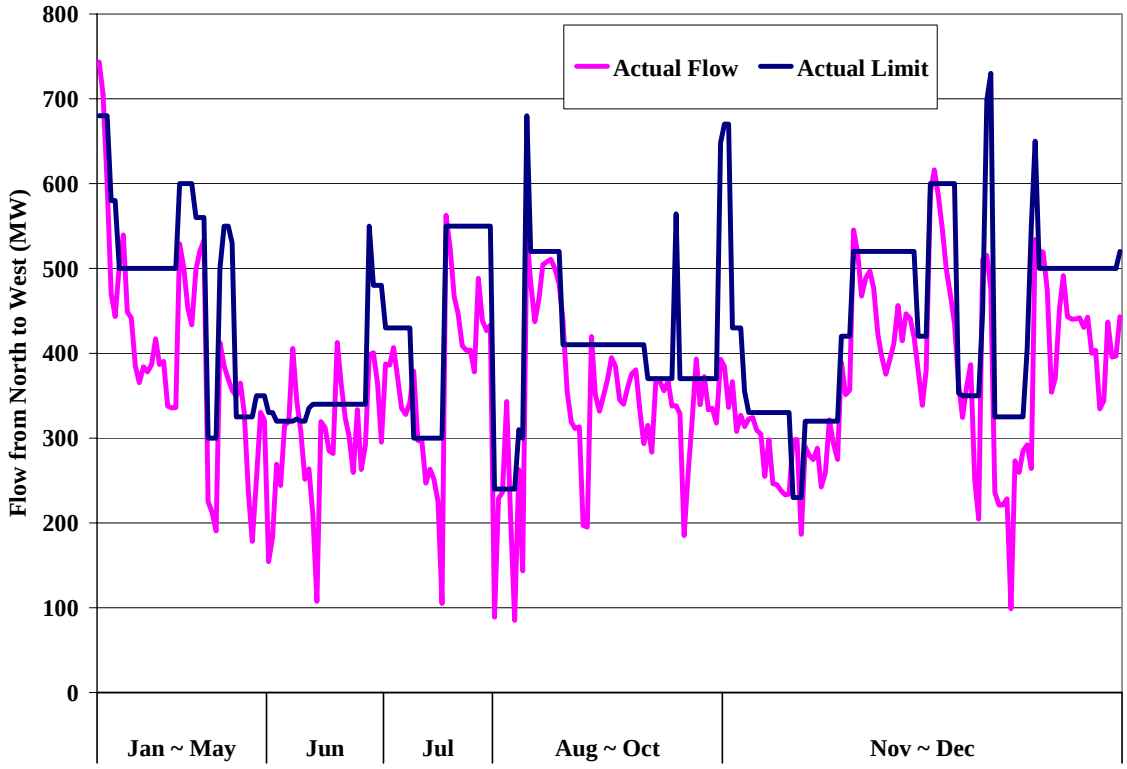


Figure 60 compares the actual flow with the actual limit for the North to West CSC. During the congestion intervals, the actual flow over the CSC was less than the physical limit by an average of 78 MW.

6. Congestion on West to North CSC

This CSC was congested in 424 intervals during 2007, much more than the congestion frequency in 2006 of 48 intervals. Most of the increase occurred in the last quarter of 2007, and is associated with the significant increases in wind generation in the West Zone during this time period. Different from other CSCs, the TCRs allocated were almost always higher than the actual SPD flow during congestion intervals. The average SPD flow during congestion intervals was 373 MW and the average TCR sold on the CSC was 795 MW. The main reason for the difference is due to planned and unplanned transmission outages that are not accounted for in the TCR auctions that significantly reduce the real-time transfer capability of the CSC. In addition, at times there are dynamic stability limits for West-to-North transfers that may result in limits that are much lower than the transfer capability used to determine TCR auction quantities.

Figure 61: Congestion Rights Allocated vs. SPD Flows during Constrained Intervals West to North

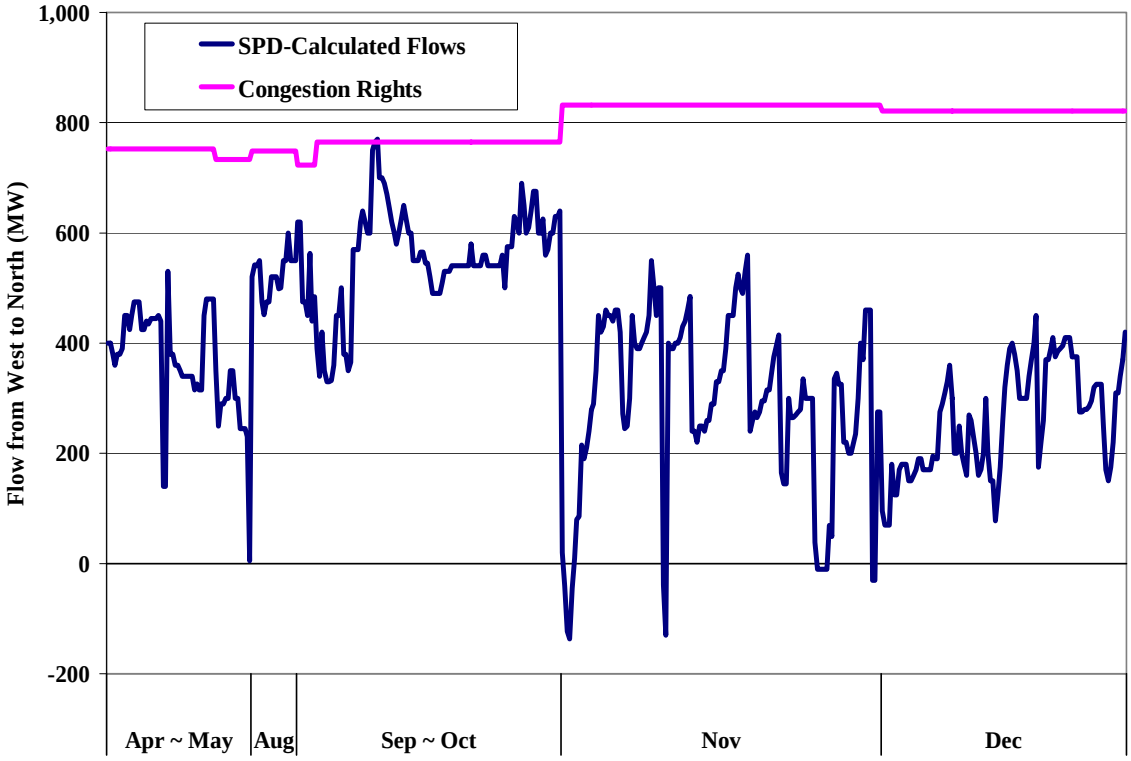


Figure 62: Actual Flows versus Physical Limits during Congestion Intervals West to North

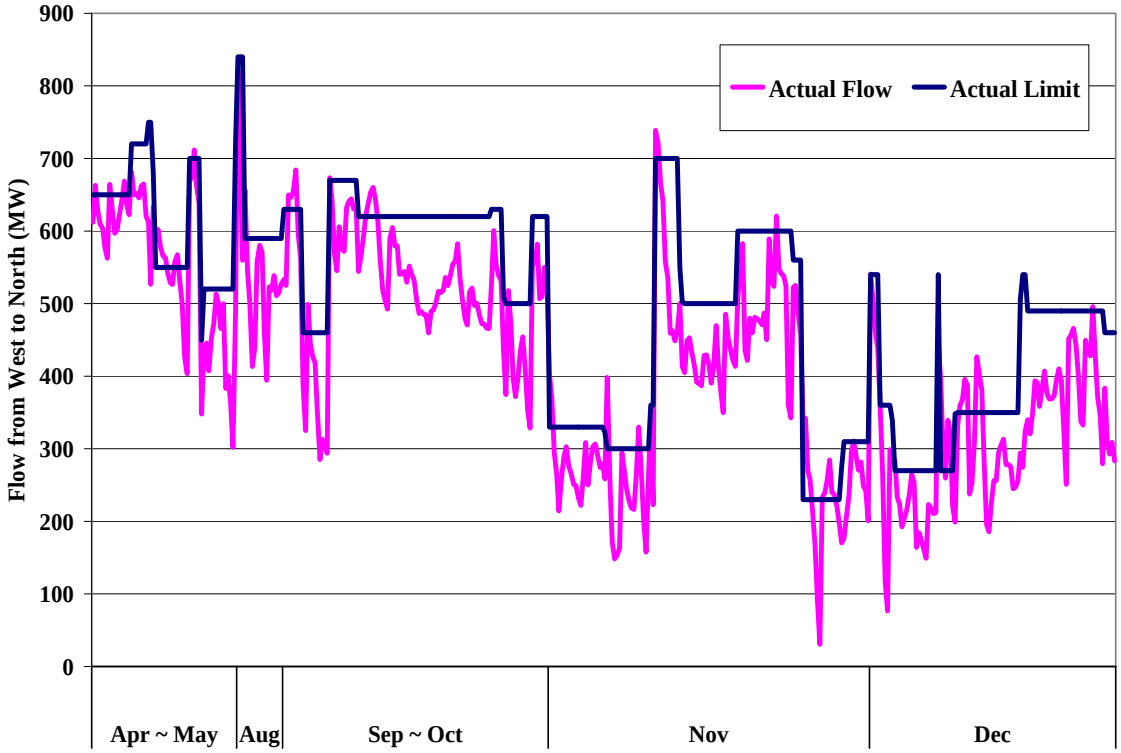


Figure 62 compares the actual flow with the actual limit for the West to North CSC. During the congestion intervals, the actual flow over the CSC was less than the physical limit by an average of 81 MW.

C. Congestion Rights Market

In this subsection, we review ERCOT's process to establish the quantity of congestion rights allocated or sold to participants. ERCOT performs transmission planning studies to determine the capability of each interface under peak summer conditions. This summer planning study is the basis for designating 40 percent of the transmission congestion rights sold in the annual auction. These rights are auctioned in December for the coming year. The remaining 60 percent of the transmission congestion rights are designated based on monthly updates of the summer study.[29] Since the monthly studies tend to more accurately reflect conditions that will prevail in the coming month, the monthly designations tend to more closely reflect actual transmission limits.

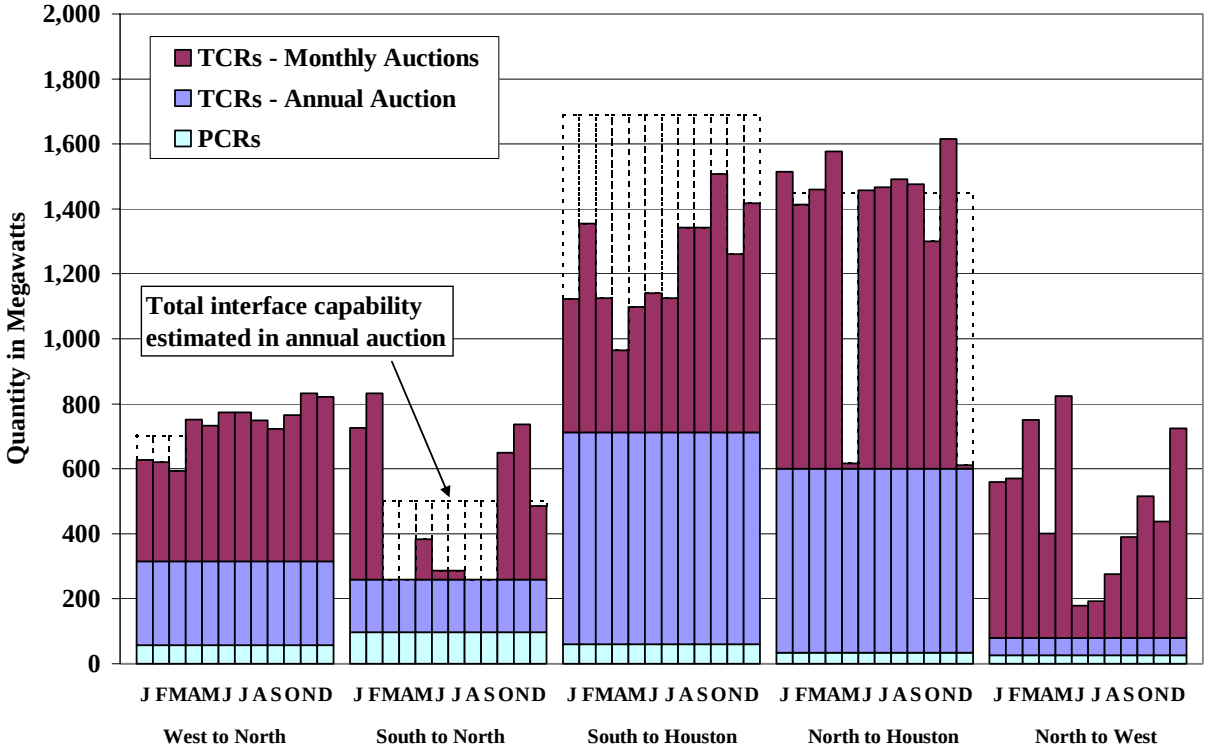
However, the summer monthly studies used to designate the TCRs do not always accurately reflect transmission conditions that can arise in real-time. This happens for two main reasons. First, transmission and generation outages can occur unexpectedly and significantly reduce the transfer capability of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

To examine how these processes have together determined the total quantity of rights sold over each interface, Figure 63 shows the quantity of each category of congestion rights for each month during 2007. The quantities of PCRs and annual TCRs are constant across months and

[29] Prior to 2005, 60 percent of estimated capability (after accounting for Pre-assigned Congestion Rights which are assigned to NOIEs) was sold in the annual auction. The remaining 40 percent was sold in the monthly auctions. This was changed because there were instances when the capability estimated before the monthly auction was more than 40 percent lower than the capability estimated before the annual auction. In these cases, no congestion rights could be sold in the monthly auction because no unsold capacity remained.

were determined before the beginning of 2007, while monthly TCR quantities can be adjusted monthly.

Figure 63: Quantity of Congestion Rights Sold by Type
2007

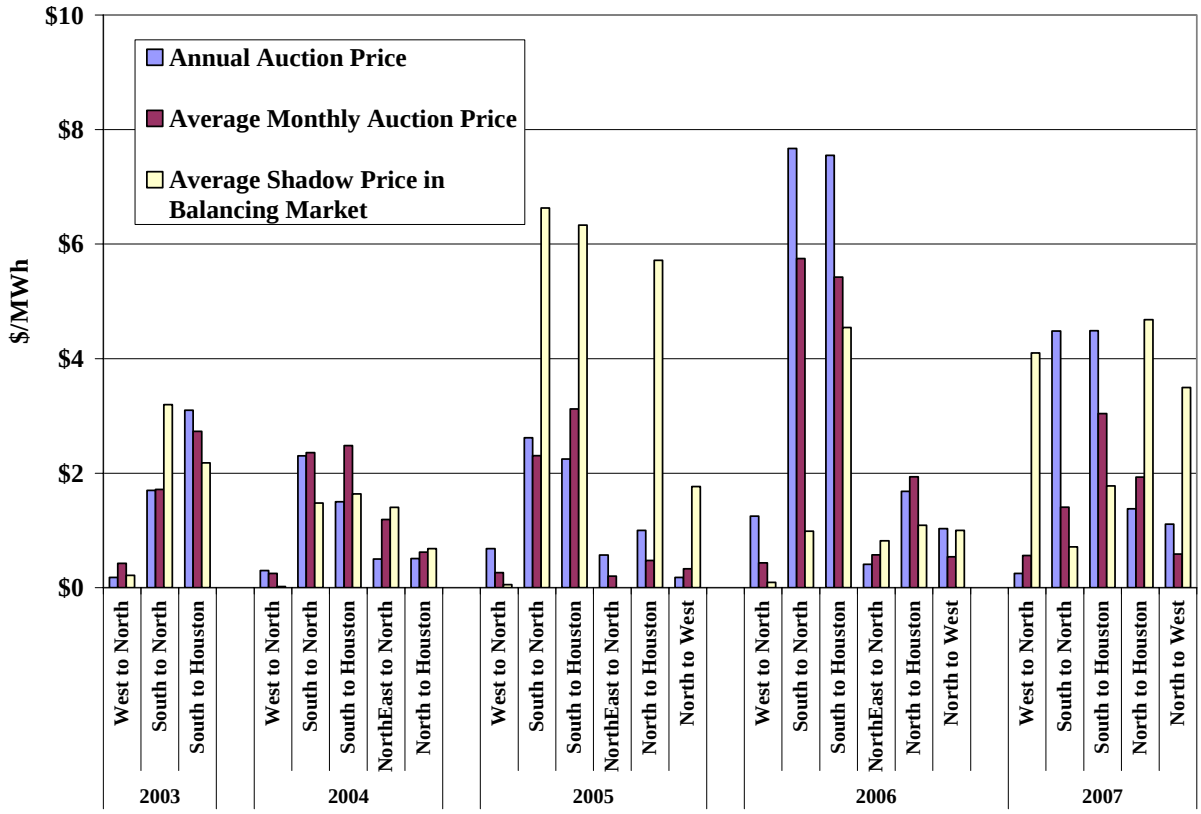


When the monthly planning studies indicate changes from the summer study, revisions are often made to the estimated transmission capability. Therefore, the auctioned congestion rights may increase or decrease relative to the amount estimated in the summer study. The shadow boxes in the figure represent the capability estimated in the summer study that is not ultimately sold in the monthly auction. When there is no shadow box in Figure 63, the total quantity of PCRs and TCRs sold in the annual and monthly auctions equaled or exceeded the summer estimate and therefore no excess capability is shown.

The South to North, South to Houston and North to Houston interfaces experienced the largest fluctuations in the estimates of transmission capacity from the annual auction to the monthly auction. In fact, the South to North TCRs were not even auctioned during four of the monthly auctions. The divergence between annual and monthly estimates of transmission capacity on the other interfaces was smaller.

Market participants who are active in congestion rights auctions are subject to substantial uncertainty. Outages and other contingencies occur randomly that can substantially change the market value of a congestion right. Real-time congestion prices reflect the cost of interzonal congestion and are the basis for congestion payments to congestion rights holders. In a perfectly efficient system with perfect forecasting by participants, the average congestion price should equal the auction price. However, we would not expect full convergence in the real-world, given uncertainties and imperfect information. To evaluate the results of the ERCOT congestion rights market, in Figure 64 we compare the annual auction price for congestion rights, the average monthly auction price for congestion rights, and the average congestion price for each CSC.

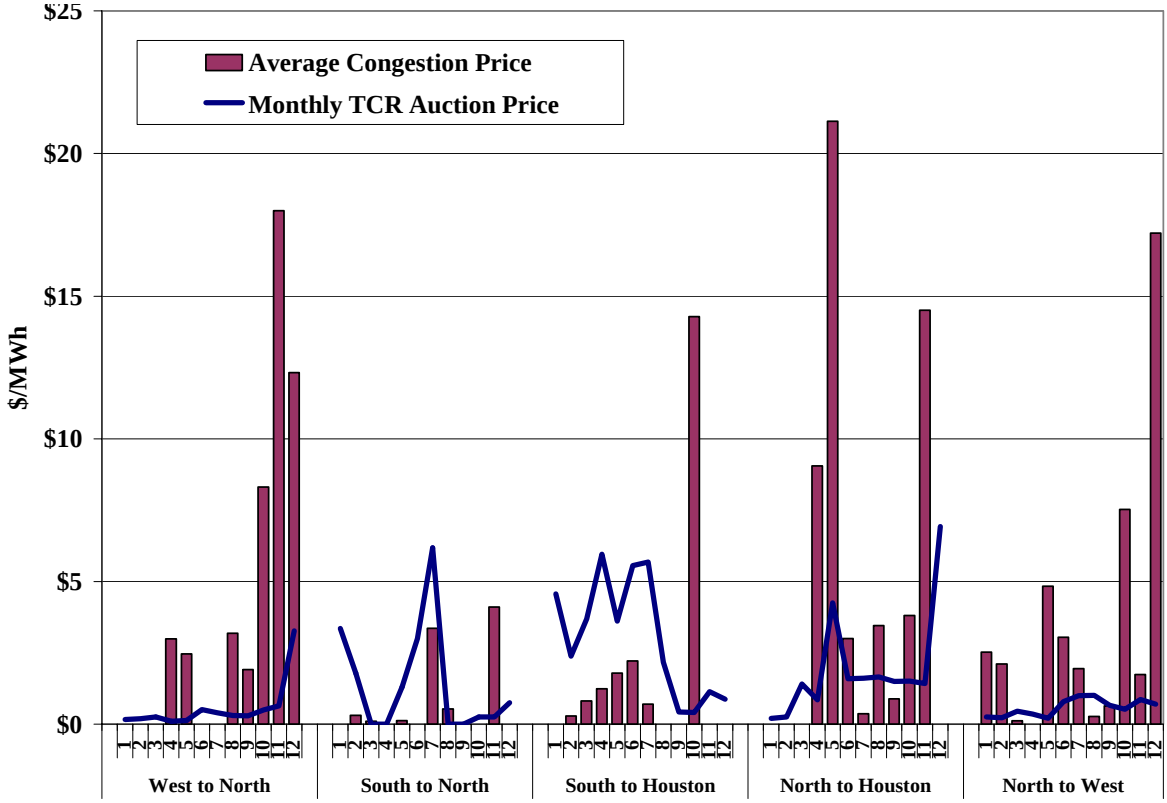
Figure 64: TCR Auction Prices versus Balancing Market Congestion Prices 2003 to 2007



This figure shows that there is a tendency for the TCRs to settle at prices that are closer to the previous years' value, but that real-time congestion prices often diverge significantly from auction prices. This suggests that participants are not able to forecast annual interzonal congestion costs and accurately value the TCRs in the annual auction, and instead rely more upon historical market outcomes.

Figure 65 compares monthly TCR auction prices with monthly average real-time CSC shadow prices from SPD for 2007. The TCR auction prices are expressed in dollars per MWh.

Figure 65: Monthly TCR Auction Price and Average Congestion Value 2007



The TCR price trends for North to Houston CSCs correlated well with the actual congestion prices, although the TCR prices for this CSC are far below the congestion prices. Overall, market participants did a poor job predicting fluctuations in congestion during 2007, particularly on the South to Houston interfaces. For South to Houston interfaces, there was one month when balancing market congestion spiked when balancing prices far exceeding the TCR prices.

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

collected in the real-time market from participants scheduling transfers between zones or power flows resulting from the balancing energy market.

The congestion rent from the balancing energy market is associated with the schedules and balancing deployments that result in interzonal transfers during constrained intervals (when there are price differences between the zones). For instance, suppose the balancing energy market deployments result in exports of 600 MWh from the West Zone to the North Zone when the price in the West Zone is $40/MWh and the price in the North Zone is $55/MWh. The customers in the North Zone will pay $33,000 (600 MWh * $55/MWh) while suppliers in the West Zone will receive $24,000 (600 MWh * $40/MWh). The net result is that ERCOT collects $9,000 in congestion rent ($33,000 – $24,000) and uses it to fund payments to holders of TCRs.[30] If the quantity of TCRs perfectly matches the capability of the CSC in the balancing energy market, the congestion rent will perfectly equal the amount paid to the holders of TCRs.

Figure 66 reviews the results of these processes by showing (a) monthly and annual revenues from the TCR auctions, (b) credit payments earned by the holders of TCRs based on real-time outcomes, and (c) congestion rent from schedules and deployments in the balancing energy market.

[30] This explanation is simplified for the purposes of illustration. However, congestion rents would also depend on the net imports into and net exports from the other three zones as well as the zonal prices. Furthermore, the net exports from the West Zone do not necessarily match the net imports into the North Zone in real-time operation.

Figure 66: TCR Auction Revenues, Credit Payments, and Congestion Rent[31]
2003 to 2007

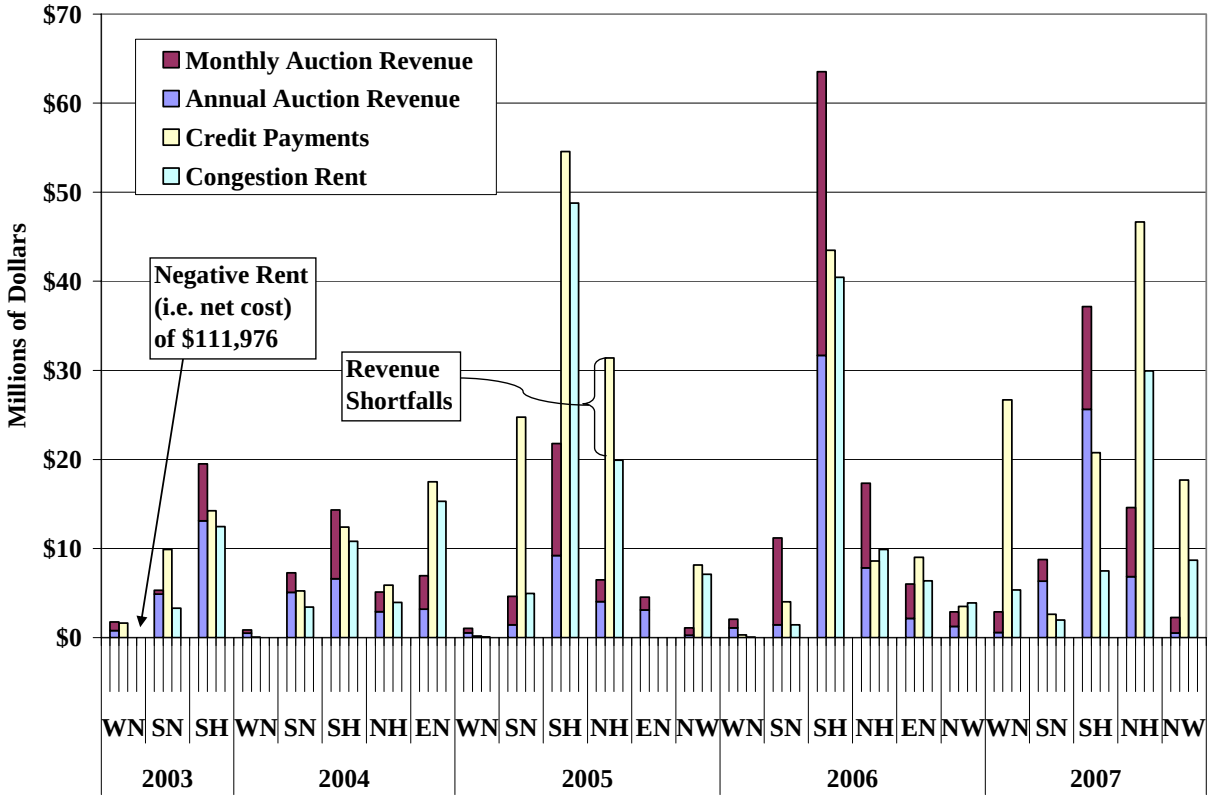


Figure 66 shows that in 2004, the auction revenues were consistent with credit payments for the three CSC that existed in 2003. This appeared to be due to market participant basing their valuations of the TCRs on their value in the prior year. The auction revenues for the North to Houston CSC, which was added for the first time in 2004, were quite close to credit payments. However, market participants substantially under-valued congestion on the Northeast to North interface, which was also new in 2004.

In 2005, the auction revenues were greatly exceeded by credit payments for the four interfaces with significant congestion. This was because the TCR market under-estimated the volume of congestion that would occur in the balancing market. TCR prices were generally consistent between 2004 and 2005, suggesting that market participants based their expectations on the levels of congestion that occurred in 2004. Since interzonal congestion in the balancing market

[31] The source for congestion rents is the ERCOT TCR Program Report. However, this source incorporates an additional term based on the revenue impact of using generation-weighted shift factors for loads instead of the load-weighted shift factor.

was far greater in 2005 than in previous years, payments to TCR holders exceeded TCR auction revenues by a significant margin.

In contrast to 2005, auction revenues for the South to North, South to Houston and North to Houston interfaces exceeded credit payments in 2006. As shown in Figure 66, for those interfaces, auction prices exceeded the congestion prices. The magnitude of credit payments are in the same trend as in 2005, but the 2006 South to North and North to Houston interfaces exhibited far less credit payments and congestions rent compared to 2005. Northeast to North interfaces experienced more congestion than 2005 and hence the credit payments went up compared to 2005.

In 2007, the South to North and South to Houston interfaces exhibited similar pattern as in 2006, where the auction revenue exceeded credit payments. In contrast, the West to North, North to West and the North to Houston interface show signification higher credit payments than auction revenue, while there are still revenue short falls on those three interfaces since credit payments also exceeded congestion rent.

Figure 66 also shows that payments to TCR holders have consistently exceeded the congestion rents that have been collected from the balancing market since the creation of the TCR market. The difference was relatively modest in 2004 when congestion rents covered 81 percent of payments to TCR holders. However, in 2003 and 2005, congestion rents covered only 61 percent and 68 percent, respectively, of payments to TCR holders. In 2006, congestion rents covered 90 percent of payments to TCR holders, which is an improvement from previous years. In 2007, however, congestion rents only covered 47 percent of payments to TCR holders. When congestion rents fall significantly below payments to TCR holders, it implies that the SPD-calculated flows across constrained interfaces have been systematically lower than the amount of TCRs sold for the interfaces.

As described above, a revenue shortfall exists when the credit payments to congestion rights holders exceed the congestion rent. This shortfall is caused when the quantity of congestion rights exceeds the SPD-calculated flow limits in real-time.[32] These shortfalls are included in the

[32] For instance, if the shadow price on a particular CSC is $10 per MWh for one hour and the SPD flow limit is 300 MW, ERCOT will collect $3,000 in congestion rents. However, if the holders of congestion rights

Balancing Energy Neutrality Adjustment charge and assessed to load ERCOT-wide. Collecting substantial portions of the congestion costs for the market through such uplift charges reduces the transparency and efficiency of the market. It also increases the risks of transacting and serving load in ERCOT because uplift costs cannot be hedged.

D. Local Congestion and Local Capacity Requirements

In this subsection, we address local congestion and local reliability requirements by evaluating how ERCOT manages the dispatch and commitment of generators when constraints and reliability requirements arise that are not recognized or satisfied by the current zonal markets. Local (or intrazonal) congestion occurs in ERCOT when a transmission constraint is binding that is not defined as part of a CSC or CRE. Hence, these constraints are not managed by the zonal market model. ERCOT manages local congestion by requesting that generating units adjust their output quantities (either up or down). When insufficient capacity is committed to meet reliability, ERCOT commits additional resources to provide the necessary capacity in either the day-ahead or real-time. Some of this capacity is instructed to be online through Reliability Must Run ("RMR") contracts.

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

When a unit is dispatched out of merit (OOME up or OOME down), the unit is paid for a quantity equal to the difference between the scheduled output based on the unit's resource plan

own a total of 800 MW, then ERCOT must pay out $8,000 worth of credit payments. Thus, the revenue shortfall for ERCOT would be $5,000.

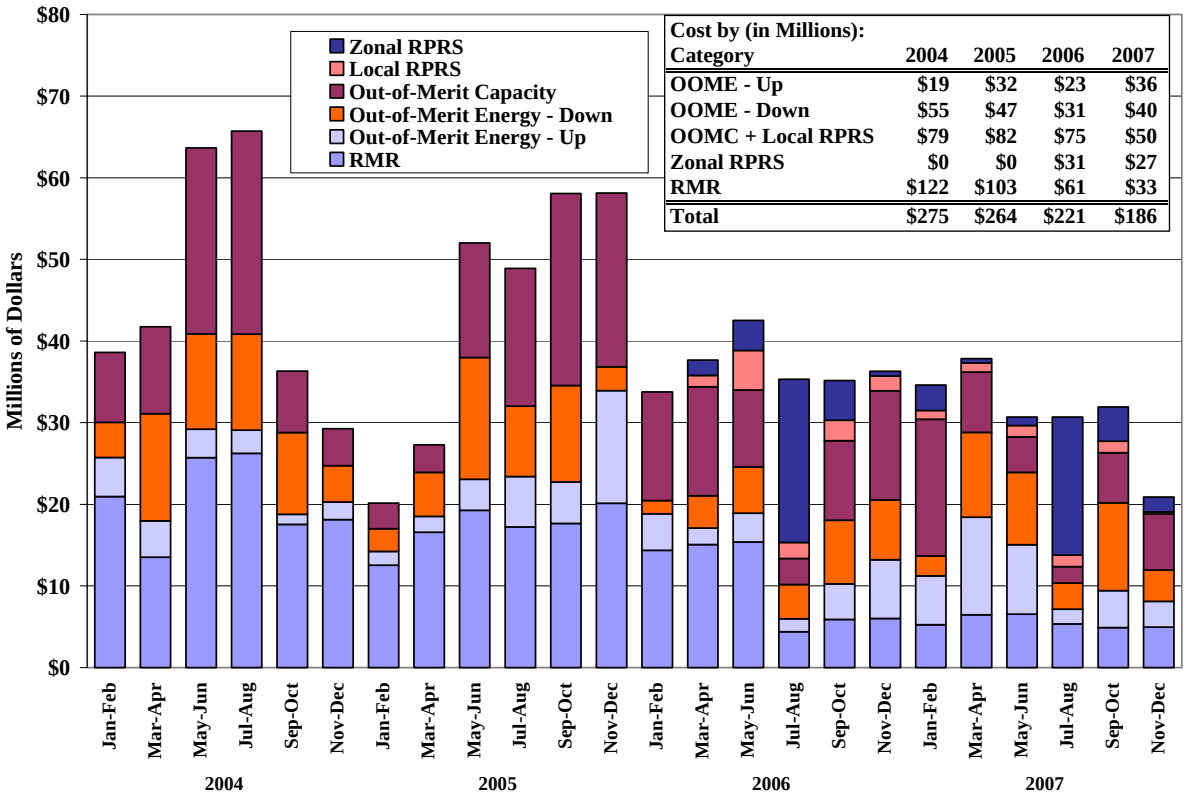
and the actual output resulting from the OOME instruction from ERCOT. The payment per MWh for OOME is a pre-determined amount specified in the ERCOT Protocols based on the type and size of the unit, the natural gas price, and the balancing energy price. The net payment to a resource receiving an OOME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

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

A unit providing capacity under an OOMC instruction is paid a pre-determined amount, defined in the ERCOT Protocols, based on the type and size of the unit, natural gas prices, the duration of commitment, and whether the unit incurred start-up costs. Owners of a resource receiving an OOMC instruction from ERCOT are obligated to offer any available energy from the resource into the balancing energy market.

Finally, RMR units committed or dispatched pursuant to their RMR agreements receive cost-based compensation. Since October 2002, ERCOT has entered into several RMR agreements with older, inefficient units that were planned to be retired. However, as a part of the RMR exit strategy process, all but three units were removed from RMR status by mid-2006. In 2007, there were only three RMR units (Laredo units 1, 2 and 3). Units contracted to provide RMR service to ERCOT are compensated for start-up costs, energy costs, and are also paid a standby fee. Figure 67 shows each of the four categories of uplift costs from 2004 to 2007.

Figure 67: Expenses for Out-of-Merit Capacity and Energy 2004 to 2007



Cost by (in Millions): Category	2004	2005	2006	2007
OOME - Up	$19	$32	$23	$36
OOME - Down	$55	$47	$31	$40
OOMC + Local RPRS	$79	$82	$75	$50
Zonal RPRS	$0	$0	$31	$27
RMR	$122	$103	$61	$33
Total	$275	$264	$221	$186

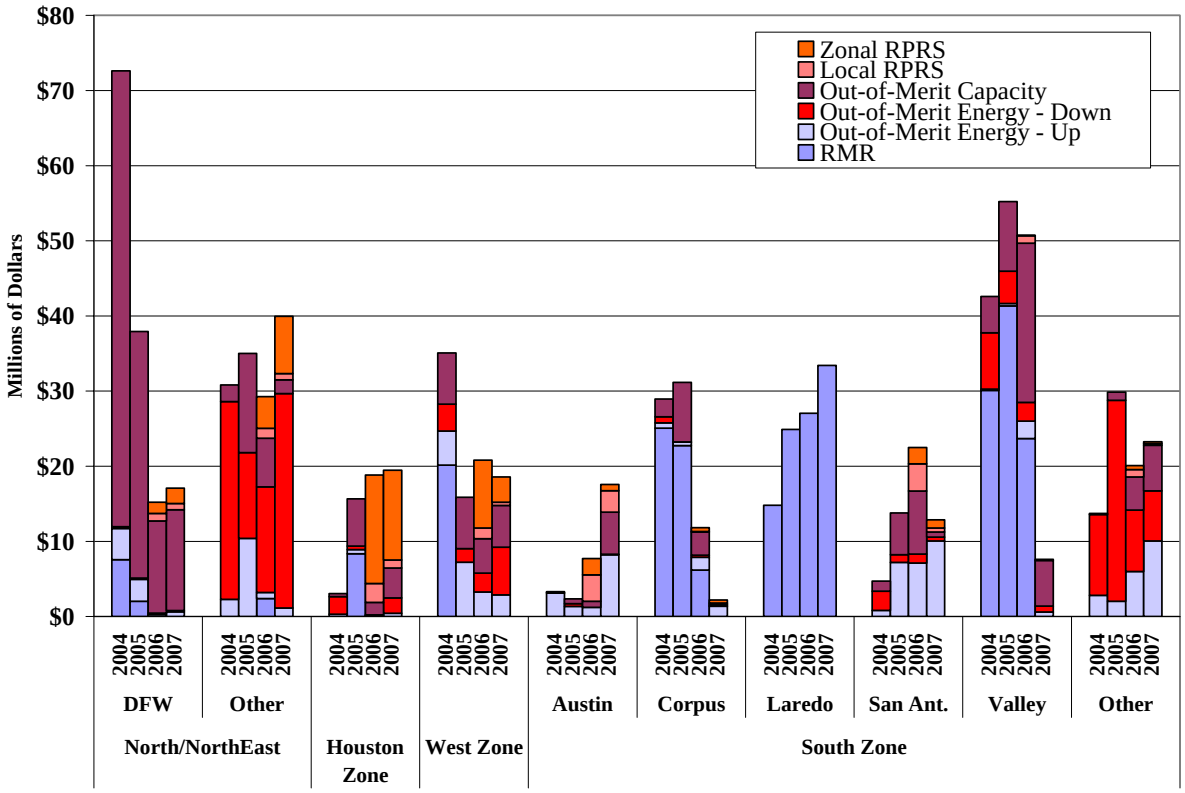
The results in Figure 67 show that overall uplift costs for RMR units, OOME units, OOMC/Local RPRS and Zonal RPRS[33] units decreased in 2007 from the 2006 level. The costs decreased by $74 million in 2006 from $264 million to $221 million. The cost further decreased by $35 million in 2007. As previously noted, there were substantial reductions to RMR cost due to the expiration of RMR agreements, which accounts for $28 million of the $35 million decrease from 2006 to 2007. Total OOME Up and OOME Down costs increased from $54 million in 2006 to $76 million in 2007. A sizable portion of this increase can be attributed to the management of North-to-South congestion in 2007 during which there was not a CSC defined for this interface. Unit commitment cost decreased in 2007 by $29 million from 2006. Notably, zonal RPRS costs for system adequacy were the highest in the peak system demand months of

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

July and August for both 2006 and 2007. These results were also likely influenced by the day-ahead load forecast issues discussed in Section I of this report.

Although the costs are borne by load throughout ERCOT, the costs are caused in specific locations because these actions, with the exception of zonal RPRS, are taken to maintain local reliability. The rest of the analyses in this section evaluate in more detail where these costs were caused and how they have changed between 2004 and 2007. Figure 68 shows these payments by location.

Figure 68: Expenses for OOME, OOMC and RMR by Region 2004 – 2007



Uplift costs decreased dramatically from 2004 to 2007 in the Dallas/Ft. Worth ("DFW") area, in the West zone and in the South zone Corpus Christi and Valley area. In DFW, the reduction was due to less frequent OOMC commitments, whereas uplift was reduced in the West zone by the elimination of RMR status for units located in that area. Corpus Christi area uplift cost reduction was primarily caused by the decrease of RMR payments, from $23 million in 2005 to $0 in 2007. RMR costs in the Laredo area increased from 2004 to 2007 due to increased fuel costs, as the number of RMR units in that area remained constant during this time period. The Austin area

exhibited the highest increase of uplift costs, from $3 million in 2004 to $18 million in 2007. This increase is most likely associated with the increase in the frequency of North-to-South congestion in 2007 that was discussed previously in this section.

V. ANALYSIS OF COMPETITIVE PERFORMANCE

In this section, we evaluate competition in the ERCOT market by analyzing the market structure and the conduct of the participants during 2007. We examine market structure using a pivotal supplier analysis, which indicates that suppliers were pivotal in the balancing energy market at a significantly smaller frequency in 2007 than in 2006. This analysis also shows that the frequency with which a supplier was pivotal increased with the level of demand. To evaluate participant conduct, we estimate measures of physical and economic withholding. We examine withholding patterns relative to the level of demand and the size of each supplier's portfolio. Based on theses analyses, we find that the overall competitive performance of the market continued its trend of improvement in 2007.

A. Structural Market Power Indicators

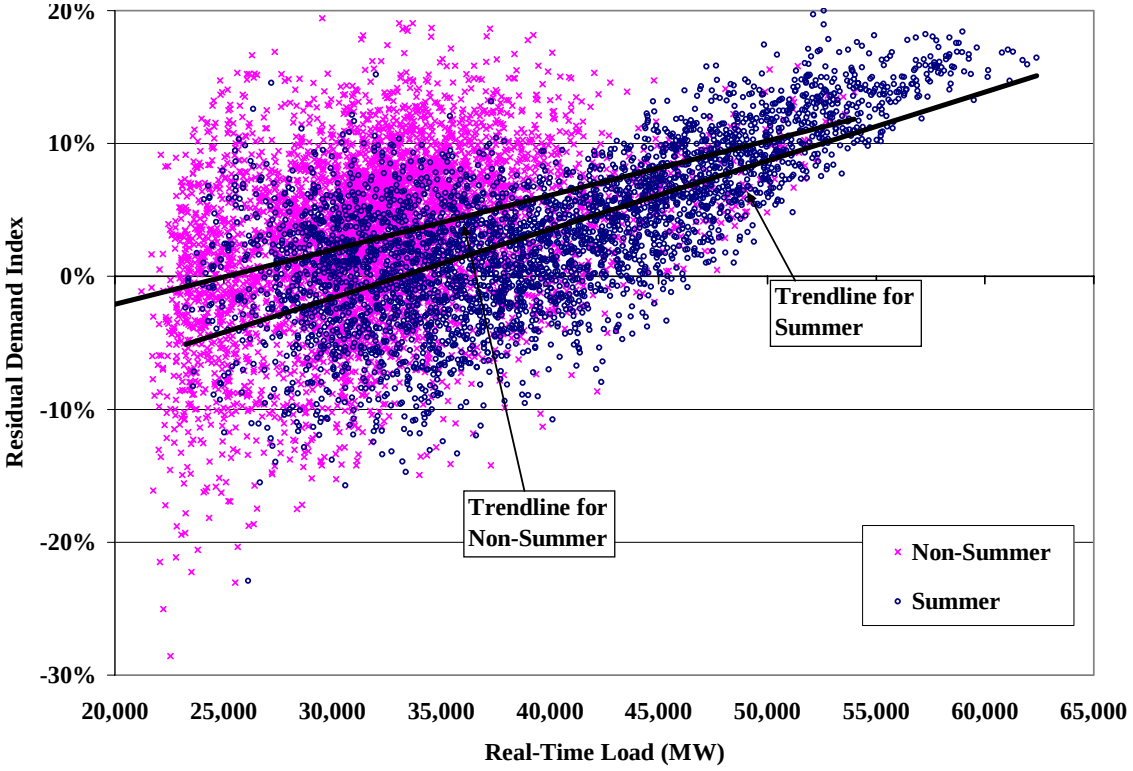
We analyze market structure using the Residual Demand Index ("RDI"), a statistic that measures the percentage of load that could not be satisfied without the resources of the largest supplier. When the RDI is greater than zero, the largest supplier is pivotal (*i.e.*, its resources are needed to satisfy the market demand). When the RDI is less than zero, no single supplier's resources are required to serve the load as long as the resources of its competitors are available.

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

Figure 69 shows the RDI relative to load on an hourly basis in 2007. The data is divided into two groups: (i) hours during the summer months (from May to September) are shown using darker points, while (ii) hours during other months are shown using lighter points. The trend lines for each data series are also shown and indicate a strong positive relationship between load and the RDI. This analysis is done at the QSE level because the largest suppliers that determine the RDI values shown below own a large majority of the resources they are scheduling or

offering. It is possible that they also control the remaining capacity through bilateral arrangements, although we do not know whether this is the case. To the extent that the resources scheduled by the largest QSEs are not controlled or providing revenue to the QSE, the RDIs will tend to be slightly overstated.

Figure 69: Residual Demand Index
2007



The figure shows that the RDI for the summer (i.e. May to September) was usually positive in hours when load exceeded 45,000 MW. During the summer, the RDI was greater than zero in approximately 70 percent of hours. The RDI was typically positive at lower load levels during the spring and fall due to the large number of generation planned outages and less commitment. Hence, although the load was lower outside the summer, our analysis shows that a QSE was pivotal in approximately 71 percent of hours during the non-summer period. It is important to recognize that inferences regarding market power cannot be made solely from this data. Retail load obligations can affect the extent of market power for large suppliers, since such obligations cause them to be much smaller net sellers into the wholesale market than the analysis above would indicate. Bilateral contract obligations can also affect a supplier's potential market power. For example, a smaller supplier selling energy in the balancing energy market and through short-

term bilateral contracts may have a much greater incentive to exercise market power than a larger supplier with substantial long-term sales contracts. The RDI measure shown in the previous figure does not consider the contractual position of the supplier, which can increase a supplier's incentive to exercise market power compared to the load-adjusted capacity assumption made in this analysis.

In addition, a supplier's ability to exercise market power in the current ERCOT balancing energy market may be higher than indicated by the standard RDI. Hence, a supplier may be pivotal in the balancing energy market when it would not have been pivotal according to the standard RDI shown above. To account for this, we developed RDI statistics for the balancing energy market. Figure 70 shows the RDI in the balancing energy market relative to the actual load level.

Ordinarily, the RDI is used to measure the percentage of load that cannot be served without the resources of the largest supplier, assuming that the market could call upon all committed and quick-start capacity[34] owned by other suppliers. Figure 70 limits the other supplier's capacity to the capacity offered in the balancing energy market. When the RDI is greater than zero, the largest supplier's balancing energy offers are necessary to prevent a shortage of offers in the balancing energy market. Figure 71 shows the same data as in Figure 70 except that the balancing energy offers are limited by portfolio ramp constraints in each interval.

[34] For the purpose of this analysis, "quick-start" includes off-line simple cycle gas turbines that are flagged as on-line in the resource plan with a planned generation level of 0 MW that ERCOT has identified as capable of starting-up and reaching full output after receiving a deployment instruction from the balancing energy market.

Figure 70: Balancing Energy Market RDI vs. Actual Load
2007

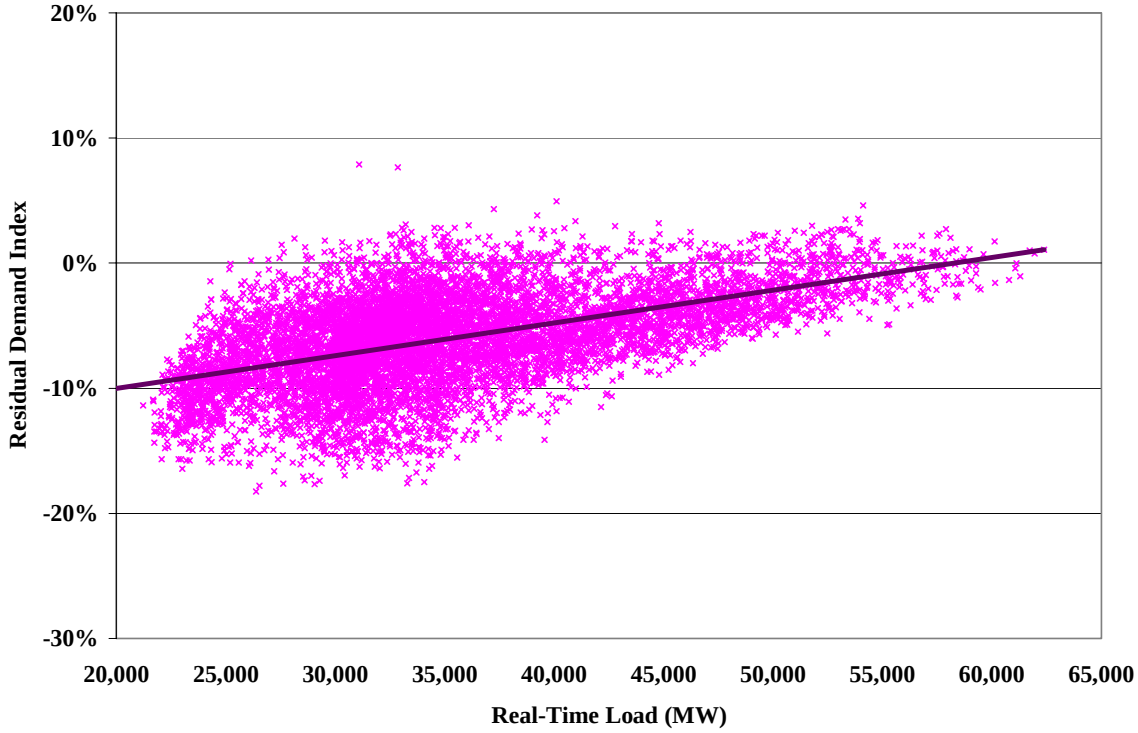
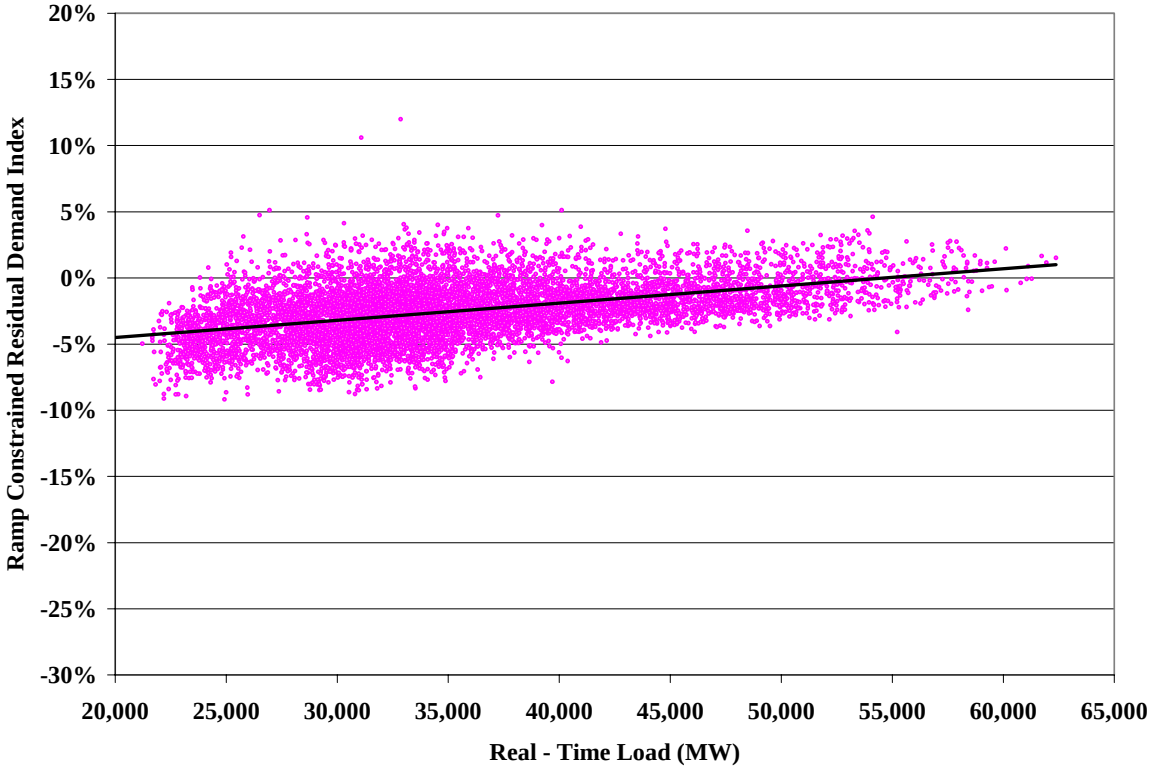
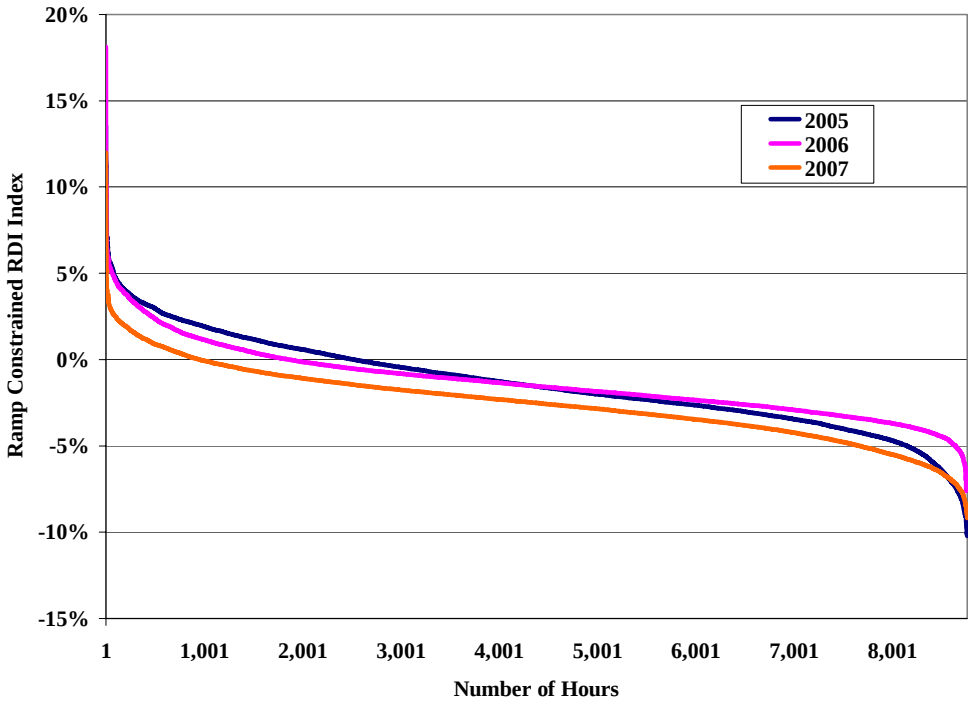


Figure 71: Ramp-Constrained Balancing Energy Market RDI vs. Actual Load
2007



In 2007, the instances when the RDI was positive occurred over a wide range of load levels, from 26 GW to 63 GW. The RDI results for the balancing energy market shown in the preceding two figures help explain how transient price spikes can occur under mild demand while large amounts of capacity are available in ERCOT. The balancing energy market RDI data and trend line for 2007 are similar in shape to 2006, although the frequency of data points that are positive is significantly lower in 2007 than in 2005 and 2006. This difference is highlighted in Figure 72 which compares the balancing energy market RDI duration curves for 2005 -2007.

Figure 72: Ramp-Constrained Balancing Energy Market RDI Duration Curve 2005 - 2007



The frequency with which at least one supplier was pivotal in the balancing energy market *(i.e.*, an RDI greater than zero) has fallen consistently from 29 percent of the hours on 2005 to 21 percent of the hours in 2006 and less than 11 percent of the hours in 2007. These results indicate that the structural competitiveness of the balancing energy market continued to improve in 2007. Figure 73 examines how the balancing energy market RDIs are correlated with balancing energy market prices as adjusted for gas prices in 2007, and Figure 74 shows the same data for 2006.

Figure 73: 2007 Ramp-Constrained Balancing Energy Market RDI vs. Balancing Energy Price Adjusted for Fuel Price

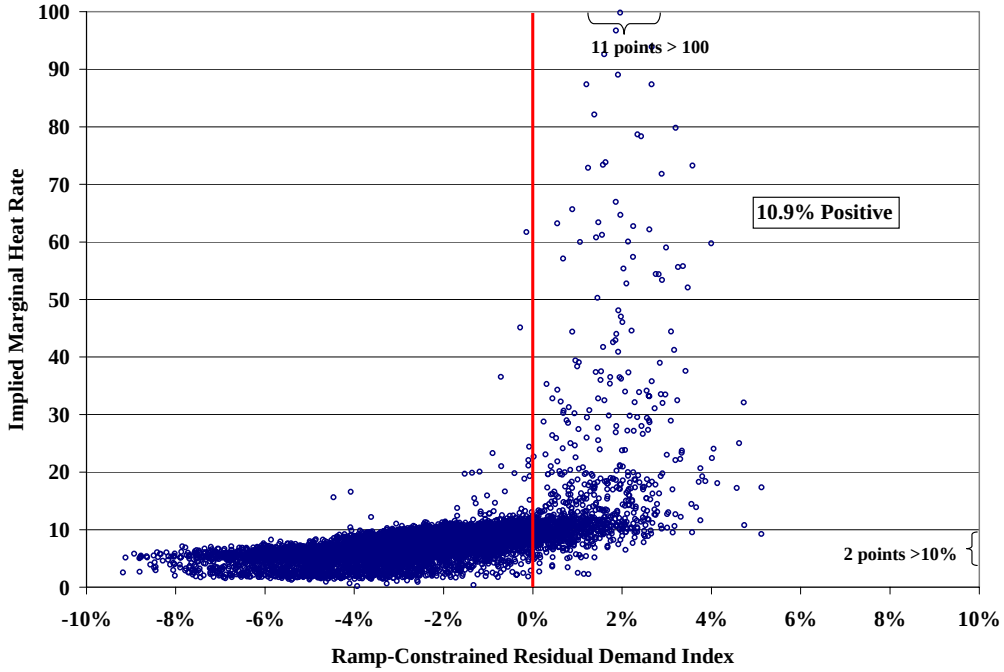
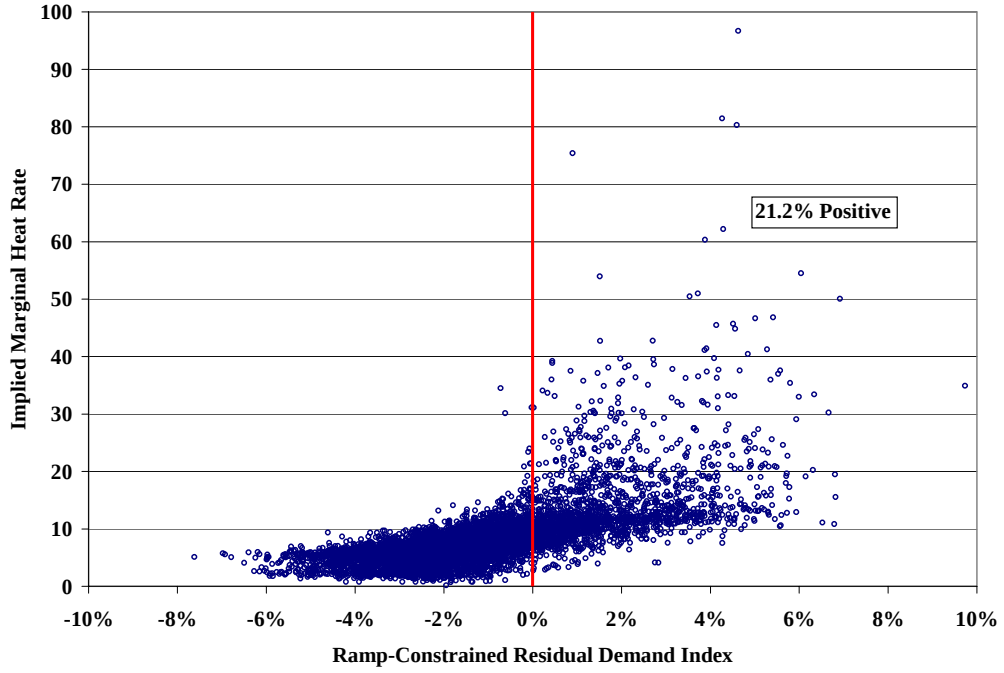


Figure 74: 2006 Ramp-Constrained Balancing Energy Market RDI vs. Balancing Energy Price Adjusted for Fuel Price

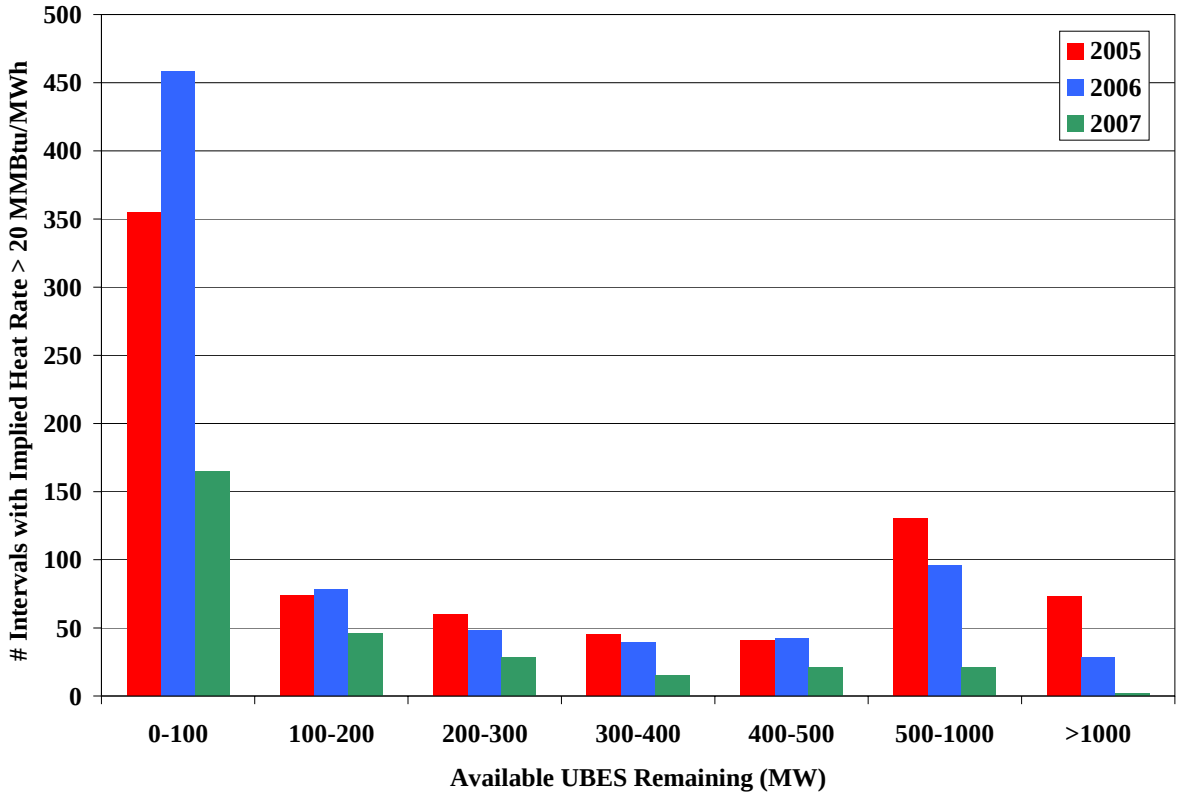


The figures above show a similar relationship between the ramp-constrained balancing energy market RDI and the gas price-adjusted balancing energy market price in 2006 and 2007, with the

rate of change becoming exponentially larger as the balancing energy market RDI enters the positive range. However, Figure 72 reveals that the number of data points with positive ramp-constrained balancing energy market RDIs is over 50 percent less in 2007 than in 2006.

A final structural measure used to evaluate the potential for economic withholding analyzes the number of balancing energy market price spikes compared to the available UBES remaining. If the market is operating competitively, price spikes should occur during shortage and near shortage conditions, and the number of price spikes should reduce significantly as the amount of available surplus energy increases.

Figure 75: Price Spikes vs. Available UBES Remaining



The results in Figure 75 indicate very competitive market outcomes in 2007, with over 92 percent of the price spikes occurring during intervals with less than 500 MW of available UBES remaining.[35] These results show significant improvement over 2005 and 2006 when only 74 and 84 percent, respectively, of the price spikes occurred during intervals with less than 500 MW of

[35] The data in Figure 75 exclude intervals where there was zonal congestion or when non-spinning reserves were deployed.

available UBES remaining. The significant number of price spikes in 2005 and 2006 in intervals with significant available surplus energy (*i.e.*, available UBES remaining greater than 500 MW) give rise to competitive concerns, although the performance improved in 2006 relative to 2005.

B. Evaluation of Supplier Conduct

The previous sub-section presented a structural analysis that supports inferences about potential market power. In this section we evaluate actual participant conduct to assess whether market participants have attempted to exercise market power through physical and economic withholding. In particular, we examined unit deratings and forced outages to detect physical withholding and we evaluate the "output gap" to detect economic withholding.

In a single-price auction like the balancing energy market auction, suppliers may attempt to exercise market power by withholding resources. The purpose of withholding is to cause more expensive resources to set higher market clearing prices, allowing the supplier to profit on its other sales in the balancing energy market. Because forward prices will generally be highly correlated with spot prices, price increases in the balancing energy market can also increase a supplier's profits in the bilateral energy market. The strategy is profitable when the withholding firm's incremental profit is greater than the lost profit from the foregone sales of its withheld capacity.

1. Evaluation of Potential Physical Withholding

Physical withholding occurs when a participant makes resources unavailable for dispatch that are otherwise physically capable of providing energy and that are economic at prevailing market prices. This can be done by derating a unit or designating it as a forced outage. In any electricity market, deratings and forced outages are unavoidable. The goal of the analysis in this section is to differentiate justifiable deratings and outages from physical withholding. We test for physical withholding by examining deratings and forced outage data to ascertain whether the data is correlated with conditions under which physical withholding would likely be most profitable.

The RDI results shown in Figure 69 through Figure 74 indicate that the potential for market power abuses rises as load rises and RDI values become more positive. Hence, if physical withholding is a problem in ERCOT, we would expect to see increased deratings and forced

outages at the highest load levels. Conversely, because competitive prices increase as load increases, deratings and forced outages in a market performing competitively will tend to decrease as load approaches peak levels. Suppliers that lack market power will take actions to maximize the availability of their resources since their output is generally most profitable in these peak periods.

Figure 76 shows the average relationship of short-term deratings and forced outages as a percentage of total installed capacity to real-time load level during the summer months for large and small suppliers. Portfolio size is important in determining whether individual suppliers have incentives to withhold available resources. Hence, the patterns of outages and deratings of large suppliers can be usefully evaluated by comparing them to the small suppliers' patterns.

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

Figure 76: Short-Term Deratings by Load Level and Participant Size
June to August, 2007

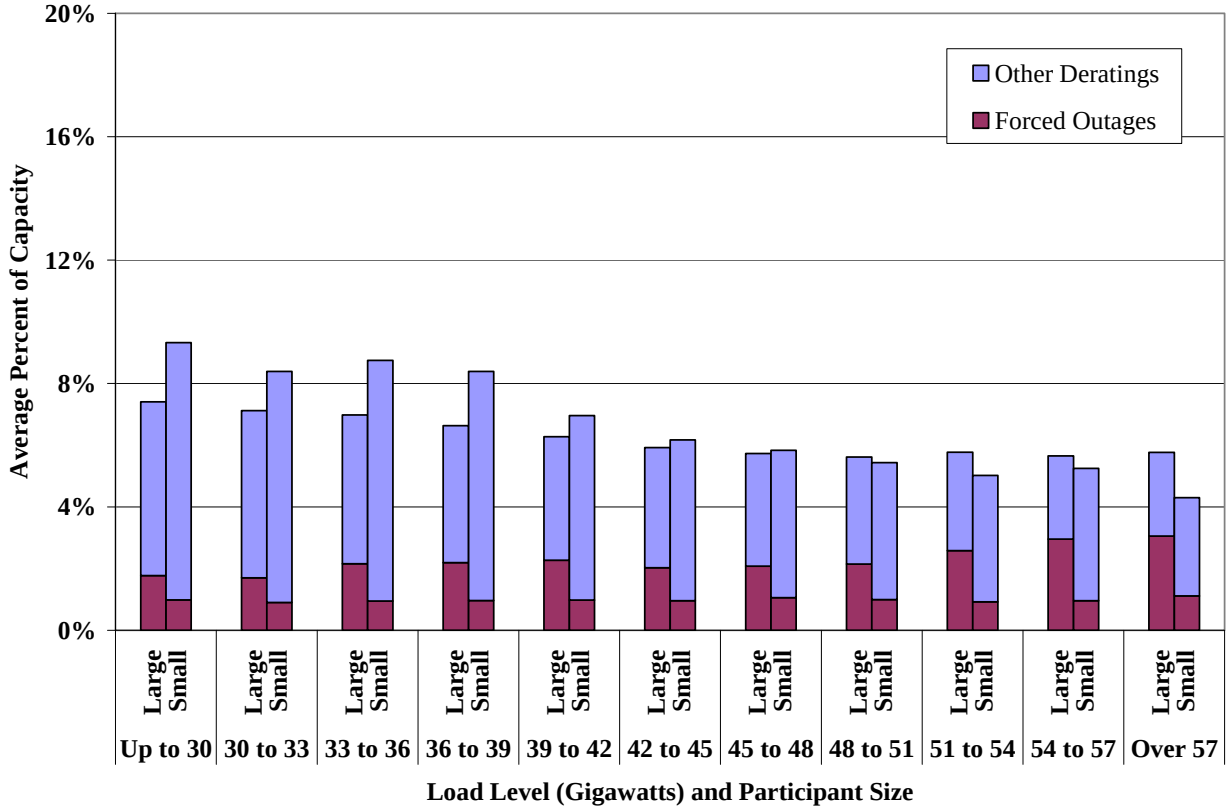


Figure 76 suggests that as electricity demand increases, both large and small market participants tend to make more capacity available to the market. For both large and small suppliers, the short-term derating and forced outage rates decreased from approximately 7 and 9 percent respectively at low demand levels to about 6 and 4 percent respectively at load levels above 57 GW.

Large suppliers have derating and outage rates that are lower than those of small suppliers across the range of load levels up to 48 GW. Furthermore, large suppliers' deratings and outages generally decline as load levels increase. Given that the market is more vulnerable to market power at the highest load levels, these derating patterns do not provide evidence of physical withholding by the large suppliers. The average derating rates for large and small suppliers are approximately 2 and 3 percent lower, respectively, than in 2006 at load levels greater than 51 GW. Further, although the forced outage rate for large suppliers increases slightly at higher load levels, the highest forced outage rate for large suppliers is approximately 3 percent, which is within the range of expected outcomes.

2. Evaluation of Potential Economic Withholding

To complement the prior analysis of physical withholding, this subsection evaluates potential economic withholding by calculating an "output gap". The output gap is defined as the quantity of energy that is not being produced by in-service capacity even though the in-service capacity is economic by a substantial margin given the balancing energy price. A participant can economically withhold resources, as measured by the output gap, by raising the balancing energy offers so as not to be dispatched (including both balancing up and balancing down offers) or by not offering unscheduled energy in the balancing energy market.

Resources can be included in the output gap when they are committed and producing at less than full output or when they are uncommitted and producing no energy. Unscheduled energy from committed resources is included in the output gap if the balancing energy price exceeds the marginal production cost of the energy by at least $50 per MWh. The output gap excludes capacity that is necessary for the QSE to fulfill its ancillary services obligations. Uncommitted capacity is considered to be in the output gap if the unit would have been profitable given published zonal day-ahead bilateral market prices.[36] The resource is counted in the output gap for commitment if its net revenue (market revenues less total cost, which includes startup and operating costs) exceeds the total cost of committing and operating the resource by a margin of at least 25 percent for the standard 16 hour delivery time associated with on-peak bilateral contracts.[37]

As was the case for outages and deratings, the output gap will frequently detect conduct that can be competitively justified. Hence, it is important to evaluate the correlation of the output gap patterns to those factors that increase the potential for market power, including load levels and portfolio size. Figure 77 shows the relationship between the output gap from committed resources and real-time load for all hours during 2007.

[36] Day-ahead bilateral prices are from Megawatt Daily.

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

Figure 77: Output Gap from Committed Resources vs. Actual Load 2007

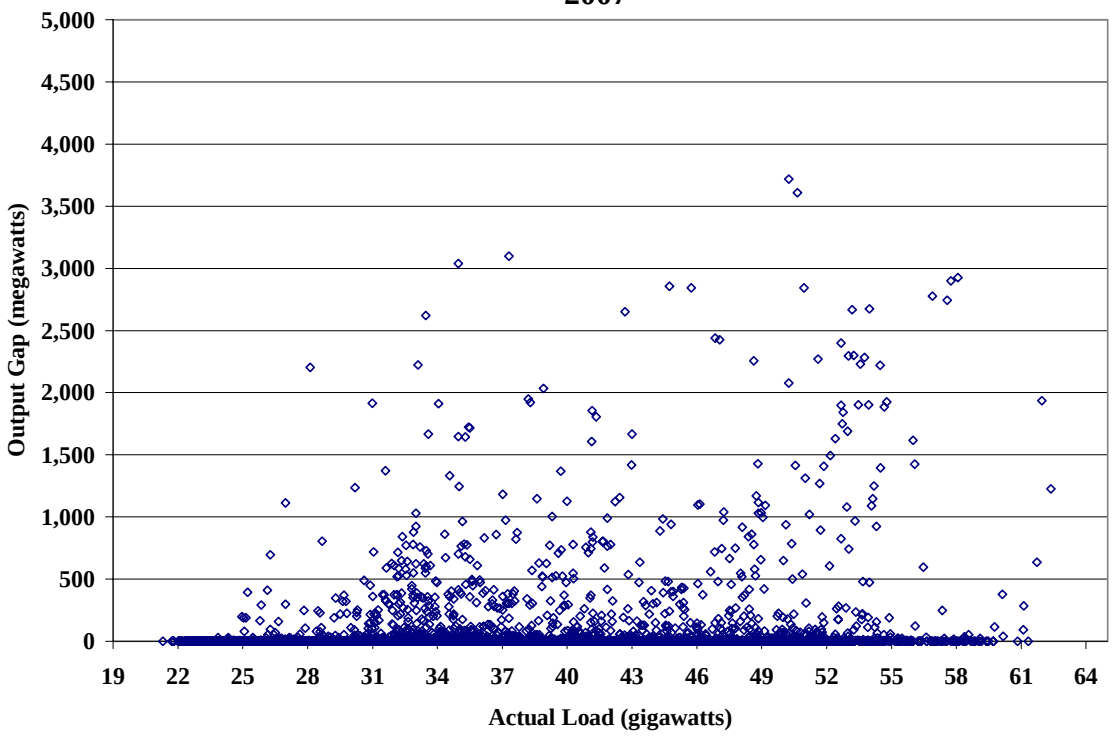


Figure 77 shows that the output gap from committed resources ranged from zero in most hours to a maximum of around 3,700 MW during 2007. As more clearly shown in Figure 78, the average output gap from committed resources rises slightly with real-time demand. This is not surprising given that clearing prices tend to be higher at higher load levels. Many of the high output gap values occurred during transitory price spikes under a wide range of demand levels that make most of the unscheduled energy appear economic. The transitory nature of most of these instances would make a large share of the identified output unavailable due to the resources' ramp limitations. Ramp limitations prevent resources from responding instantaneously to an unpredicted price spike. The next analysis further examines the output gap results by size of supplier and load level.

Figure 78 compares real-time load to the average output gap as a percentage of total installed capacity by participant size. The large supplier category includes the four largest suppliers in ERCOT, whereas the small supplier category includes the remaining suppliers that each controls more than 300 MW of capacity. The output gap is separated into (a) quantities associated with

uncommitted resources and (b) quantities associated with incremental output ranges of committed resources.

Figure 78: Output Gap by Load Level and Participant Size 2007

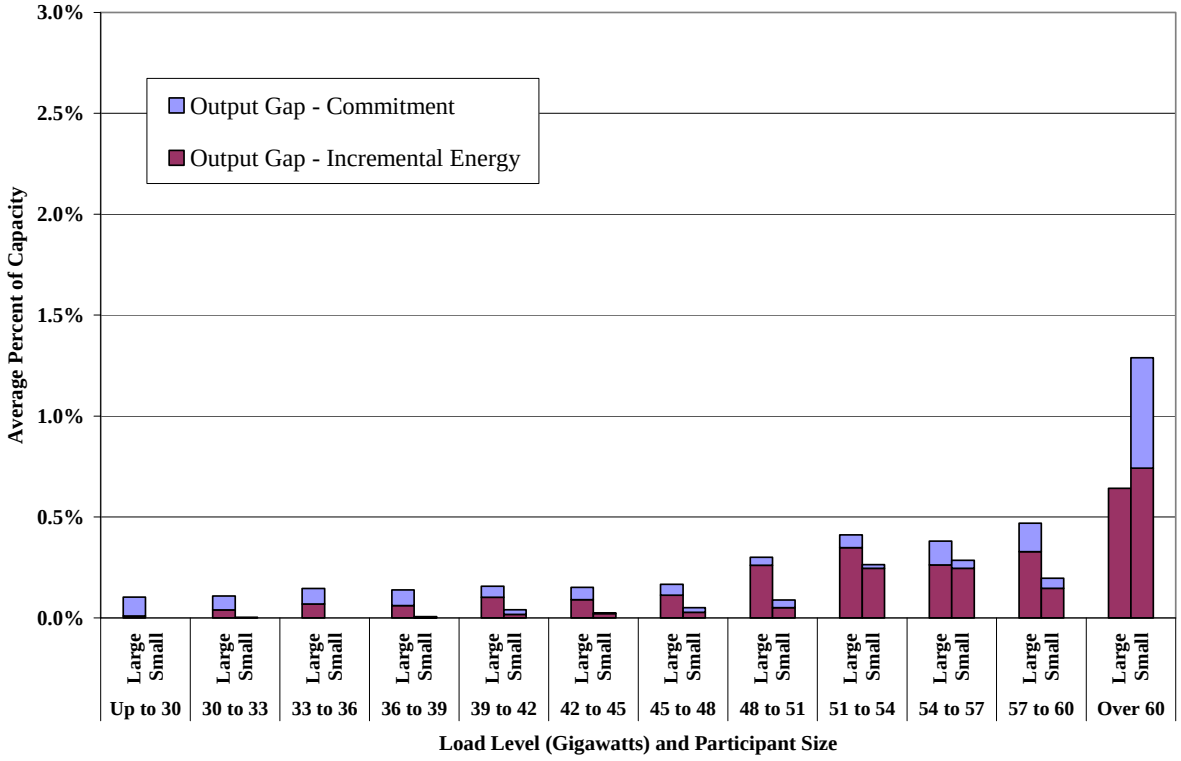


Figure 78 shows that the output gap quantities for incremental energy of large and small suppliers were comparable across all load levels. Overall, the output gap measures in 2007 were comparable with the levels in 2006, with both years showing improvement over 2005.

Overall, based upon the analyses in this section, we find that the ERCOT wholesale market performed competitively in 2007.