POTOMAC ECONOMICS

2006 STATE OF THE MARKET REPORT
FOR THE
ERCOT WHOLESALE ELECTRICITY MARKETS

POTOMAC ECONOMICS, LTD.

ERCOT Independent Market Monitor

August 2007

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 2006. 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. Additionally, balancing energy prices decreased by over 24 percent in 2006 due to lower fuel prices (particularly natural gas) and improved competitive performance of the market. However, the report generally confirms prior findings that the current market rules and procedures are resulting in systematic inefficiencies.

These findings can be found in four 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, which is currently being developed for implementation by 2009.

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 pricing will result in price signals that provide incentives to build new generation where it is most needed for managing congestion and

[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"); and "ERCOT State of the Market Report 2005", Potomac Economics, July 2006 (hereafter "2005 SOM Report").

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

A. Review of Market Outcomes

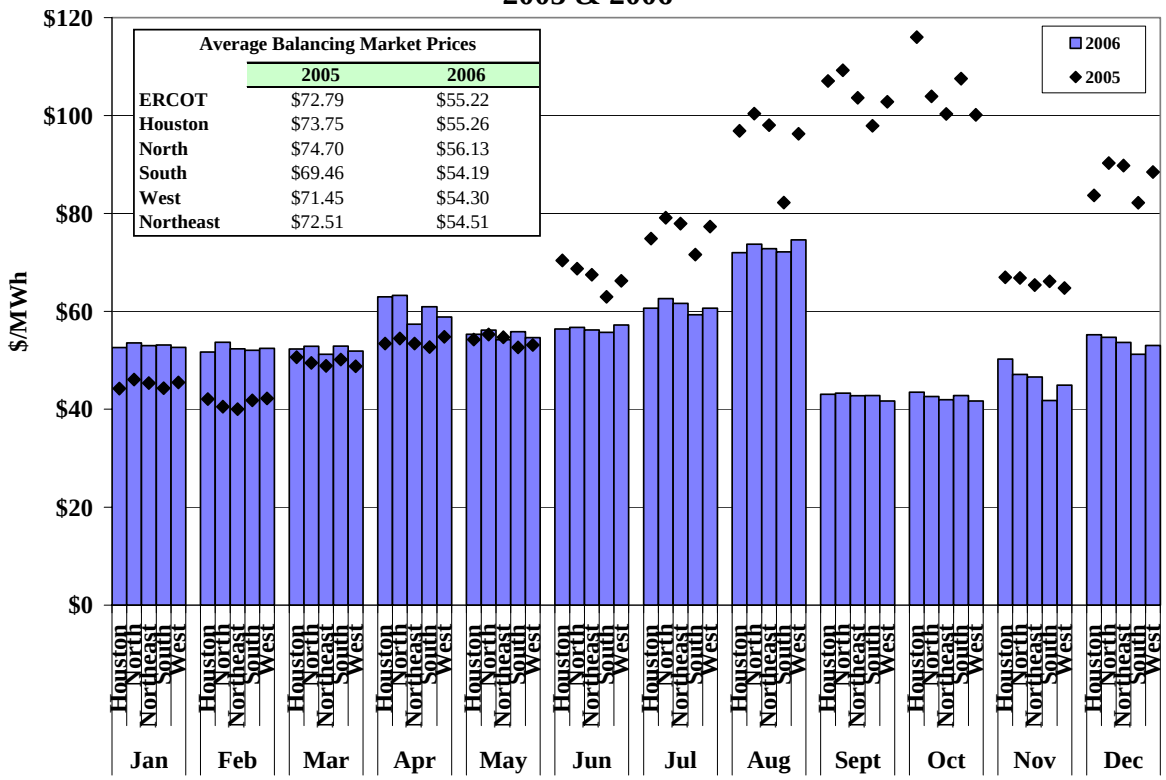
1. Balancing Energy Prices

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 in order 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 over 24 percent lower in 2006 than in 2005, with the latter half of the year showing the largest reductions from 2005. In 2005, natural gas prices began to rise significantly during the summer and remained at high levels through the end of the year. This increase was largely due to the effects of the hurricanes on the productive capability of the Gulf Coast region. However, natural gas prices settled to relatively lower levels in 2006, especially during the latter half of the year. Natural gas is typically the marginal fuel in the ERCOT market. Hence, the changes in energy prices from 2005 to 2006 were largely a function of natural gas price movements.

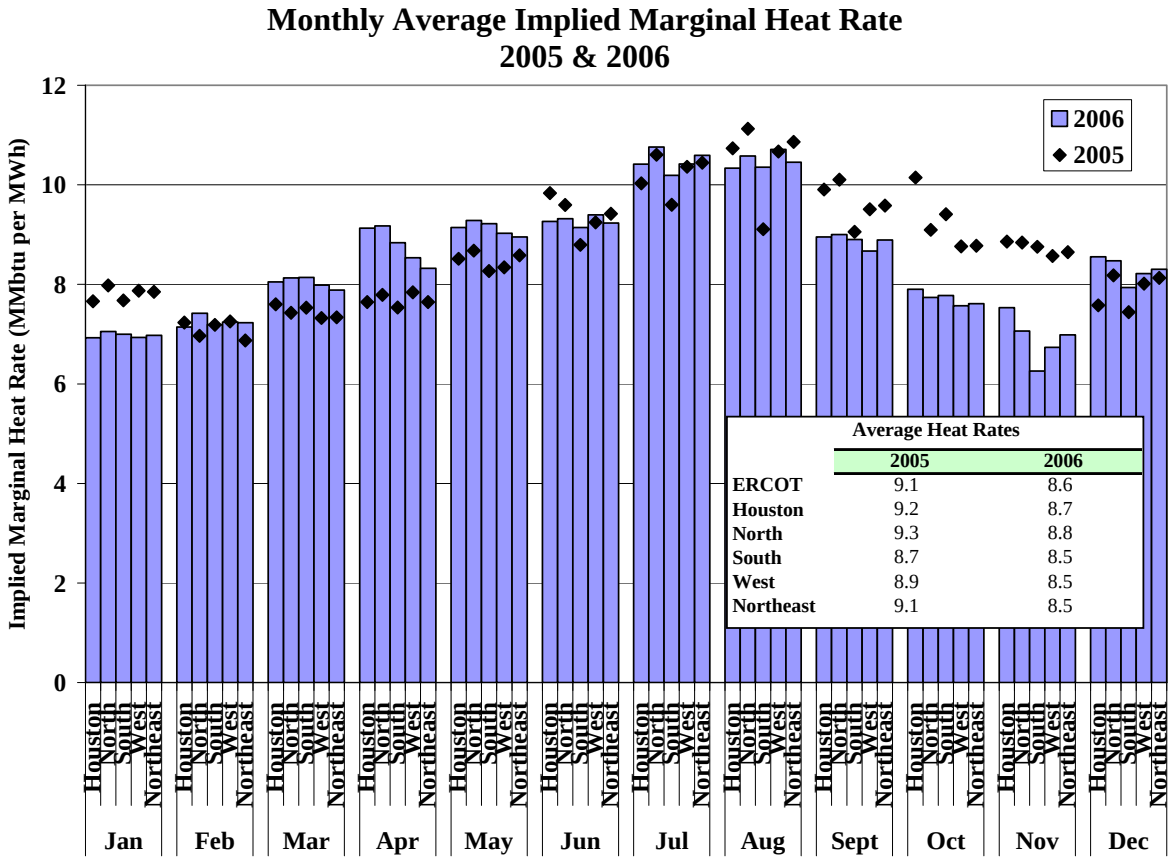
Balancing Energy Market Prices
2005 & 2006



Average Balancing Market Prices

	2005	2006
ERCOT	$72.79	$55.22
Houston	$73.75	$55.26
North	$74.70	$56.13
South	$69.46	$54.19
West	$71.45	$54.30
Northeast	$72.51	$54.51

Although fuel price fluctuations have been the dominant factor driving the decreases in electricity prices in 2006, fuel prices alone do not explain all of the price changes. At least three other factors contributed to price changes in 2006. First, ERCOT demand increased in 2006, while the supply remained relatively static. Second, ERCOT generally committed less excess capacity on a daily basis in 2006. Third, the overall competitive performance of the market improved in 2006 relative to 2005. The first two factors will tend to produce an upward pressure on prices in 2006 relative to 2005, and these factors are discussed in greater detail in Section III of this report. In contrast, the third factor will tend to lower prices and is examined in Section V. To account for changes in fuel prices, the following figure compares the implied marginal heat rate in 2005 and 2006. The implied marginal heat rate is calculated by dividing the balancing energy price by the natural gas price.



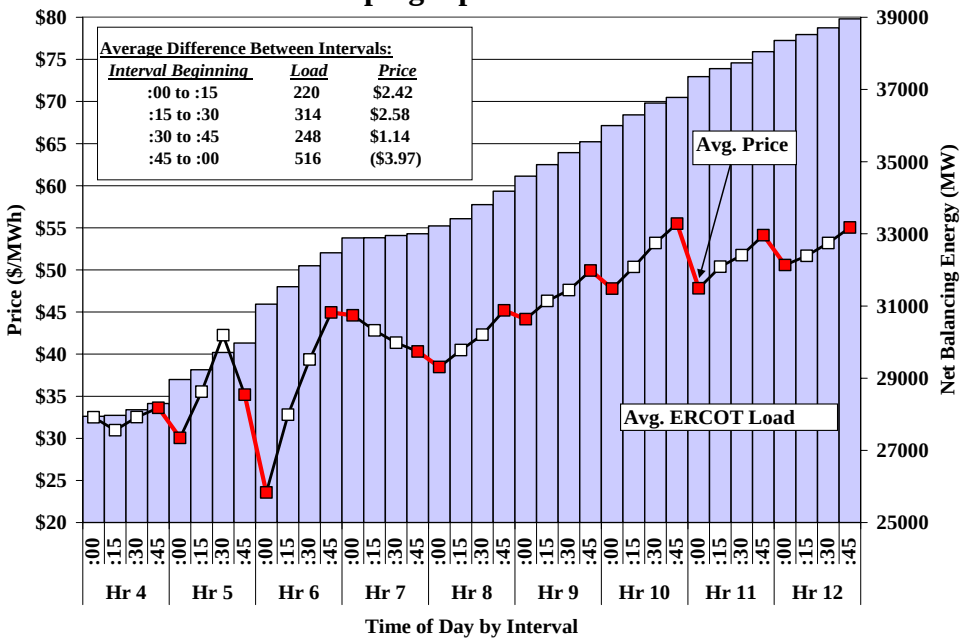
Adjusted for gas price influence, the above figure shows that average implied heat rate for all hours of the year decreased by 5.5 percent from 9.1 in 2005 to 8.6 in 2006. On average, the implied heat rate was lower in 2006 than in 2005 for the months of June through November. With the exception of January, the average implied heat rate for the remaining months was higher in 2006 than in 2005.

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 2006.

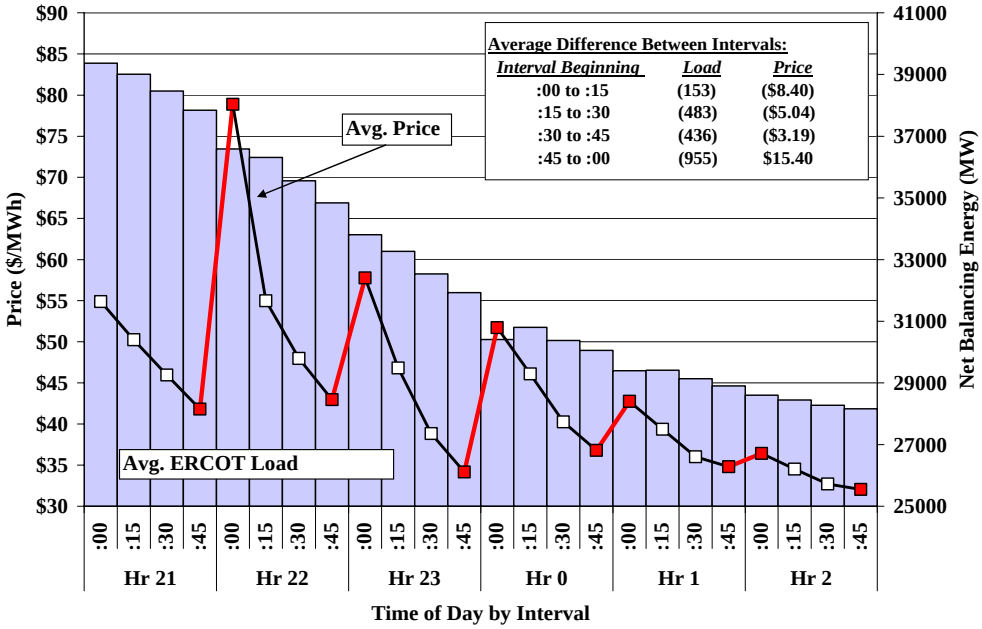
As discussed in prior reports, we continue to observe in 2006 a clear relationship between the net balancing energy deployments and the balancing energy prices. This is not expected in a well-functioning market. This relationship is partly due to the hourly scheduling patterns of most of

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.

Average Balancing Energy Prices and Load by Time of Day
Ramping-Up Hours – 2006



Average Balancing Energy Prices and Load by Time of Day
Ramping-Down Hours – 2006

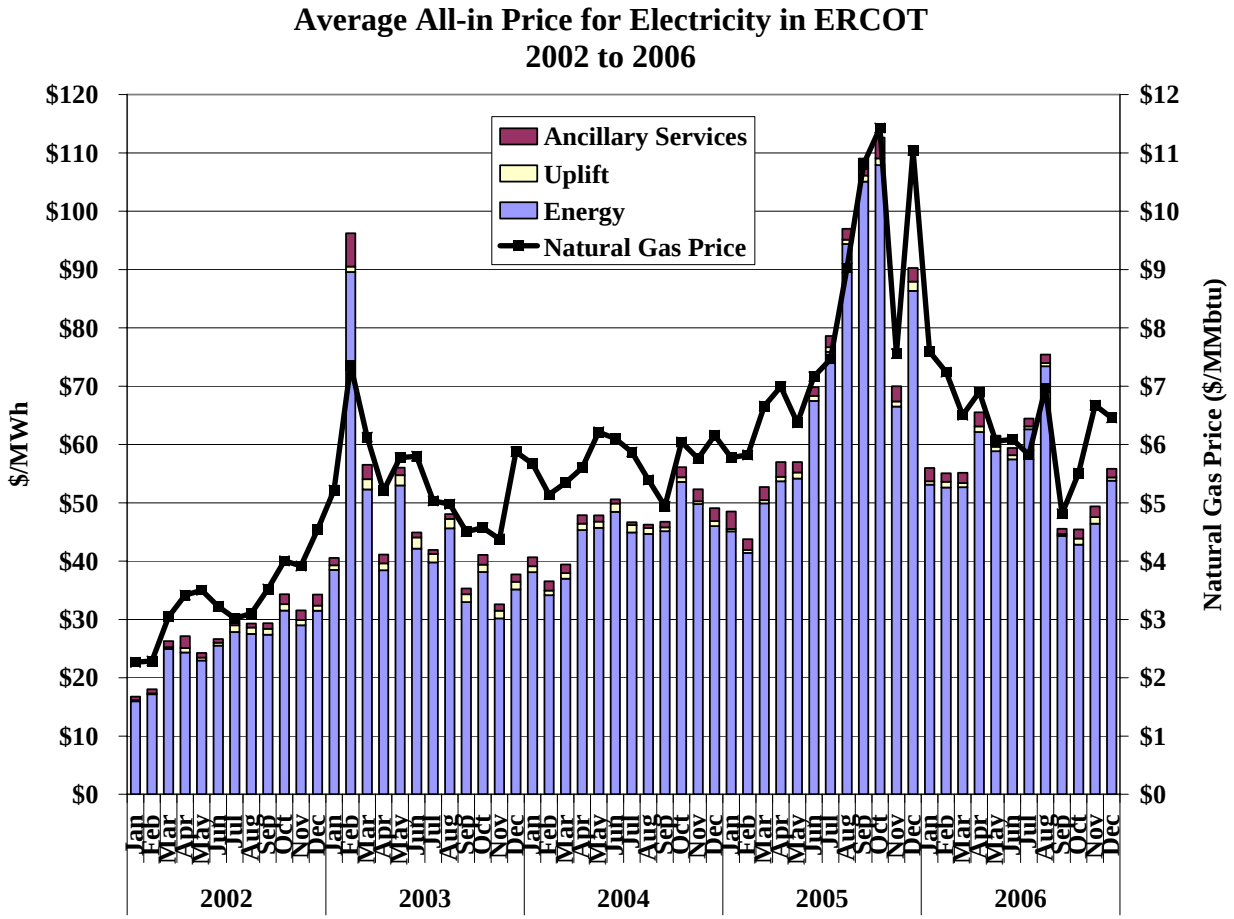


The previous two figures summarize these erratic price patterns by showing the balancing energy prices and actual load in each 15-minute interval during the morning "ramping-up" hours and evening "ramping-down" hours. These pricing patterns raise significant efficiency concerns regarding the operation 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.[2] In prior reports, we have made several recommendations to address the issue under the current zonal design. However, significant modifications to the zonal market design may not be practical at this time given the scheduled implementation of the nodal market by 2009. The nodal market will provide for a comprehensive solution to the operational issues described in this and prior reports.

2. All-In Electricity Prices

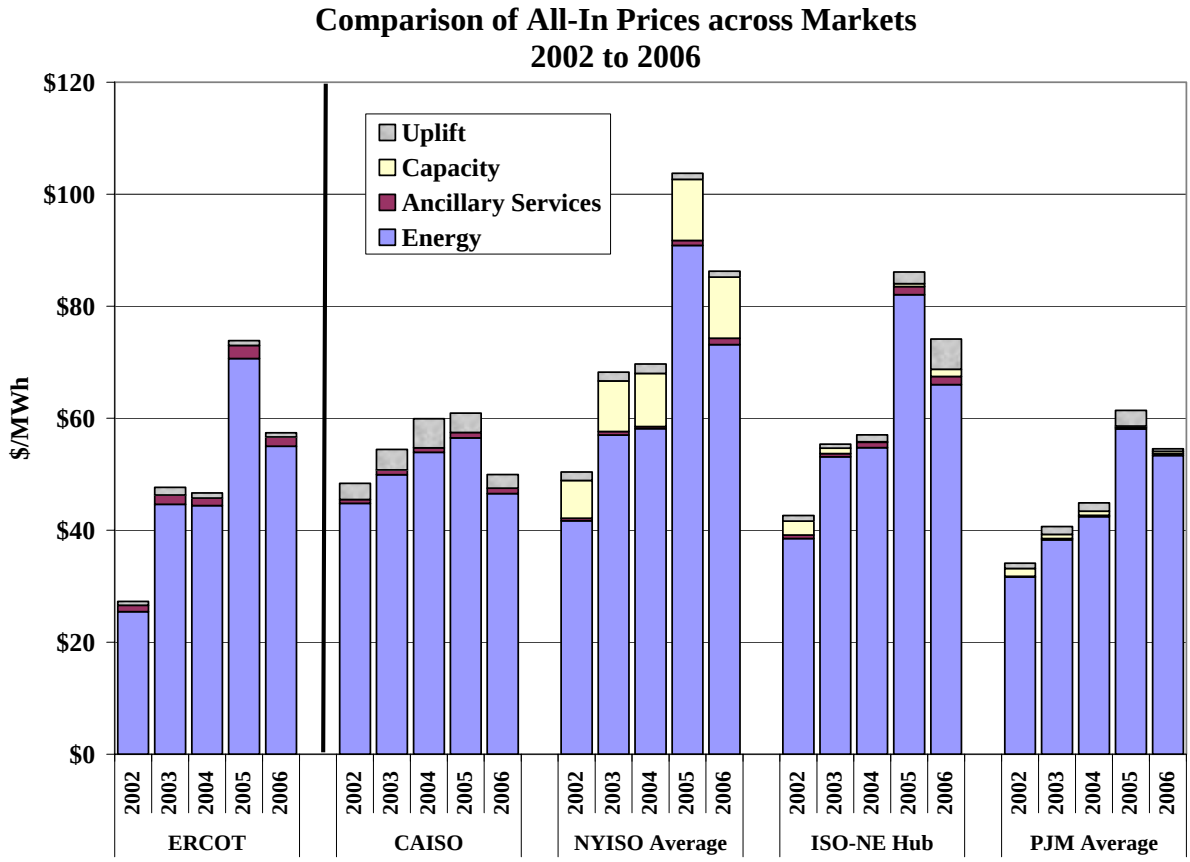
In addition to the costs of energy, loads incur costs associated with operating reserves, regulation, and uplift. The uplift costs include payments for out-of-merit capacity ("OOMC"), Replacement Reserve ("RPRS") out-of-merit energy ("OOME"), and reliability must run agreements ("RMR"). These costs, 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.

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



The figure indicates that natural gas prices were a primary driver of the trends in electricity prices from 2002 to 2006. Natural gas prices increased in 2003 by more than 65 percent from 2002 levels on average while the all-in price for electricity increased by 72 percent. Again, 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.

To provide some 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 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. Although the markets vary substantially in the portion of their generating capacity that is fueled by natural gas, these units are on the margin and setting the wholesale spot prices in a large share of the hours in each of the markets. The largest decreases in electricity prices occurred in ERCOT, indicating natural gas resources are on the margin more frequently in this market than other markets. PJM had the smallest percentage decrease in electricity price in 2006 from 2005. Coal-fired generation is on the margin in a larger share of the hours in PJM, making prices in that market less sensitive to changes in natural gas prices.

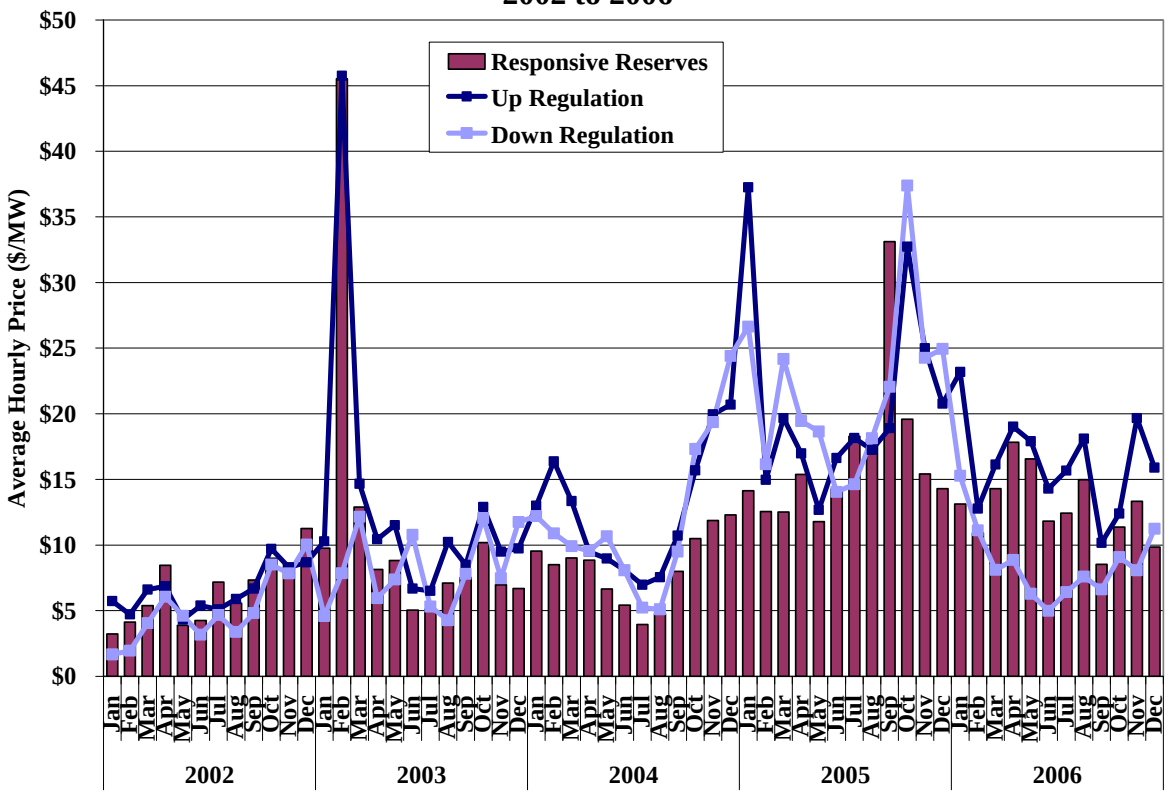
3. Ancillary Services Markets

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 2006.

Ancillary services prices have risen considerably since 2002, peaked in 2005 and dropped in 2006, consistent with long-term trends in natural gas and electricity prices. 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 average prices for regulation and responsive reserve services from 2002 to 2006.

Monthly Average Ancillary Service Prices
2002 to 2006



Although ancillary services prices have generally risen over the last few 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 2006, the requirement was reduced to an average of 1,950 MW during ramping hours and 1,500 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.

Currently, ERCOT's regulation procurement methodologies group regulation procurement quantities into 4 to 6 blocks of hours and procure the same quantity in each block for each day in each month. In late 2006, we initiated discussions with ERCOT to investigate modifications to this methodology that would allow 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. The ERCOT Board approved the changed methodology in June 2007 to be implemented in August 2007. It is expected that this change will reduce the overall quantities of regulation procured over all hours, but may increase the regulation quantities procured in certain hours. This change should result in more efficient procurement of regulation up and down service while maintaining or even improving reliability.

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 by 2009.

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. However, although the functionality will not be implemented at the inception in the nodal market in 2009, we also recommend the development of real-time markets that co-optimize ancillary services and energy to further enhance the efficient dispatch of resources and pricing in real-time.

4. Net Revenue Analysis

A final analysis of the outcomes in the ERCOT markets in 2006 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.

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 2004 to 2006 for four types of units, a natural gas combined-cycle generator, a simple-cycle gas turbine, a coal-fired steam turbine with scrubbers, and a nuclear unit. The net revenue increased significantly from 2002 to 2005, largely due to rising natural gas prices and more frequent price spikes in the balancing energy market.

In contrast to 2005, net revenue was insufficient to support new entry for gas-fired units in 2006, although the net revenue for gas-fired units in 2006 remained significantly higher than years prior to 2005. As in 2005, net revenue for coal and nuclear units remained above the levels required to support new entry. These outcomes were primarily affected by the following factors:

- Although continuing to decline relative to prior years, planning reserve margins in 2006 were approximately 16.5 percent, which is well 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 moderated in 2006, but remained at levels significantly higher than years prior to 2005. Thus, net revenue for coal and nuclear units continued to be at levels sufficient to support new entry.

- The Modified Competitive Solution Method ("MCSM") triggered price adjustments more frequently in 2006. MCSM is a PUCT-approved mechanism that was in effect in 2005 and through September 2006 that provided for an *ex post* reduction to the resulting market prices when all dispatchable balancing energy was exhausted. The average number of MCSM intervals per month almost doubled to over 26 per month in 2006 compared to less than 16 per month in 2005 for the months in which MCSM was in effect.
- The competitive performance of the ERCOT market improved in 2006.

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, MCSM was eliminated by the new rules.

Unlike markets with a long-term capacity market, 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 for demand response and efficient incentives for new investment when required. During non-shortage conditions (*i.e.*, most of the time), the expectation of competitive market outcomes is no different in energy-only than in capacity markets.

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 1000 to 4000 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.

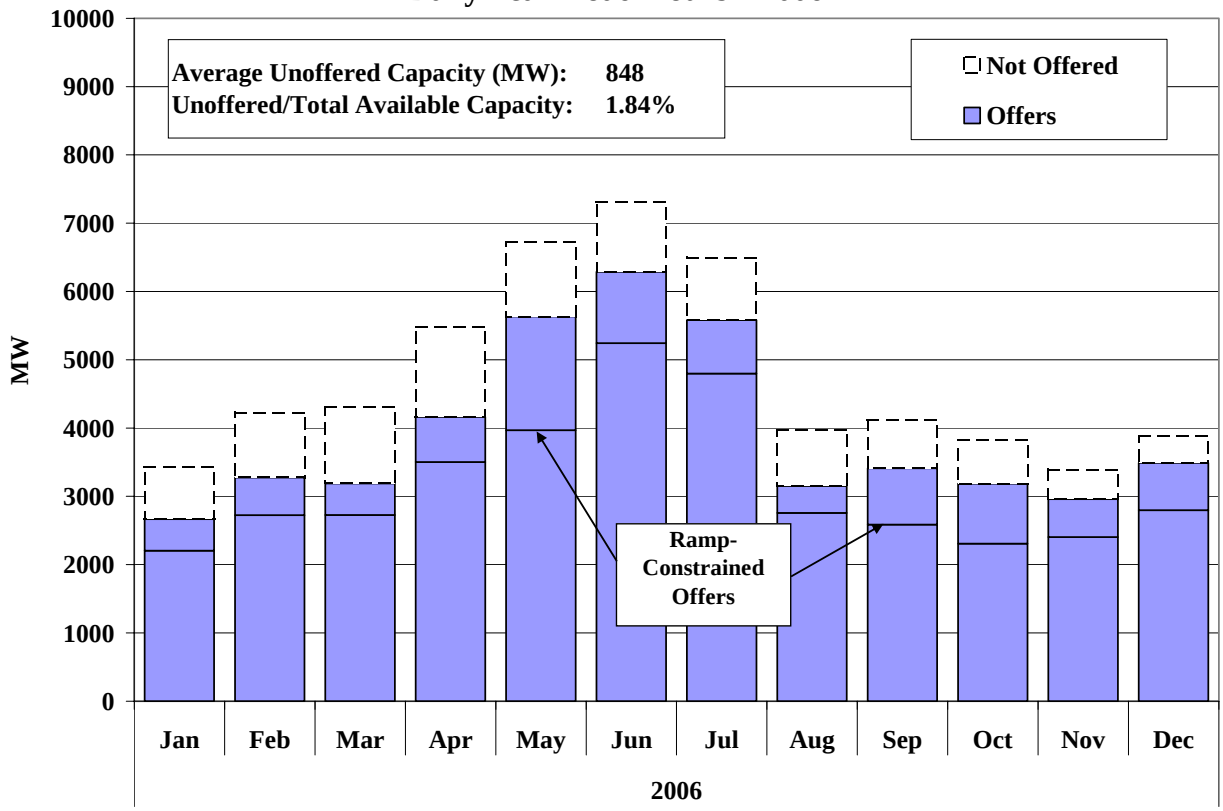
In many cases, the lack of ramp capable resources offered to the balancing energy market results in unnecessary price spikes (as well as large negative prices). There are three aspects of the current market design that inhibit QSEs from fully utilizing the ramp capability of their portfolio. These are: (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 2006 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 analysis in this section differs from similar analyses in prior reports in the following important respect. In prior reports, un-offered capacity calculations included capacity that existed but was not offered. They did not attempt to quantify the amount of un-offered capacity that was actually available, and practicable to offer, given the ERCOT scheduling timelines, operating rules and conditions, and technical or commercial limitations that might limit a QSE's ability to offer capacity in the ERCOT market. In contrast, the approach used for the analysis of un-offered capacity in this section is focused on online,

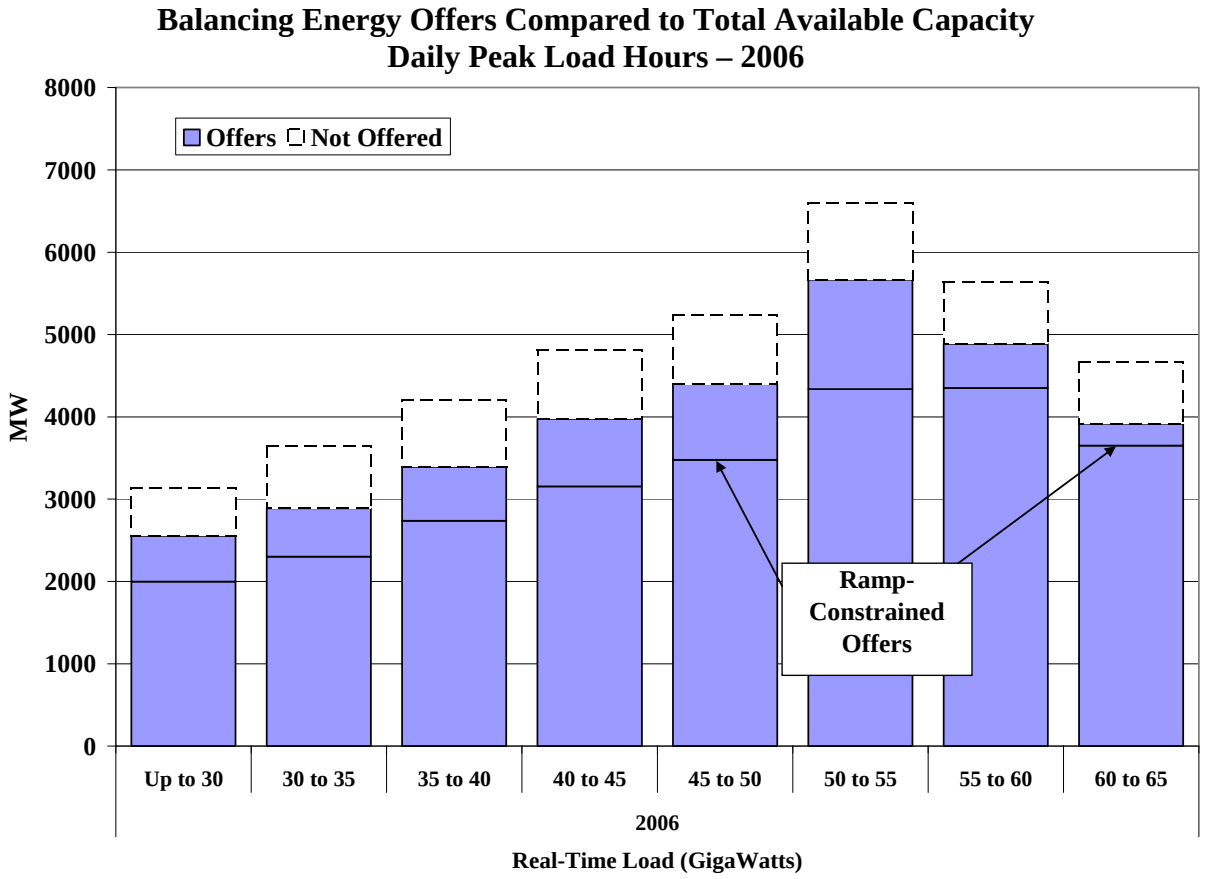
available capacity for which there is a reasonable expectation that the energy can be produced in light of the factors and considerations listed above.

Balancing Energy Offers Compared to Total Available Capacity
Daily Peak Load Hours – 2006



In regard to the residual un-offered capacity, the report identifies several structural impediments that could not be specifically quantified in the figure above. These impediments are largely a function of the zonal market design and serve to explain the quantity of un-offered capacity that could not be specifically quantified in the figure above.

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 2006. 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 2006, the average amount of capacity available to the balancing market increased gradually up to 55 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.

4. Resource Plan Analysis

QSEs submit resource plans to inform ERCOT about which resources they plan to use to satisfy their energy and ancillary services obligations. While QSEs are expected to make their best effort to accurately forecast how they will operate their units, the resource plans are not financially binding. Resource plans are used by ERCOT in some of its reliability assessments before real-time and to make additional commitments to maintain reliability. Therefore, it is important for ERCOT to have accurate information in the resource plans that QSEs submit in order to avoid taking unnecessary and sometimes costly actions to maintain reliability.

Resource plans are not financially binding, yet they are used by ERCOT to make commitment decisions that can have significant cost implications. Hence, a market participant can affect ERCOT's actions and the revenue it receives by submitting resource plans that do not represent efficient generator commitment and dispatch. We analyzed market participants' resource plans to evaluate whether the market protocols may provide incentives for such strategic conduct. Specifically, we evaluated units that are frequently committed out-of-merit or frequently dispatched out-of-merit. Such units receive additional payments from ERCOT and we investigated whether market participants may engage in strategies to increase these payments.

This analysis indicates that most QSEs receiving substantial OOMC or OOME Up payments have commitment and scheduling patterns consistent with the market as a whole. However, some QSEs may delay the commitment of some resources that are frequently committed by ERCOT for reliability purposes or under-schedule resources that are frequently receive OOME Up instructions. In contrast, the analysis of units that are frequently called on for OOME Down indicates that these units are generally scheduled in a manner similar to the market as a whole.

The incentives for participants to submit resource plans that do not reflect anticipated real-time operations stem from the lack of nodal prices to signal the value of capacity and energy in local areas. In the absence of nodal prices, market participants may act strategically to garner additional uplift payments.

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 2006, the load served by ERCOT reached a peak of over 63 GW.[3] This was a relatively significant increase over previous years when the peak was approximately 60, 59, and 61 GW in 2003, 2004 and 2005, respectively. Changes in these 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, as evidenced by the rolling blackout events of April 17, 2006.

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 76 percent of generation capacity in ERCOT as a whole, and 86 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 more than 8 GW was mothballed during 2006 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 2006 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

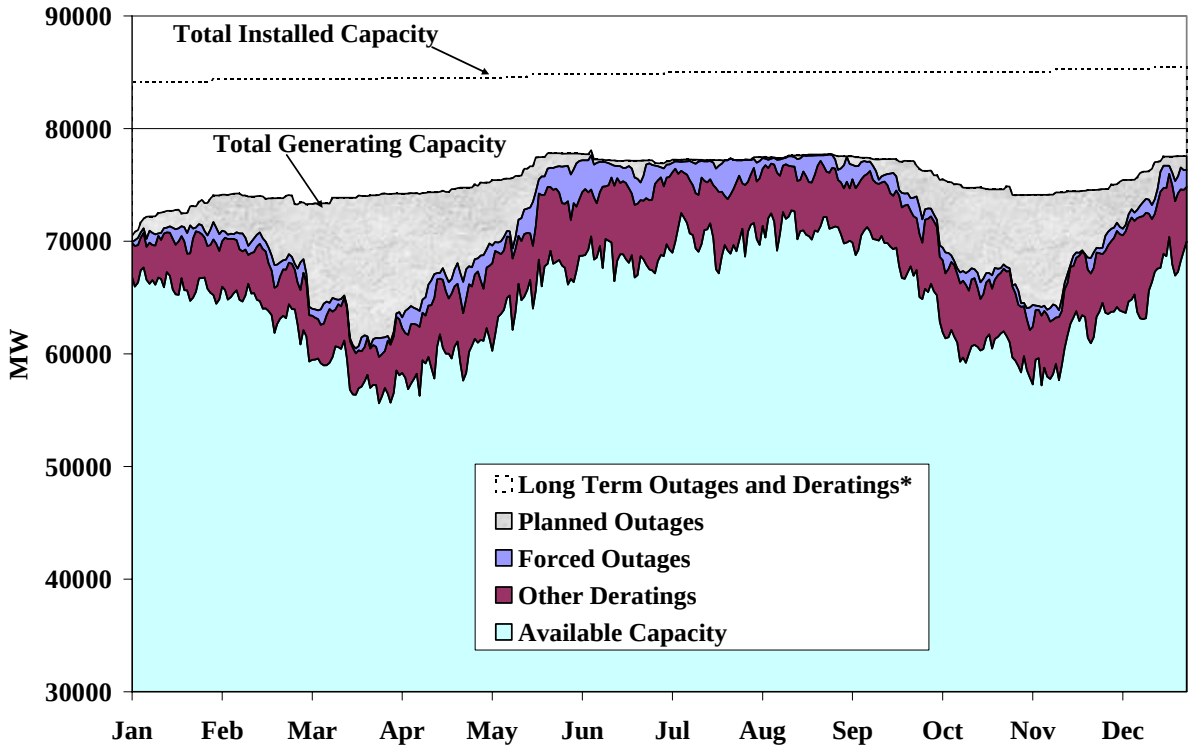
[3] This value is the total load to be served in real-time as represented in ERCOT's Scheduling, Pricing and Dispatch software (including transmission and distribution losses), and may differ from settlement values.

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
2006



* 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 deratings fluctuated between 6 GW and 13 GW. These long-term deratings reduce the effective resource margins in ERCOT from the levels reported above. Most of these deratings reflect:

- Resources out-of-service for extended periods due to maintenance requirements;
- Resources out-of-service for economic reasons (*e.g.*, mothballed units);
- Cogeneration resources typically used for purposes other than electricity generation; or
- Output ranges on available generating resources that are not capable of producing up to the full installed capability level.

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 "other deratings" shown in the figure ranged from an average of 7 percent during the summer in 2006 to as high as 12 percent in other months. These deratings include outages not reported or correctly logged by ERCOT and natural deratings due to high ambient temperature conditions and other factors. The overall pattern of outages and deratings is consistent with competitive expectations and does not raise significant concerns.

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

The report finds that the excess on-line capacity during daily peak hours on weekdays averaged 2,927 MW in 2006, which is approximately 8 percent of the average load in ERCOT. This is a significant decrease from the average of 4,313 MW in 2005 and 6,627 MW in 2004. These decreases can be attributed in part to the continued increase in ERCOT load with a relatively static available supply, fewer quick-start gas turbines that were qualified to provide balancing

energy, and a continuation of the trend from previous years of ERCOT committing fewer units via OOMC instructions and RMR.

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 planned for implementation by 2009 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 2006, 1,985 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 1,835 MW at the end of 2005. Currently, LaaRs are permitted to supply up to 1,150 MW (50 percent) of the responsive reserves requirement. Although the participants with LaaRs resources are qualified to provide non-spinning reserves and up balancing energy in real-time, in 2006, they provided only about one percent of non-spinning reserves and none of the balancing energy. 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. In addition, prices in the balancing energy market have not been high enough to attract load participation in that market. 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 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

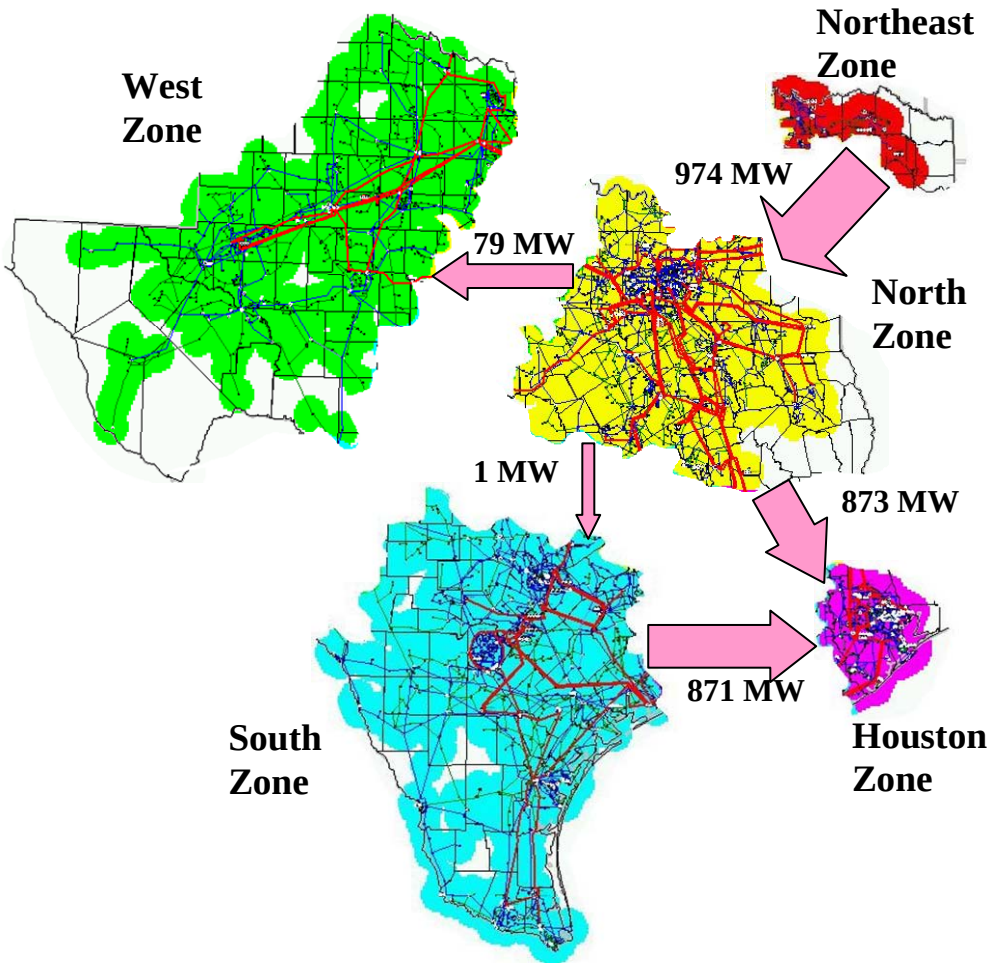
One of the most important functions of any electricity market is to manage the flows of power over the transmission network, limiting additional power flows over transmission facilities when

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 in order 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 five zone-based locations and six transmission interfaces. The transmission interfaces are referred to as Commercially Significant Constraints ("CSCs"). The following figure shows the average flows modeled in SPD during 2006 over each of these CSCs.

Average Modeled Flows on Commercially Significant Constraints
2006



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 79 MW, which is equivalent to saying that the average for the West-to-North CSC was *negative* 79 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, such as the North to West CSC that was implemented by ERCOT in 2005 to better manage congestion over that path.
- Based on modeled flows, Houston is a significant importer while the Northeast Zone and the South Zone export significant amounts of power.

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. In order 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 decreased considerably to $69 million in 2006, which reflects a decrease of $50 million from 2005. This increase was the result of less frequent congestion on the South-to-Houston, North-to-Houston, and South-to-North CSCs, as well as lower overall prices.

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 decreased considerably to $7 million in 2006, down from $38 million in 2005. This reduction was primarily due to decreased interzonal congestion in 2006 and improved accuracy in the quantity of TCRs sold in the monthly auction.

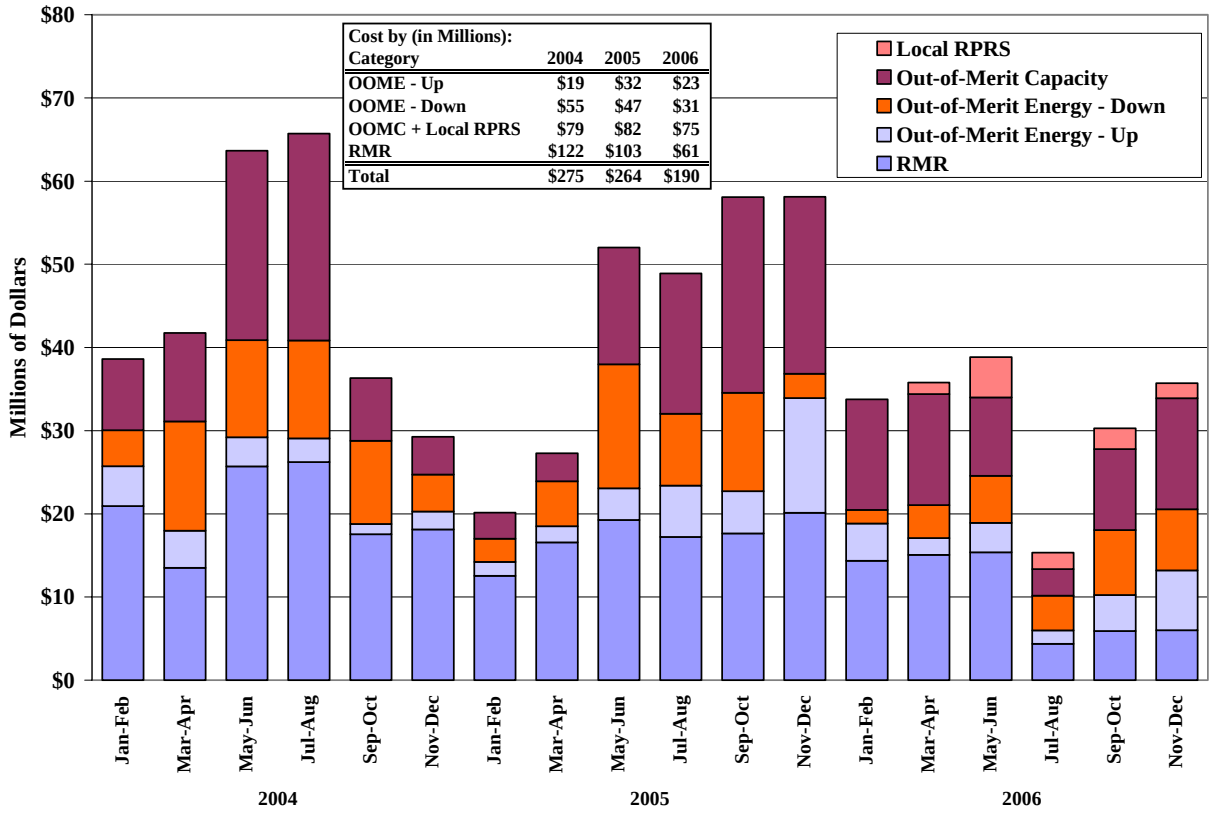
In a perfectly efficient system with no uncertainty, the average congestion cost in real-time should equal the auction price of the congestion rights. In the real world, however, we would expect only reasonably close convergence with some fluctuations from year to year due to uncertainties. In 2005, the annual and monthly TCR auctions substantially under-valued the TCRs in comparison to the balancing market congestion. In contrast to 2005, market participants

over-estimated the annual value of congestion on the South to North, South to Houston, and North to Houston CSCs in 2006. 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 2006.

Expenses for Out-of-Merit Capacity and Energy
2004-2006



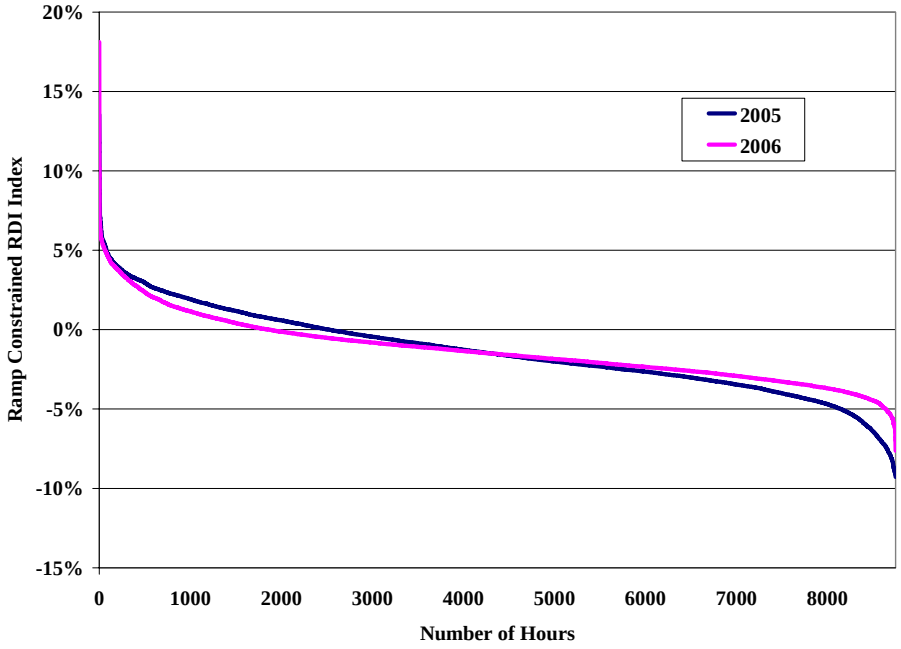
The results in the figure above show that overall uplift costs for RMR units, OOME units, and OOMC/Local RPRS units were relatively consistent between 2004 and 2005. The costs decreased by $74 million in 2006 from $264 million to $190 million, a reduction of 28 percent. There were substantial reductions to RMR cost due to the expiration of RMR agreements in 2006, which accounts for $42 million of the $74 million decrease from 2005 to 2006. Total OOME Up and OOME Down costs also decreased from $79 million in 2005 to $54 million in 2006, a reduction of 32 percent. This reduction is likely due to the continued improvements to the ERCOT transmission system resulting in less frequent local congestion, and the introduction of an enhanced replacement reserve procurement process by ERCOT in 2006.

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 2006 compared to 2005. The following figure shows the ramp-constrained balancing energy market Residual Demand Index ("RDI") duration curves for 2005 and 2006. 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 & 2006



In 2006, there were 1,861 hours (21.2 percent) when the balancing energy market RDI was greater than zero, which means a supplier was pivotal in the balancing energy market 21.2 percent of the time in 2006. In contrast, there were 2,525 hours (28.8 percent) when the balancing energy market RDI was positive in 2005. Hence, the frequency with which a supplier was pivotal in the balancing energy market decreased 26 percent in 2006 indicating that the overall competitiveness of the balancing energy market improved in 2006. Among other factors, this decrease can be attributed to an average reduction in up balancing energy deployments in 2006, which was influenced by the existence of the under-scheduled charges associated with the replacement reserve market.

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 physical and economic withholding in order to further evaluate competitive performance of the ERCOT market. Withholding patterns were examined relative to the level of demand and the size of each supplier's portfolio. Based on the analyses conducted in this area, the report found the overall output gap for both large and small suppliers was reduced considerably in 2006 as compared to 2005. Overall, we find that the competitive performance of the market improved in 2006.

I. REVIEW OF MARKET OUTCOMES

A. Balancing Energy Market

1. Balancing Energy Prices During 2006

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 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).[4] 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 2006.

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 2005 and 2006.[5] Balancing energy market prices were 24 percent lower in 2006 than in 2005, with the latter half of the year showing the largest reductions from 2005.

In 2005, natural gas prices began to rise significantly during the summer and remained at high levels through the end of the year. This increase was largely due to the effects of the hurricanes on the productive capability of the Gulf Coast region. However, natural gas prices settled to

---

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

5 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).

relatively lower levels in 2006, especially during the second half of the year. Natural gas is typically the marginal fuel in the ERCOT market. Hence, the changes in energy prices from 2005 to 2006 were largely a function of natural gas price movements.

Figure 1: Average Balancing Energy Market Prices 2005 & 2006

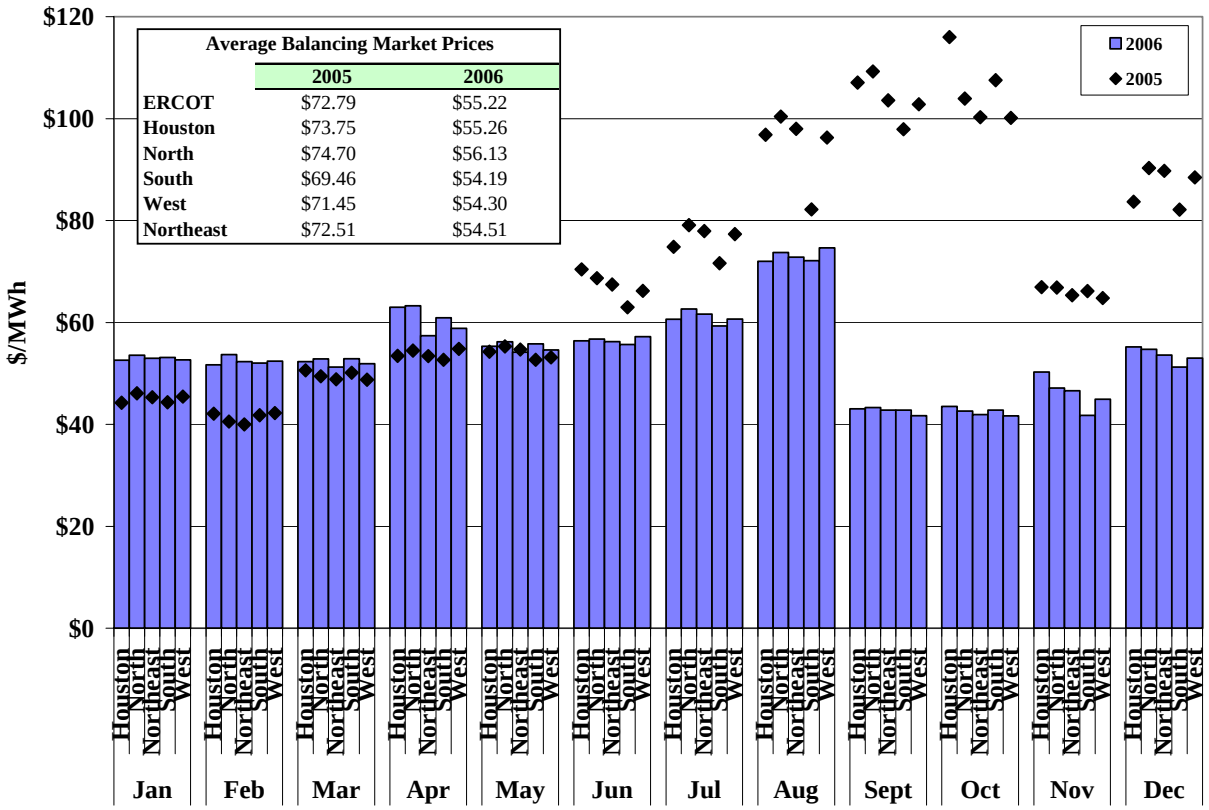


Figure 1 also shows that transmission congestion between zones decreased in ERCOT during 2006. The difference between the average North and South zones prices was approximately 3.5 percent in 2006 as compared to 7.5 percent in 2005. In individual months, the zonal price difference was also much smaller in 2006. For example, the average North zone price exceeded the South zone by approximately $18 per MWh in August 2005, while the difference was only $0.68 per MWh in August 2006.

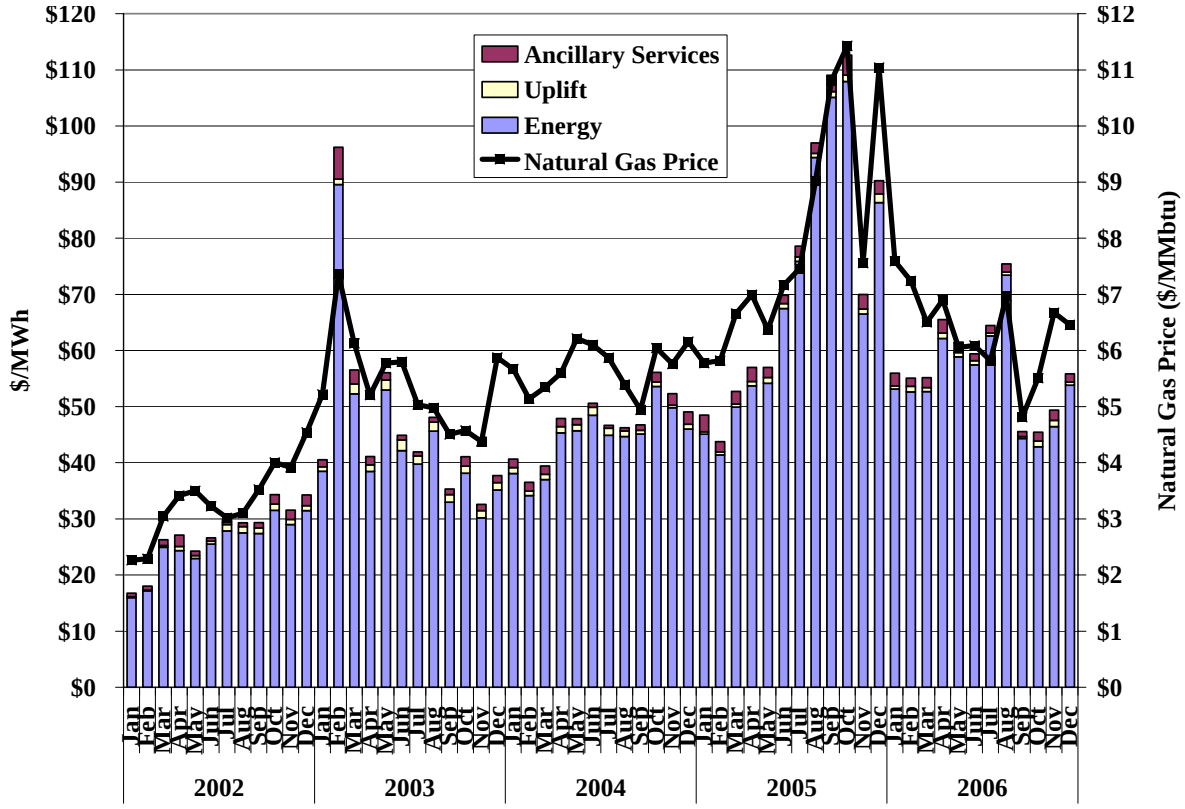
The next analysis evaluates the total cost of serving load in the ERCOT market. In addition to the costs of energy, loads incur costs associated with operating reserves, regulation, and

"uplift".[6] 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 2002 to 2006.

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 2002 to 2006



6 As discussed more below, 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.

Figure 2 indicates that natural gas prices were a primary driver of the trends in electricity prices from 2002 to 2006. 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. 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. The larger increase in electricity prices and the higher number of price spikes in 2005, as discussed later in this subsection, was due, in part, to certain participant conduct that is the subject of a PUCT enforcement proceeding. 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.

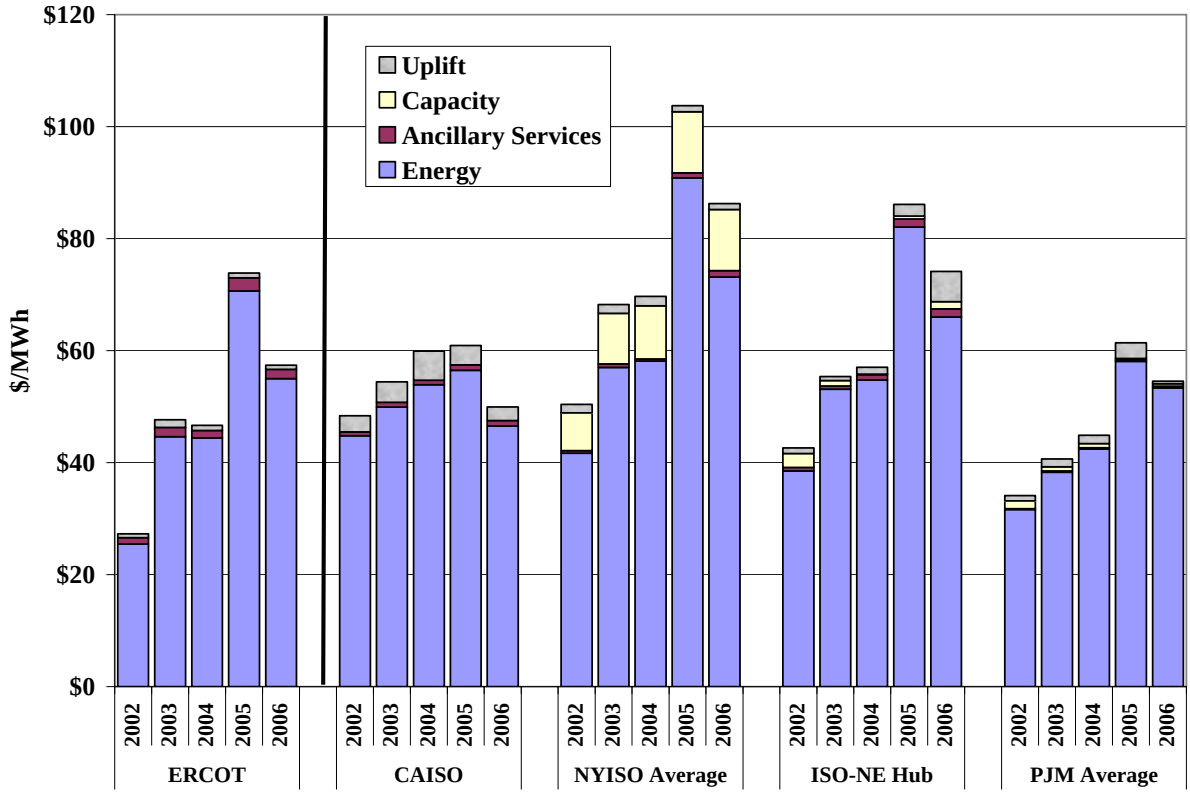
Although fuel price fluctuations have been the dominant factor driving the decreases in electricity prices in 2006, fuel prices alone do not explain all of the price changes. At least three other factors contributed to price changes in 2006. First, ERCOT demand increased in 2006, while the supply remained relatively static. Second, ERCOT generally committed less excess capacity on a daily basis in 2006. Third, the overall competitive performance of the market improved in 2006 relative to 2005. The first two factors will tend to produce an upward pressure on prices in 2006 relative to 2005, and these factors are discussed in greater detail in Section III of this report. In contrast, the third factor will tend to lower prices and is examined 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 2005 to 2006, a 30 percent decrease in ancillary services costs result in a one percent decrease in the all-in price for electricity. Ancillary services prices began to decrease in late 2005 and remained lower throughout 2006. Generally, the ancillary services prices coincided with price movements in the balancing energy market, which is to be expected since the energy and ancillary services 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 2002 to 2006

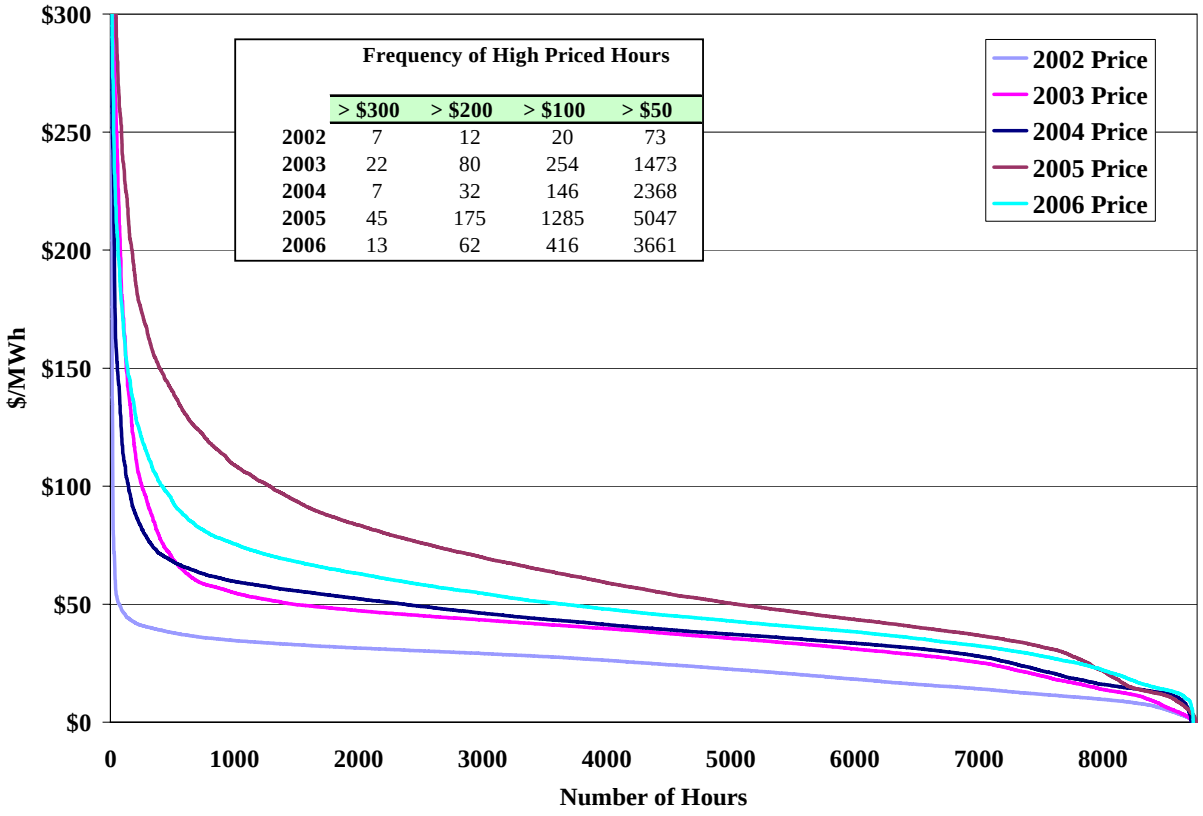


Wholesale electricity markets in the U.S. experienced substantial increases in energy prices from 2002 to 2003 and from 2004 to 2005 due to increased fuel costs. In 2006, energy prices in the U.S. dropped in every region due to decreased fuel costs. Although the markets vary substantially in the portion of their generating capacity that is fueled by natural gas, these units are on the margin and setting the wholesale spot prices in a large share of the hours in each of the markets. The largest decreases in electricity prices occurred in ERCOT, indicating natural gas resources are on the margin more frequently in this market than other markets. PJM had the smallest percentage decrease in electricity price in 2006 from 2005. Coal-fired generation is on the margin in a larger share of the hours in PJM, making prices in that market less sensitive to changes in natural gas prices.

Figure 4 presents price duration curves for the ERCOT balancing energy market in each year from 2002 to 2006. 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
2002 to 2006



Frequency of High Priced Hours

	> $300	> $200	> $100	> $50
2002	7	12	20	73
2003	22	80	254	1473
2004	7	32	146	2368
2005	45	175	1285	5047
2006	13	62	416	3661

The figure shows that, with the exception of 2005, balancing energy prices were higher in 2006 than in prior years. Balancing energy prices exceeded $50 in more than 3,000 hours in 2006 compared to more than 5,000 hours in 2005, and approximately 2,000 hours in 2003 and 2004. These large year-to-year changes reflect the effects of higher fuel prices, which impact electricity prices in a broad range of hours. Higher natural gas prices raise the marginal production costs of the generating units that set the prices in the balancing energy market in a large share of the intervals.

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 2002 and 2006 at the highest price levels.

For example, 2003 experienced considerably more price spikes (*e.g.*, prices higher than $300) than 2004 or 2006 even though prices were higher on average in 2004 and 2006. The largest number of price spikes and the highest average price of any year from 2002 to 2006 occurred in 2005. To better observe the highest-priced hours during 2004 and 2006, the following analysis focuses on the frequency of price spikes in the balancing energy market. Figure 5 shows average prices and the number of price spikes in each month of 2004 and 2006. 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).

As the figure shows, the number of price spikes increased sharply after August 2004. There was an average of 38 price spike intervals per month in 2004 (each month has over 2,900 intervals). The number of price spike intervals more than quadrupled to 127 per month during 2005. Although the number went down in 2006 to 99, it was still more than double the number in 2004. 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 $2.23 per MWh during 2004 and $6.98 per MWh during 2005. In 2006, the impact was $4.68 per MWh. Even though price spikes account for a small portion of the total intervals, they have a significant impact on overall price levels.

Figure 5: Average Balancing Energy Prices and Number of Price Spikes 2004 to 2006

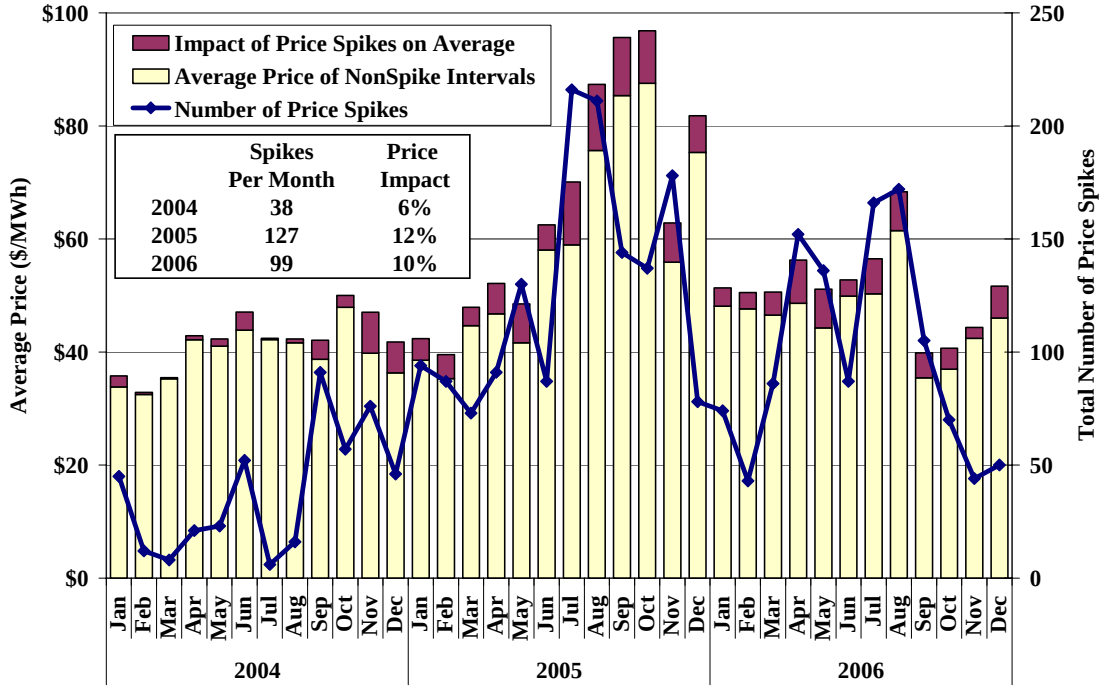


Figure 6: Average Regulation Up Prices and Number of Price Spikes 2004 to 2006

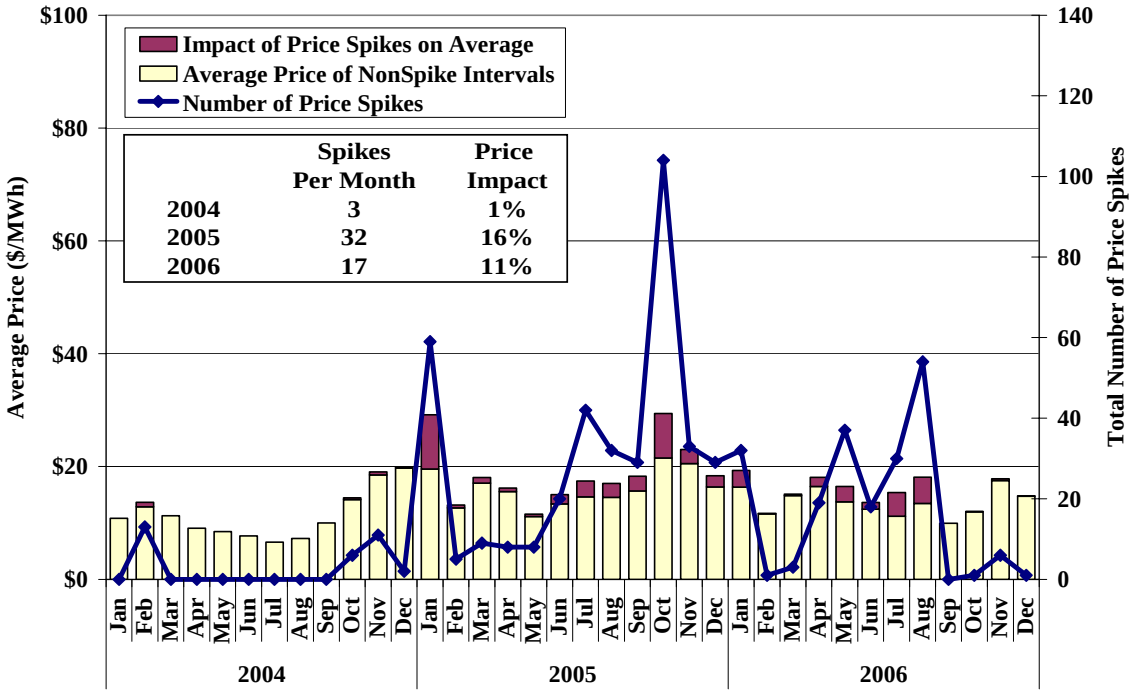


Figure 7: Average Regulation Down Prices and Number of Price Spikes 2004 to 2006

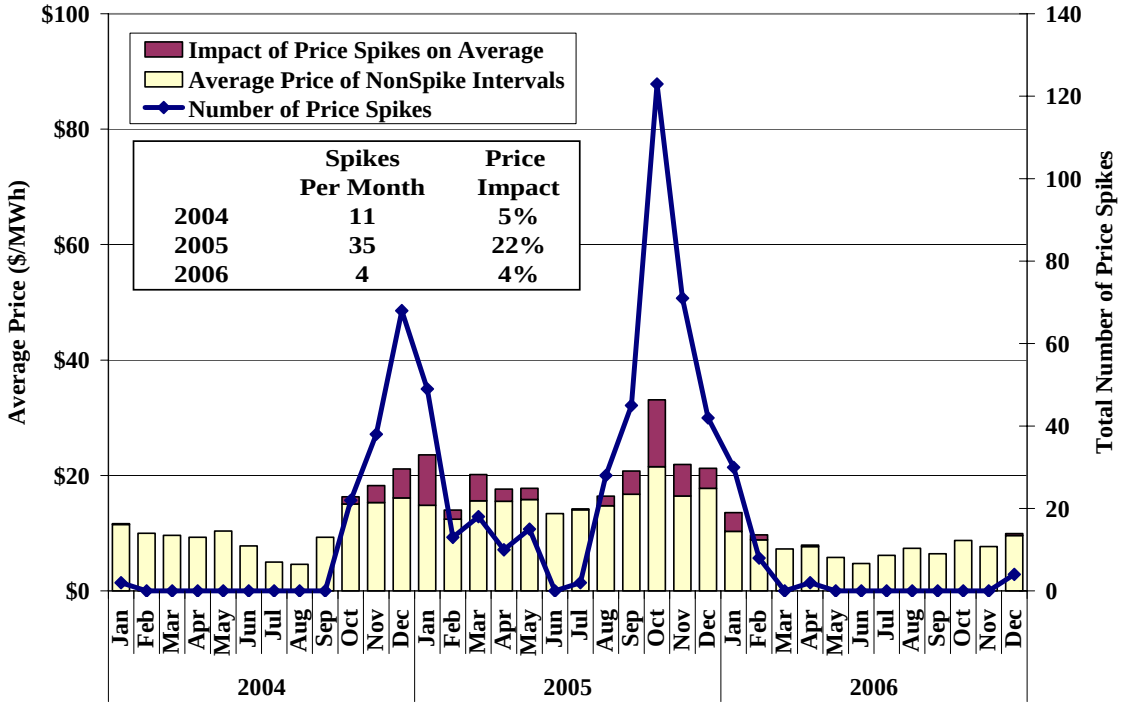
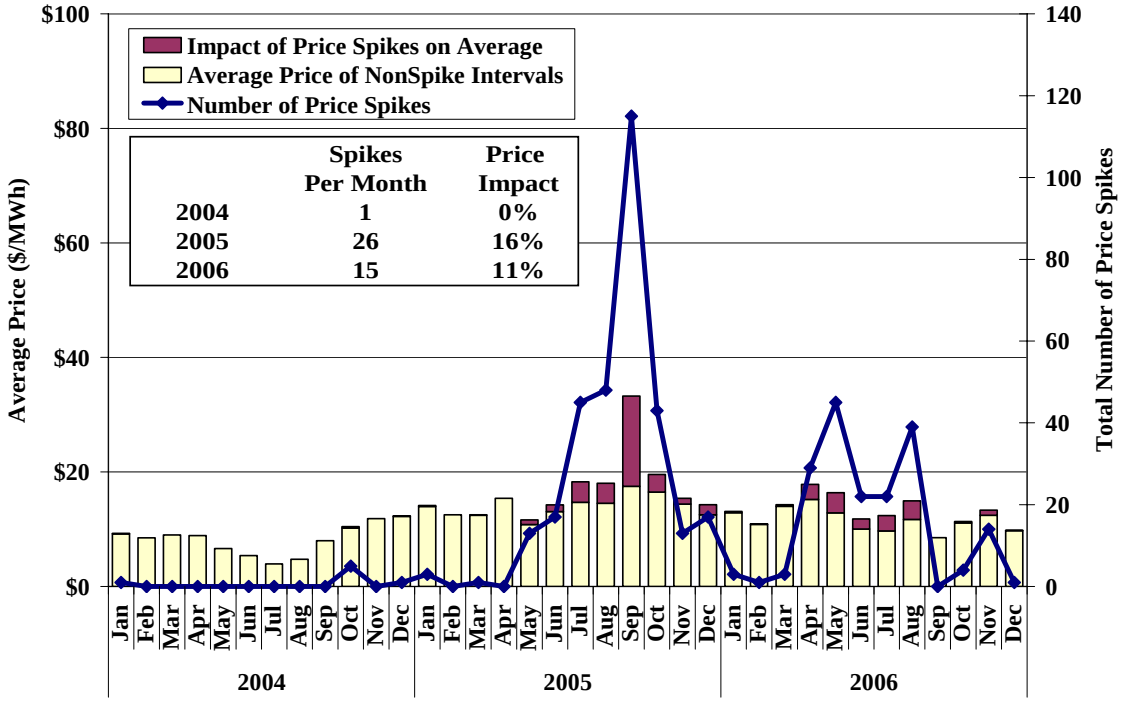


Figure 8: Average Responsive Reserve Prices and Number of Price Spikes 2004 to 2006



Price spikes in the markets for ancillary services have also risen significantly over this period. During 2004, there were three price spike hours per month for regulation up, 11 for regulation down, and one for responsive reserves. However, in 2005, the number of price spike hours rose dramatically to 32 per month for regulation up, 35 per month for regulation down, and 26 per month for responsive reserves.[7] 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. Since the same resources are used to supply ancillary services and energy, increases in energy prices should lead to corresponding increases 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. The price spikes have generally become more frequent and have become a larger component of the average balancing energy and ancillary service prices. There are several factors that have contributed to the rise in price spikes that are analyzed in detail in subsequent sections of this report. 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.

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 includes two charts showing 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

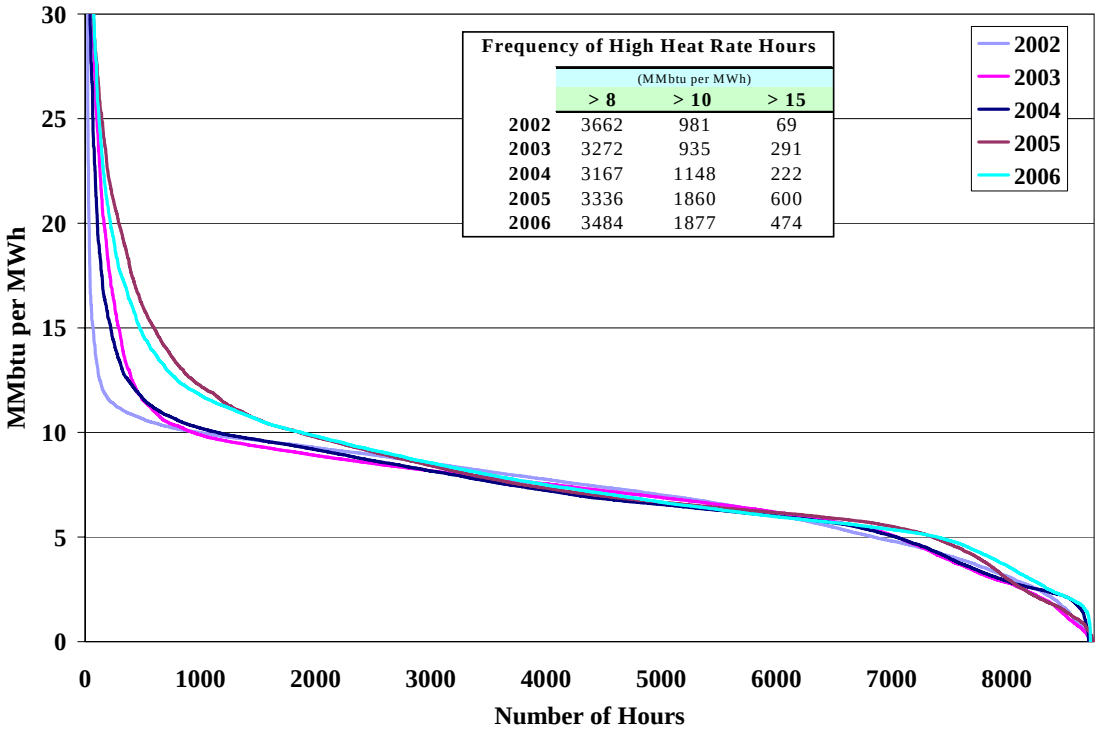
[7] Price spikes are defined as hours where the price exceeds a threshold of $50 per MW for regulation up, regulation down, and responsive reserves.

Natural Gas Price.[8] 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 2002 to 2006.

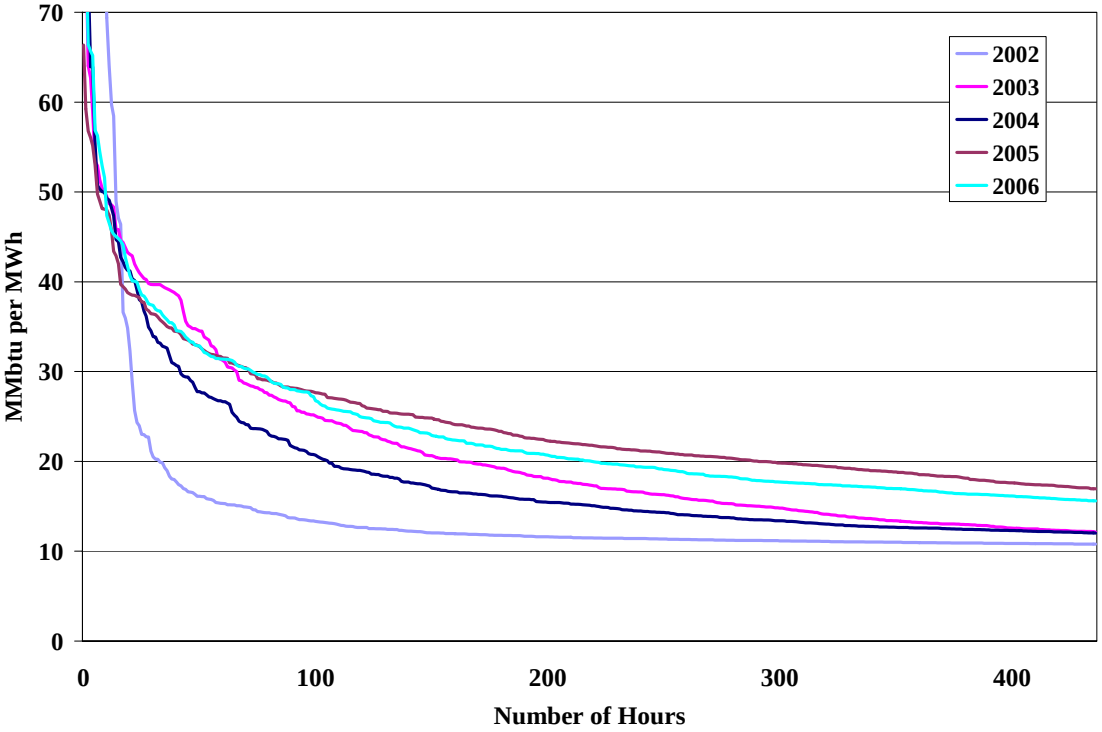
In contrast to Figure 4, Figure 9 shows that the implied marginal heat rates were relatively consistent across the majority of hours from 2002 to 2006. For instance, the table in Figure 9 indicates that the number of hours when the implied heat rate exceeded 8 MMbtu per MWh was relatively consistent across the five years. The rise in energy prices from 2002 to 2006 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 implied heat rate was over 10 was 1,148 in year 2003, whereas in 2005 and 2006 the number rose to 1,860 and 1,877, respectively. However, in 2006, the number of hours when the implied heat rate was over 15 dropped to 474 from 600 in 2005. This indicates that there are price differences that are due to factors other than changes in natural gas prices.

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

Figure 9: Implied Marginal Heat Rate Duration Curve
All Hours – 2002 to 2006

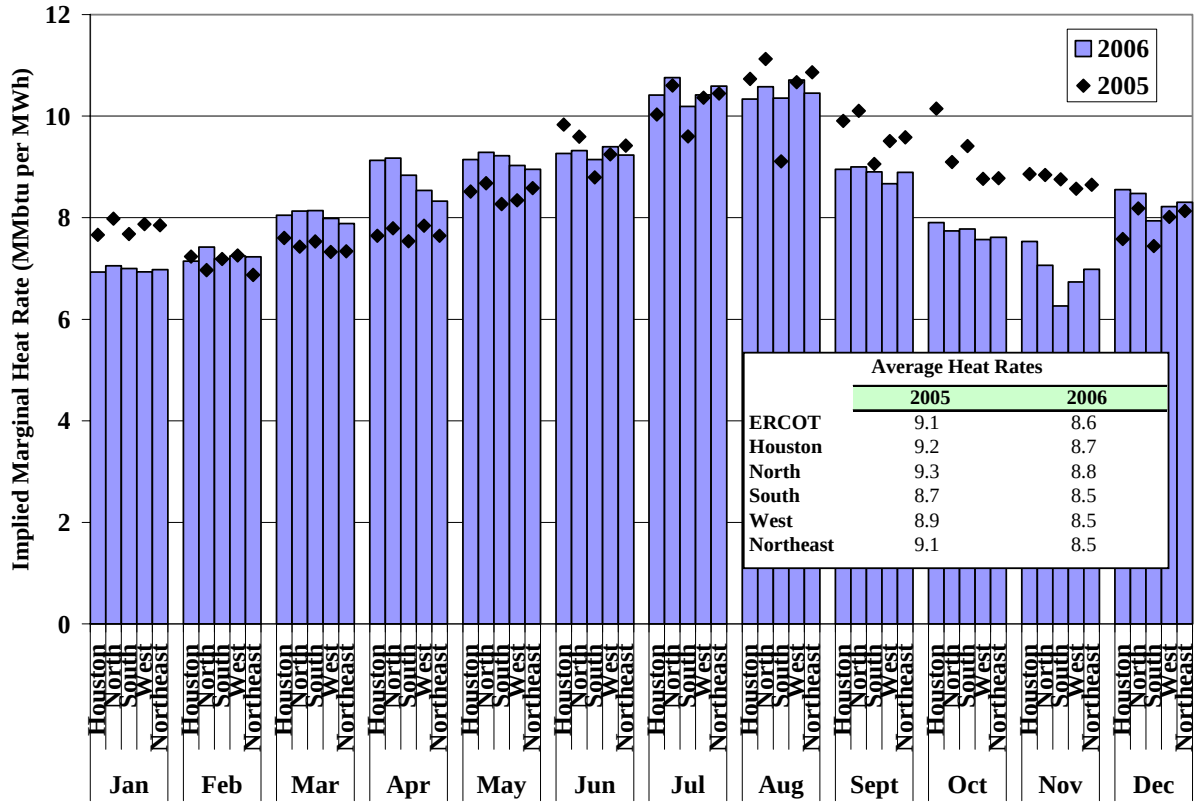


Top Five Percent of Hours in Each Year – 2002 to 2006



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 2005 and 2006. This figure is the fuel price-adjusted version of Figure 1 in the prior sub-section. Adjusted for gas price influence, Figure 10 shows that average implied heat rate for all hours of the year decreased by 5.5 percent from 9.1 in 2005 to 8.6 in 2006.

Figure 10: Monthly Average Implied Marginal Heat Rates 2005 & 2006



3. Price Convergence

One indicator of market performance is the extent to which forward and real-time spot prices converge over time. In ERCOT, there is no centralized day-ahead market so prices are formed in the day-ahead bilateral contract market. The real-time spot prices are formed in the balancing energy market. Forward prices will converge with real-time prices when two main conditions are in place: a) there are low barriers to shifting purchases and sales between the forward and real-time markets; and b) sufficient information is available to market participants to allow them to develop accurate expectations of future real-time prices. When these conditions are met, market participants can be expected to arbitrage predictable differences between forward prices

and real-time spot prices by increasing net purchases in the lower-priced market and increasing net sales in the higher-priced market. 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[9] and the average balancing energy price in each month between 2004 and 2006. This reveals whether persistent and predictable differences exist between forward and real-time prices, which participants should arbitrage over the long-term.

In order 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

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

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 2006.

Figure 11 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 the months of May, June, July and August of 2006, average day-ahead prices were higher than average balancing energy price, while in the other months of 2006, average day-ahead prices were exceeded by the average balancing energy prices, but by a much smaller margin on average than in 2005.

Figure 11: Convergence Between Forward and Real-Time Energy Prices 2004 to 2006

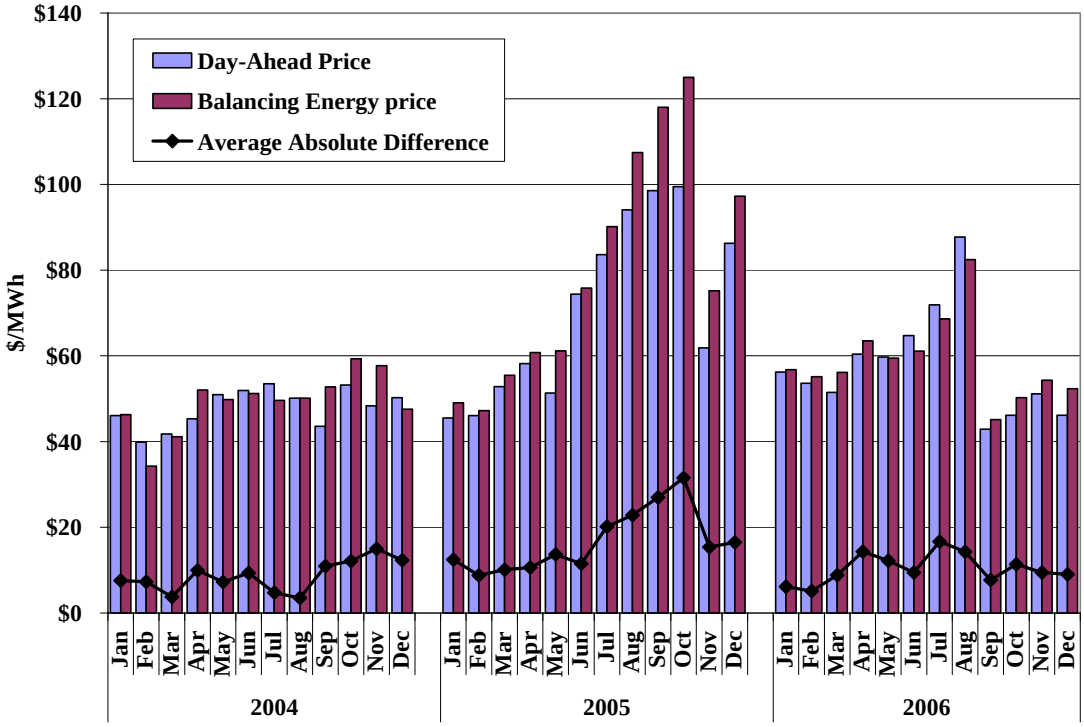


Figure 11 also shows that the average absolute price difference from 2004 to 2006. The difference (shown by the line) was relatively low during the first eight months of 2004 before rising considerably during the last four months. In 2005, the average absolute difference rose sharply in the summer and 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 and $10 in 2006. As noted above, the average absolute difference measures the volatility of the price differences.

The results in this section indicate that convergence between the day-ahead bilateral prices and the balancing energy prices has improved in 2006 from 2005. The frequency of price spikes in 2006 was less than in 2005, but still much greater than the price spike frequency in 2004. However, the average absolute difference between the day-ahead price and balancing energy price in 2006 returned closer to the average level observed in 2004.

4. Volume of Energy Traded in the Balancing Energy Market

In addition to signaling the value of power for market participants entering into forward contracts, the balancing energy market 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. 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 can 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 12 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 12: Average Quantities Cleared in the Balancing Energy Market 2002 to 2006

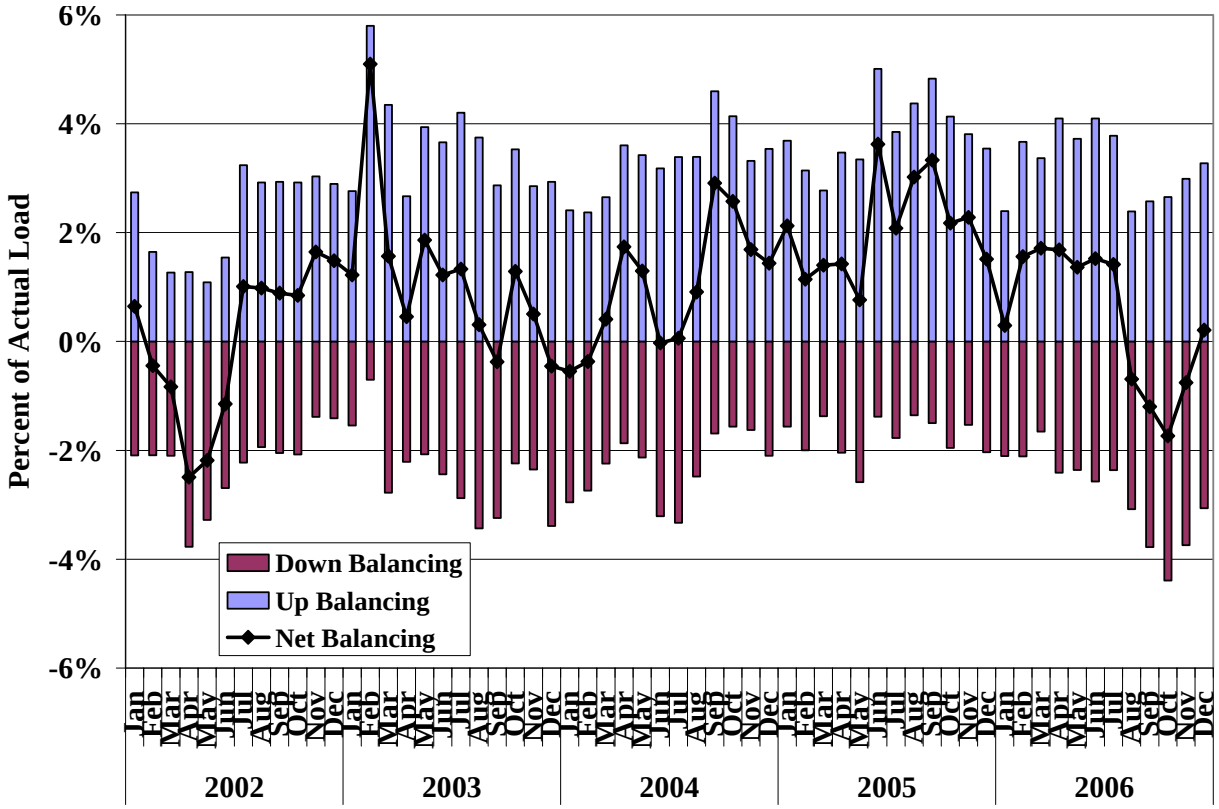


Figure 12 shows that the total volume of balancing up and balancing down energy as a share of actual load increased from an average of 4.6 percent in 2002 to 6.1 percent in 2003, 5.7 percent in 2004, 5.6 percent in 2005, and 6.1 percent in 2006. Thus, there was a general increase in trading through the balancing energy market after 2002. Over time, the volume of balancing up energy has risen relative to the volume of balancing down energy. However, starting in August 2006, the average volume of balancing down energy began to increase. In 2006, the average amount of net balancing up energy (*i.e.*, balancing up minus balancing down) was 1.3 percent. 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 after 2002 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 13 presents a distribution of the hourly net balancing energy. The distribution is shown on an hourly basis rather than by interval to minimize the effect of short-term ramp constraints and to highlight the market impact of persistent under- and over-scheduling. Each of the bars in Figure 13 shows the portion of the hours during 2006 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 14 percent of the hours in 2006.

Figure 13: Magnitude of Net Balancing Energy and Corresponding Price 2006

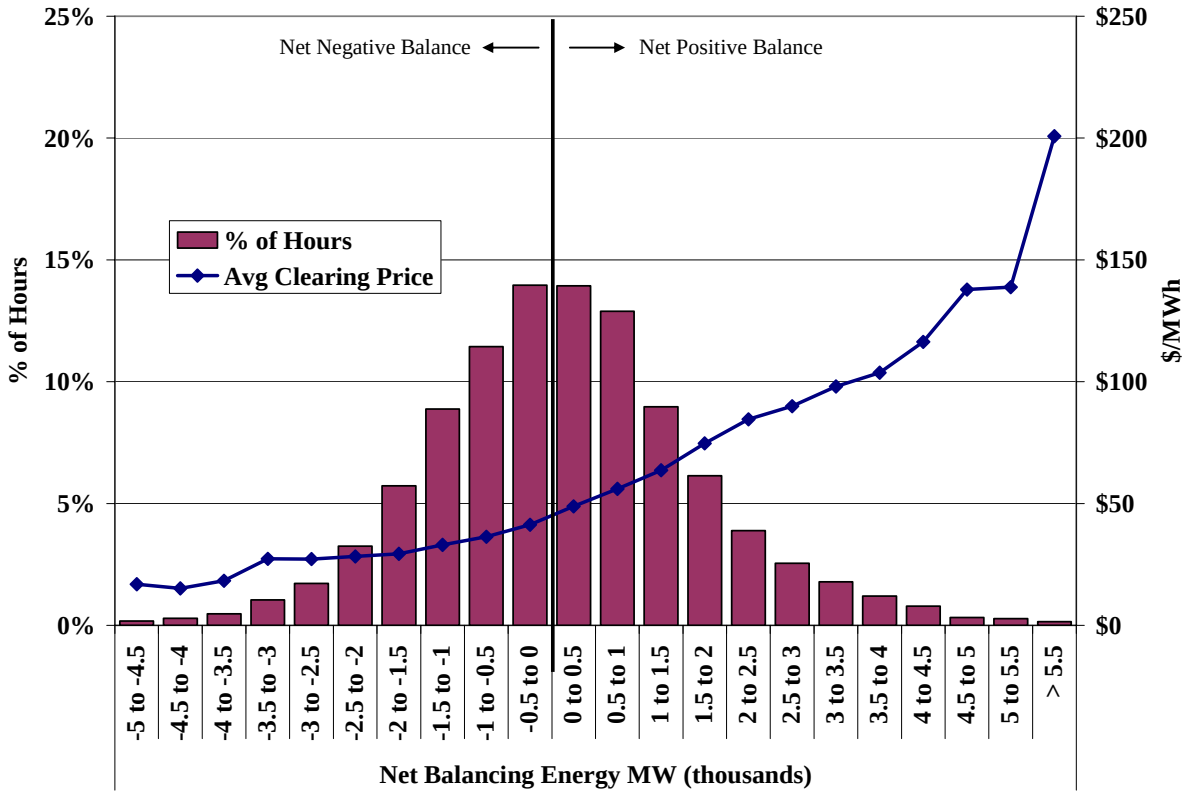


Figure 13 shows a relatively symmetrical distribution of net balancing energy purchases centered around zero gigawatts. This is consistent with Figure 12 which showed that there were comparable portions of net balancing up and down quantities on average during 2006. In approximately 52 percent of the hourly observations shown, Figure 13 also shows that net balancing energy schedules averaged between -1.0 and 1.0 gigawatts.[10] Hence, there were many hours when the net balancing energy traded was relatively low, because the total scheduled energy was frequently close to the actual load.

The line plotted in Figure 13 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,

[10] One gigawatt corresponds to roughly 3 percent of the average actual load in ERCOT.

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 2004 and 2005.[11] The pattern indicates that the balancing energy market is thinly traded, which can undermine its efficiency. 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.

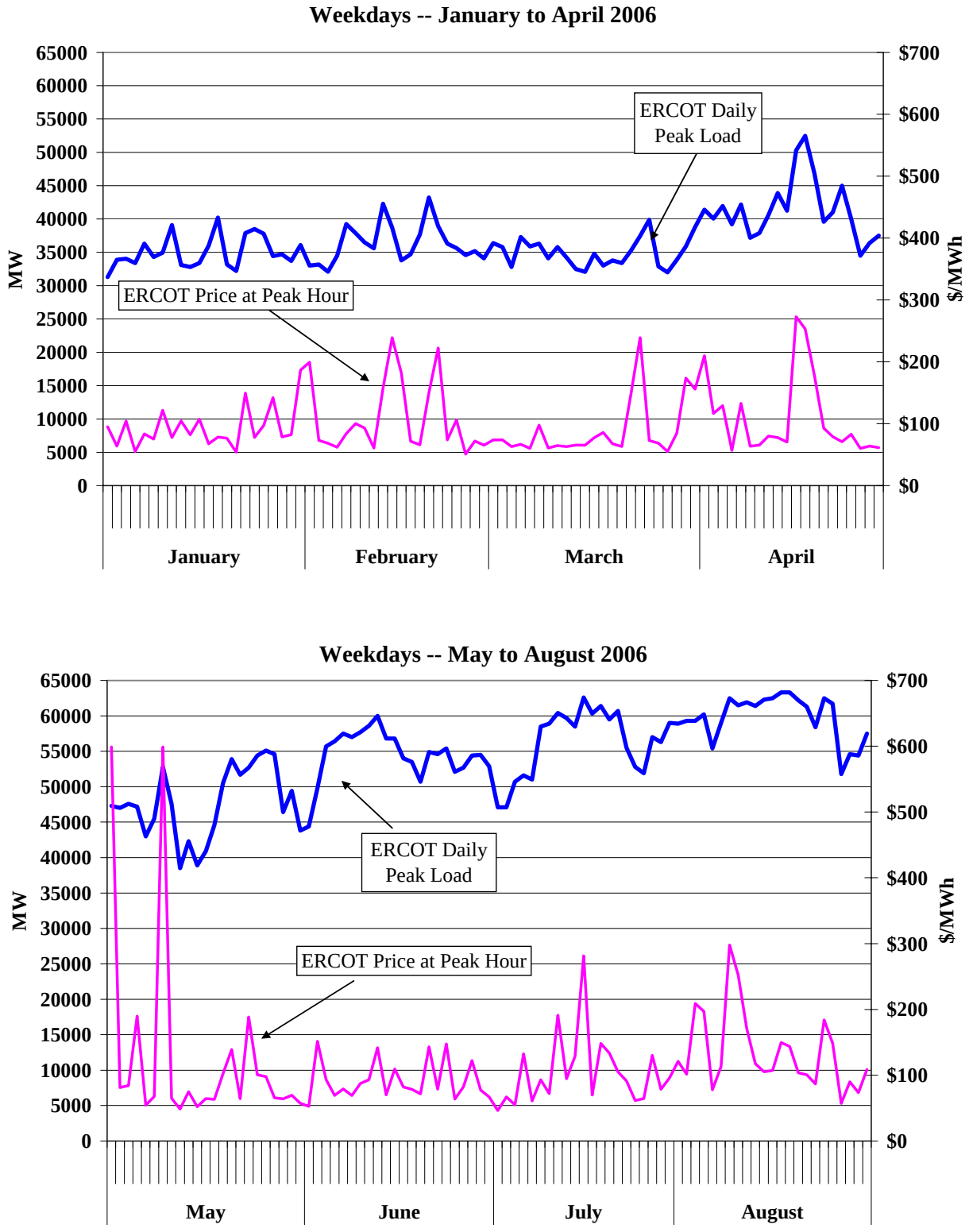
5. Determinants of Balancing Energy Prices

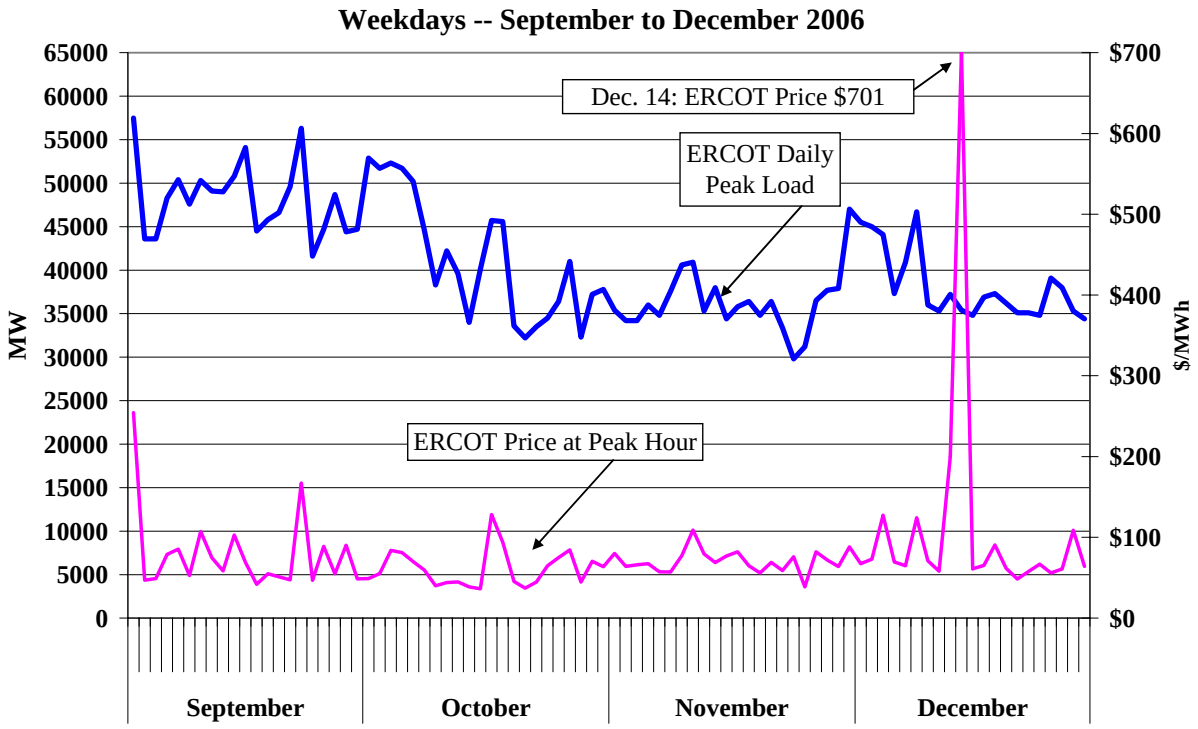
The prior section shows that the level of net sales in the balancing energy market appears to play a significant role in explaining the balancing energy prices. In this section, we examine this relationship in more detail, as well as the role of more fundamental determinants of balancing energy prices,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

Figure 14 shows the average balancing energy price and the actual load in the peak hour of each weekday during 2006.

[11] See 2004 SOM Report and 2005 SOM Report

Figure 14: Daily Peak Loads and Balancing Energy Prices





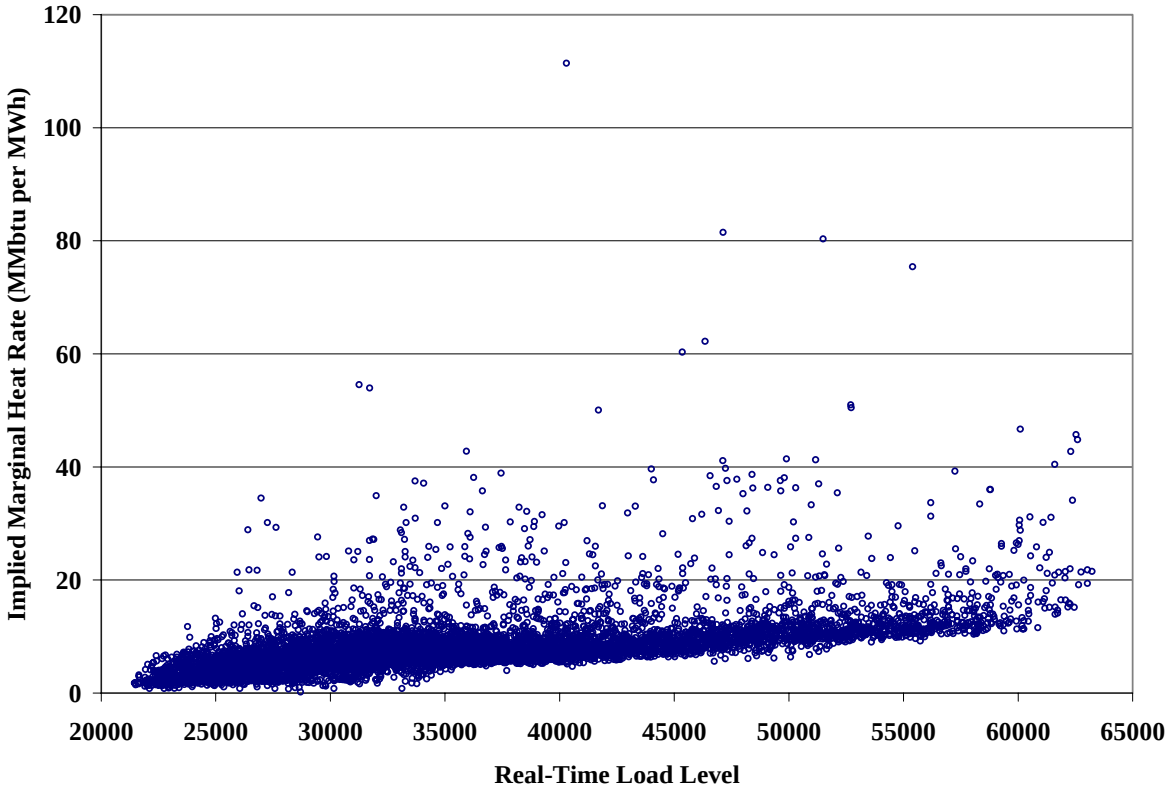
The figure shows that a large share of the days with high prices (*e.g.*, greater than $200/MWh) coincide with periods when demand is high or rising quickly relative to the previous several days. However, prices spikes also occurred during lower demand periods. For instance, on December 14, the price at peak load hour reached $701, while the peak load was lower than the previous day.

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 consistently dropped over the last several years.

To further examine the relationship between actual load in ERCOT and balancing energy prices, Figure 15 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 15: Hourly Gas Price-Adjusted Balancing Energy Price vs. Real-Time Load 2006



The figure indicates a positive correlation between real-time load and the clearing price in the balancing market. Although prices were generally higher at higher load levels, the analysis shown in Figure 13 indicates that the net volume of energy purchased in the balancing energy market is 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 27 GW at 4 AM to 38 GW at 1 PM. Thus, the change in load averages 1,280 MW per hour (320 MW per 15-minute interval) during the morning and early afternoon. Figure 16 shows the average load and balancing energy price in each interval from 4 AM through 1 PM in 2006. The price is plotted as a line in the figure

while the average load is shown with vertical bars.

Figure 16: Average Clearing Price and Load by Time of Day
Ramping-Up Hours – 2006

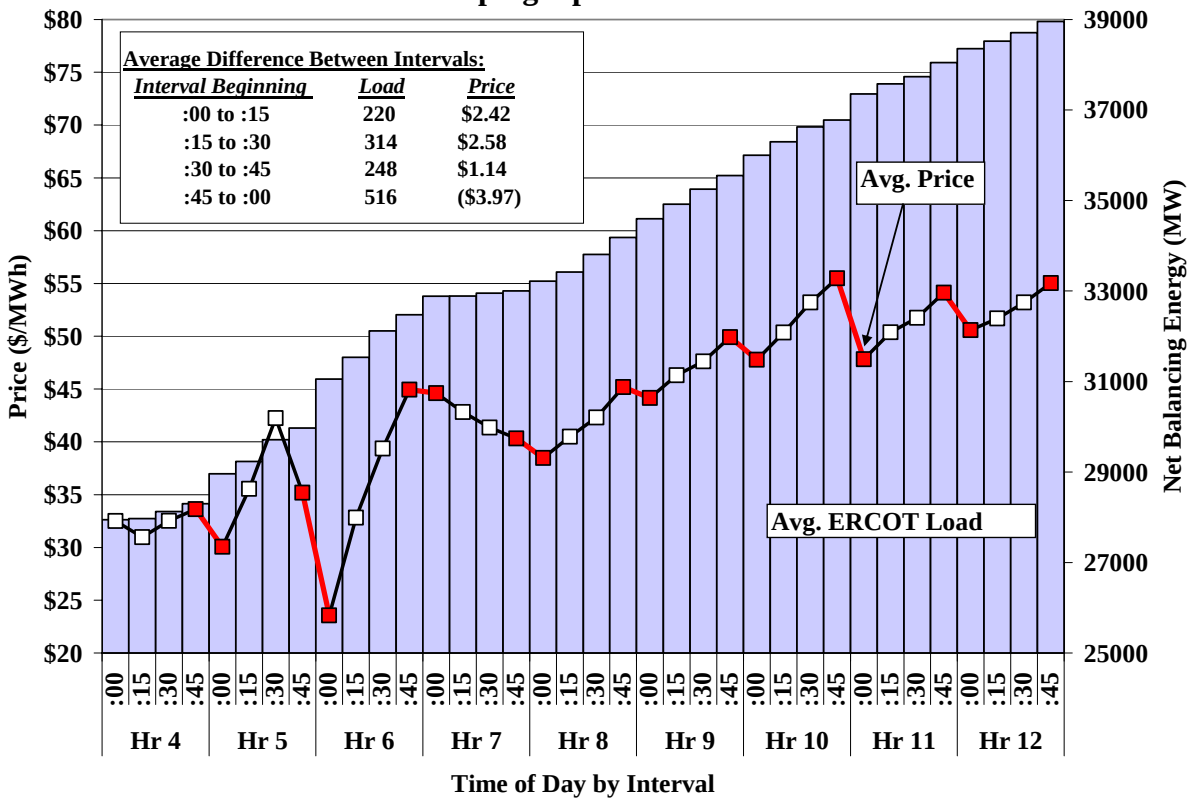


Figure 16 shows that, with the exception of hour 7, the load steadily increases in every interval and prices generally move upward from about $32 per MWh at 4:00 AM to $55 per MWh at 12:45 PM. If actual load were the primary determinant of energy prices, the balancing energy prices would rise gradually as the actual load rises. However, Figure 16 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 -$3.97 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 1,840 MW per hour (460 MW per 15-minute interval) during the late evening. Figure 17 shows this decrease in load by interval, together with the average balancing energy prices for the intervals from 9 PM to 3 AM.

Figure 17: Average Clearing Price and Load by Time of Day
Ramping-Down Hours – 2006

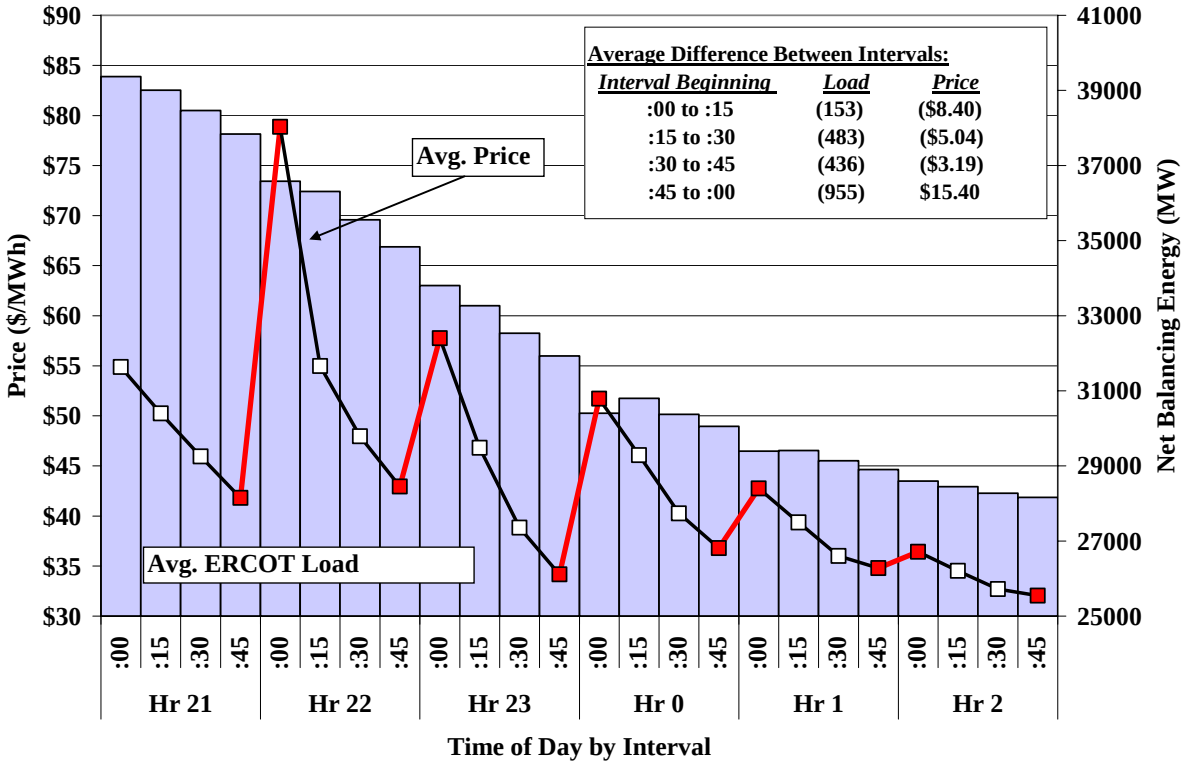


Figure 17 shows that while balancing energy prices decrease over these intervals, they follow a similar pattern as exhibited in the ramping-up hours. The balancing energy price decreases in each interval of the hour before rising substantially in the first interval of the following hour. The balancing energy price increases by an average of $15.4 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 2004 and 2005.[12]

These figures show that this pattern of balancing energy prices by interval is not explained by

[12] See 2004 SOM Report and 2005 SOM Report

changes in actual load. Rather, changes in balancing energy deployments by interval underlie this pricing pattern. Sizable changes in balancing energy deployments occur between intervals, particularly in the first interval of the hour. These changes are associated with large hourly changes in energy schedules. These scheduling and pricing patterns are examined in detail in Section II below.

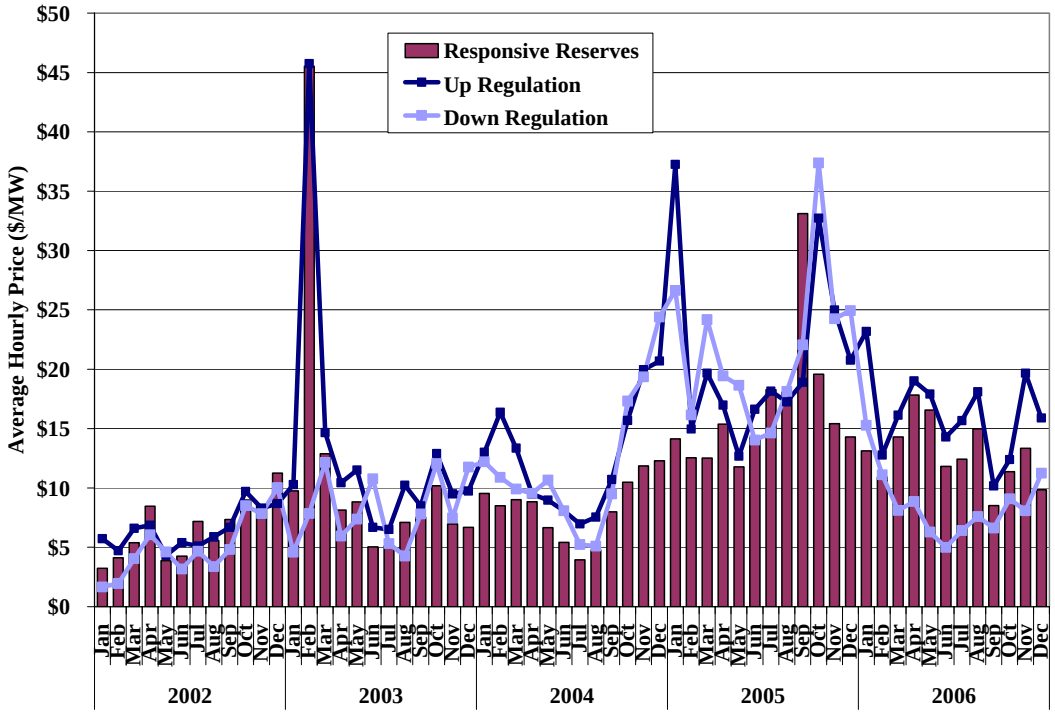
B. Ancillary Services Market Results

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 2006.

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 18 shows the monthly average ancillary services prices between 2002 and 2006. Average prices for each ancillary service are weighted by the quantities required in each hour.

Figure 18: Monthly Average Ancillary Service Prices 2002 to 2006



This figure shows that ancillary services prices have generally risen from 2002 to 2005, but that the price levels moderated in 2006. 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. Both providers of 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 2002 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, 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 four years in those hours when ERCOT procured non-spinning reserves. Non-spinning reserves were purchased in approximately 18 percent of the hours during 2002, 25 percent of hours during 2003, 24 percent of hours during 2004, 23 percent of hours during 2005, and 20 percent of hours during 2006.

Table 1: Average Hourly Responsive Reserves and Non-Spinning Reserves Prices During Hours When Non-Spinning Reserves Were Procured 2002 to 2006

	2002	2003	2004	2005	2006
Non-Spin Reserve Price	$14.51	$9.85	$6.83	$25.10	$21.75
Responsive Reserve Price	$9.20	$10.73	$9.10	$28.16	$25.55

Table 1 shows that responsive reserves prices are higher on average than non-spinning reserves prices during hours when non-spinning reserves were procured. The prices in 2002 were the exception because non-spinning reserves prices were above $990 per MWh for 13 hours on two days. 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.

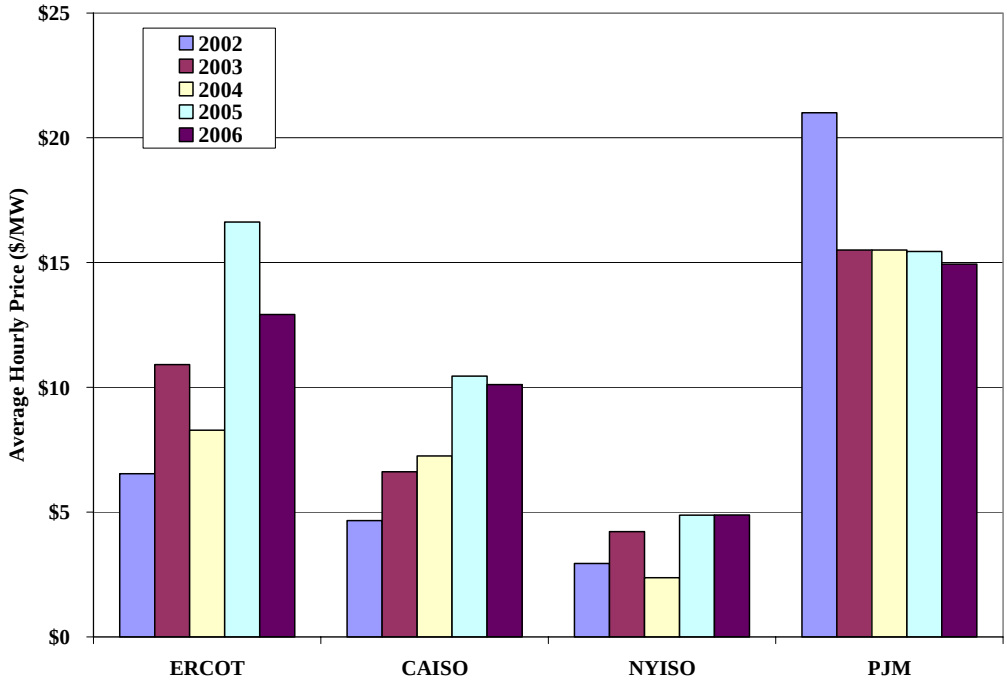
Generators incur two types of costs associated with providing reserves in the ERCOT market. First, reserves providers incur opportunity costs from any profitable sales they forego in the energy market. For generators, this is the same regardless of whether the generator is providing responsive or non-spinning 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 often 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 2006, about 2 percent of the responsive reserves were actually deployed, while 5.2 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 load. 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 2,300 MW of responsive reserves to ensure adequate protection against the loss of the two largest units.

Responsive reserve prices dropped in 2006 from 2005, but remained higher than the prices observed in 2002 to 2004. Figure 19 shows how the annual average prices in ERCOT from 2002 to 2006 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 2006.

Figure 19: Responsive Reserves Prices in Other RTO Markets 2002 to 2006



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 20 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 1,280 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 1,970 MW.

Figure 20: Regulation Prices and Requirements by Hour of Day 2006

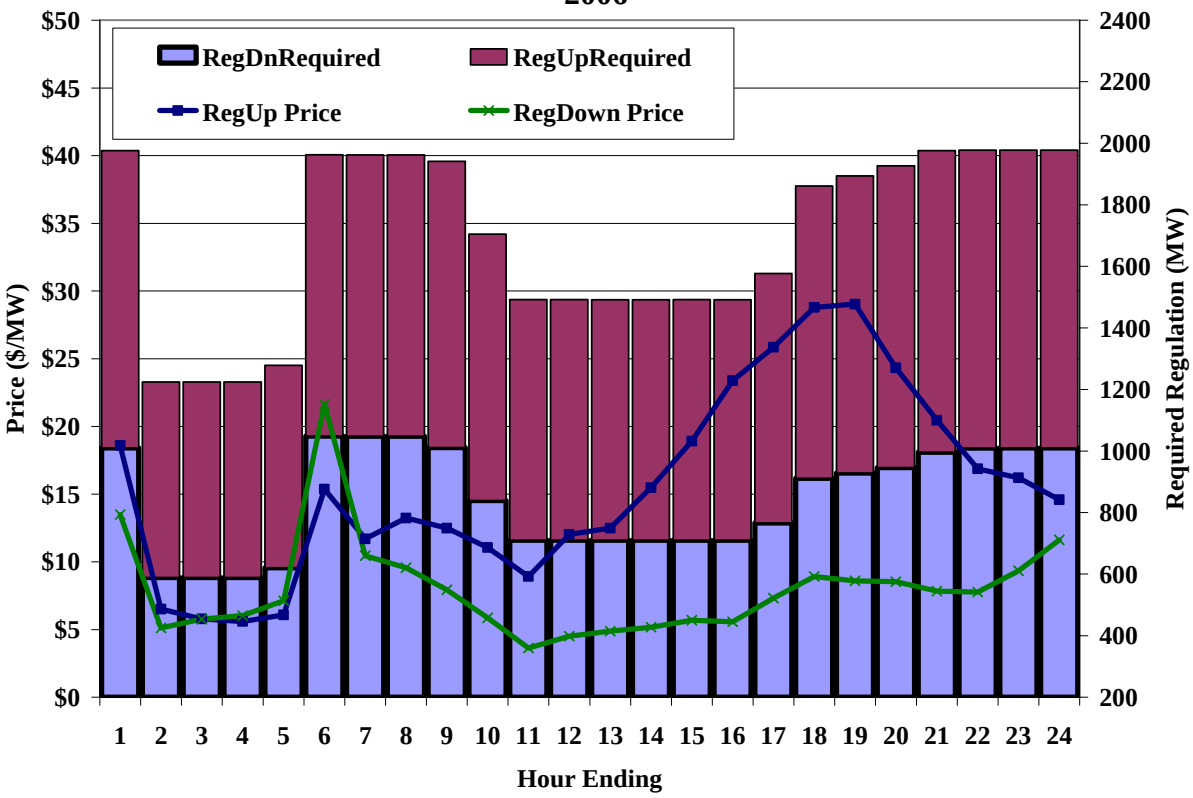


Figure 20 indicates that average regulation prices are generally correlated with the regulation quantity purchased and the typical load pattern in ERCOT. During non-ramping hours, such as overnight and late morning, regulation up and down prices range from $5 to $15 per MW. During the ramping hours in early morning and evening, average regulation up and down prices range from $10 to $20 per MW. In the afternoon hours, regulation up prices range from $20 to $30 and regulation down prices range from $5 to $10 per MW. Regulation up prices are higher on average in the 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 lower the clearing prices of regulation. Figure 21 summarizes the average amounts of regulation procured through the auction and/or bilateral arrangements on an annual basis since 2002.

Figure 21: Annual Average Regulation Procurement 2002 to 2006

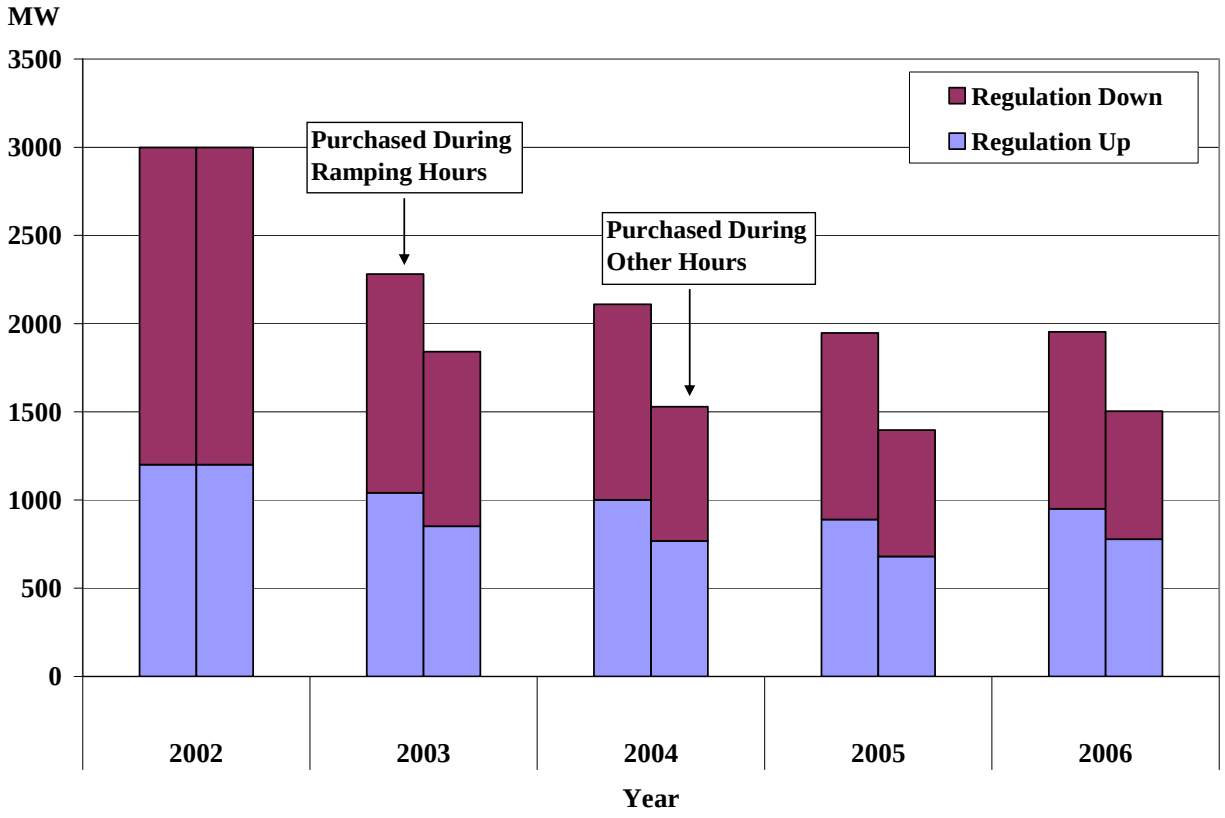


Figure 21 shows that ERCOT has reduced the average regulation quantity scheduled since 2002. The largest reduction was from 2002 to 2003, although the reductions in the remaining two years were also substantial. The regulation quantities required in 2006 was almost the same as in 2005 during ramping hours and was a slightly higher quantity than 2005 during non-ramping hours. Overall, ERCOT has lowered the required amount by 35 percent during ramping hours and 50 percent during non-ramping hours. During the same period, ERCOT also adjusted the relative

shares of regulation up and regulation down with the regulation down share decreasing from 60 percent in 2002 to close to 50 percent in 2006.

Currently, ERCOT's regulation procurement methodologies group regulation procurement quantities into 4 to 6 blocks of hours and procure the same quantity in each block for each day in each month. In late 2006, the Independent Market Monitor ("IMM") initiated discussions with ERCOT to investigate modifications to this methodology that would allow 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. The ERCOT Board approved the changed methodology in June 2007 to be implemented in August 2007. It is expected that this change will reduce the overall quantities of regulation procured over all hours, but may increase the regulation quantities procured in certain hours. This change should result in more efficient procurement of regulation up and down service while maintaining or even improving reliability.

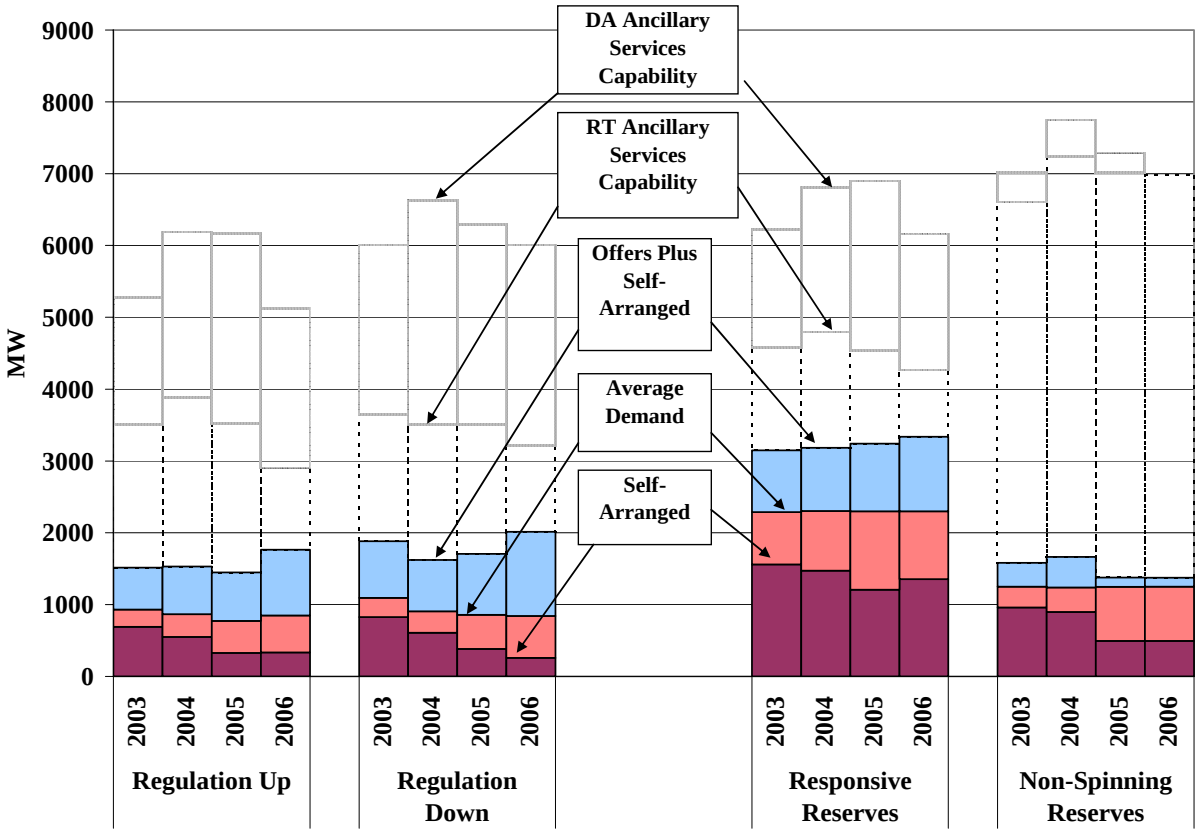
2. Provision of Ancillary Services

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 22. 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 22 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 22: Reserves and Regulation Capacity, Offers, and Schedules 2003 to 2006



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 22 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 22 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 1150 MW of the

responsive reserves capability shown above, because there is currently a requirement that no more than 50 percent of the 2300 MW requirement be met with LaaRs.

For non-spinning reserves, Figure 22 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 22 shows that except for responsive reserve in 2006, in which about 54 percent of available responsive reserve capacity was offered, less than one-half of each type of ancillary services capability was offered during 2003, 2004, 2005, and 2006. 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.

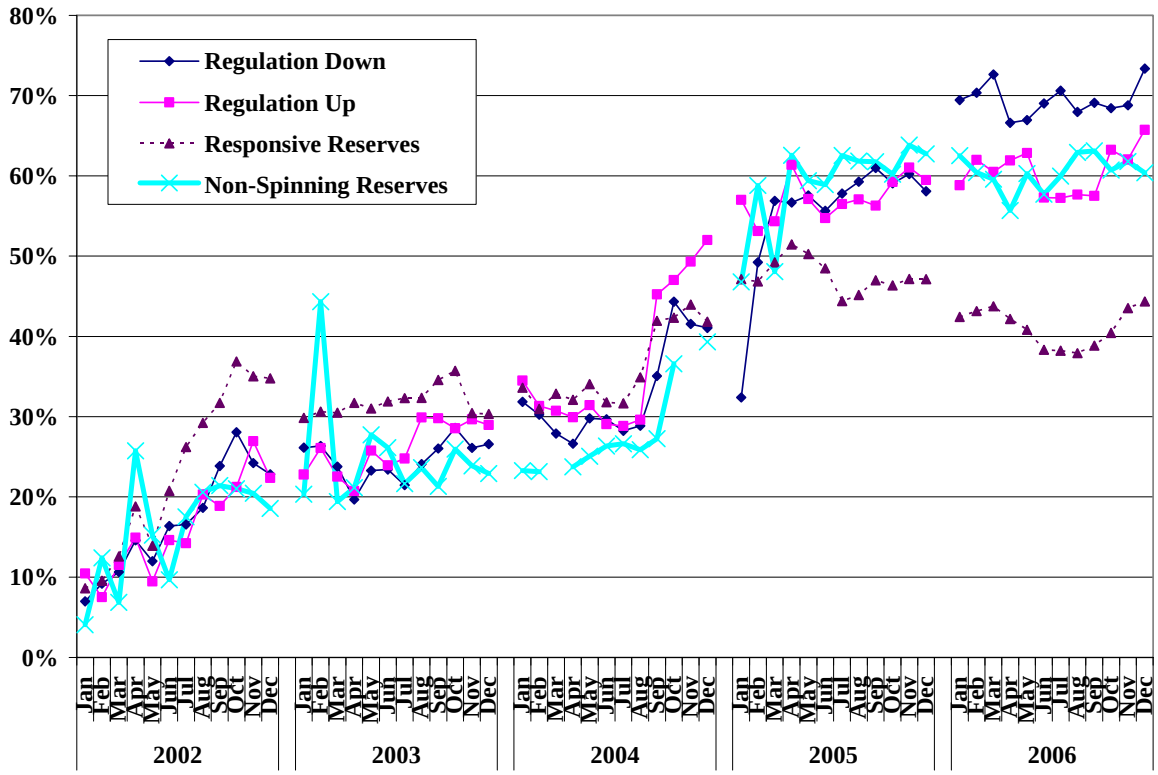
In addition, participants may not offer the capability of resources they do not expect to commit for the following day. Suppliers could submit offer prices high enough to ensure that their costs of committing additional resources to support the ancillary services offers are covered. However, under the 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. On average, there is often a substantial quantity of reserves that remain available in real time, but that is not offered. This is surprising given the relatively high prices for operating reserves in ERCOT. It is possible that some of the ancillary services capability is withheld in an attempt to increase the ancillary services clearing prices. However, this is not likely to be the primary reason, since both small

and large participants choose not to offer substantial portions of their capability in the ancillary services market.

Figure 22 shows modest changes in the amount of day-ahead ancillary services capability between 2003 and 2006. 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. The average amount of excess on-line capacity has declined each year since 2003, thereby reducing the amount of capacity available to provide ancillary services.

Finally, although market participants increasingly rely on the auction market to procure these services, Figure 23 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 23 shows the share of each type of ancillary service that is purchased through the ERCOT market.

Figure 23: Portion of Reserves and Regulation Procured Through ERCOT 2002 to 2006



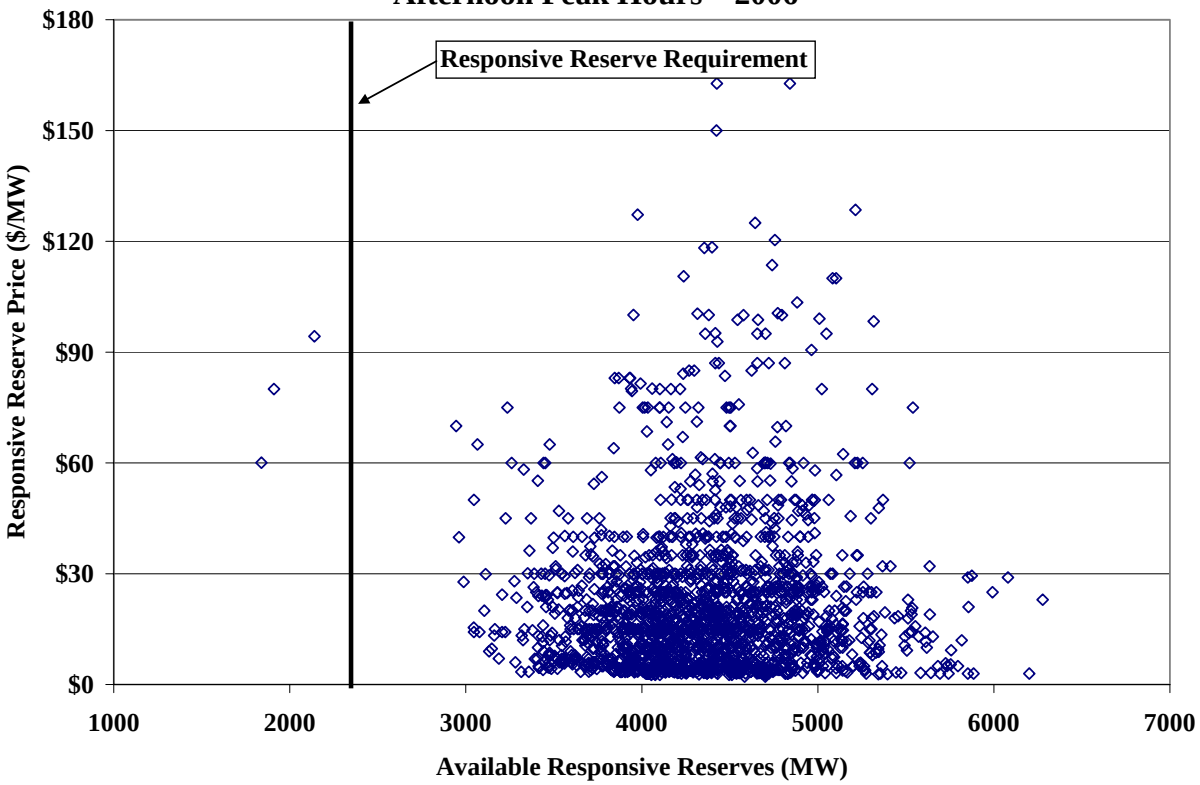
This figure shows that purchases of all ancillary services from the ERCOT markets have generally increased over time, although the purchases of responsive reserve from the ERCOT

market has dropped slightly over the last two years (*i.e.*, the quantity of self-arranged responsive reserve has increased slightly over the last two years). 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 2006. 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 in 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 24 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 24: Hourly Responsive Reserves Capability vs. Market Clearing Price
Afternoon Peak Hours – 2006



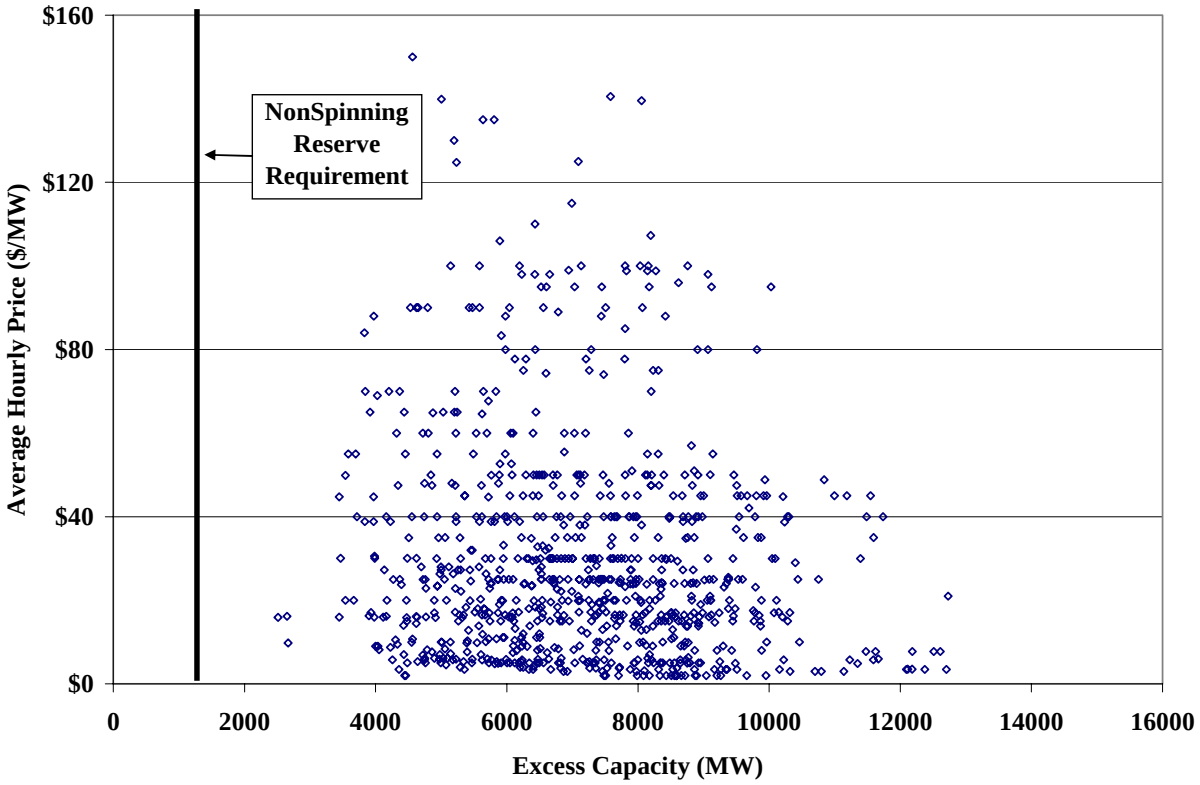
This figure indicates very little relationship between the hourly available responsive reserves capability and the responsive reserves prices in real time. In a well functioning-market for responsive reserves, we would expect excess capacity to be negatively correlated with the clearing prices, but this was not the case in 2006. Similar analyses in previous reports show the same lack of correlation between prices and available reserves. These results reinforce the potential benefits promised by jointly optimizing the operating reserves and energy markets, which is currently being developed for implementation in the nodal market by 2009 (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 25 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 2006.

Like the previous analysis of responsive reserves, the results shown in Figure 25 do not indicate a significant correlation between non-spinning reserves prices and the quantity of available reserves capability in real time. This is consistent with similar analyses in previous reports which showed a lack of correlation between prices and excess capacity in 2004 and 2005. In a well functioning-market for non-spinning reserves, we would expect excess capacity to be negatively correlated with the clearing prices.

Figure 25: Hourly Non-Spinning Reserves Capability vs. Market Clearing Price All Hours 2006

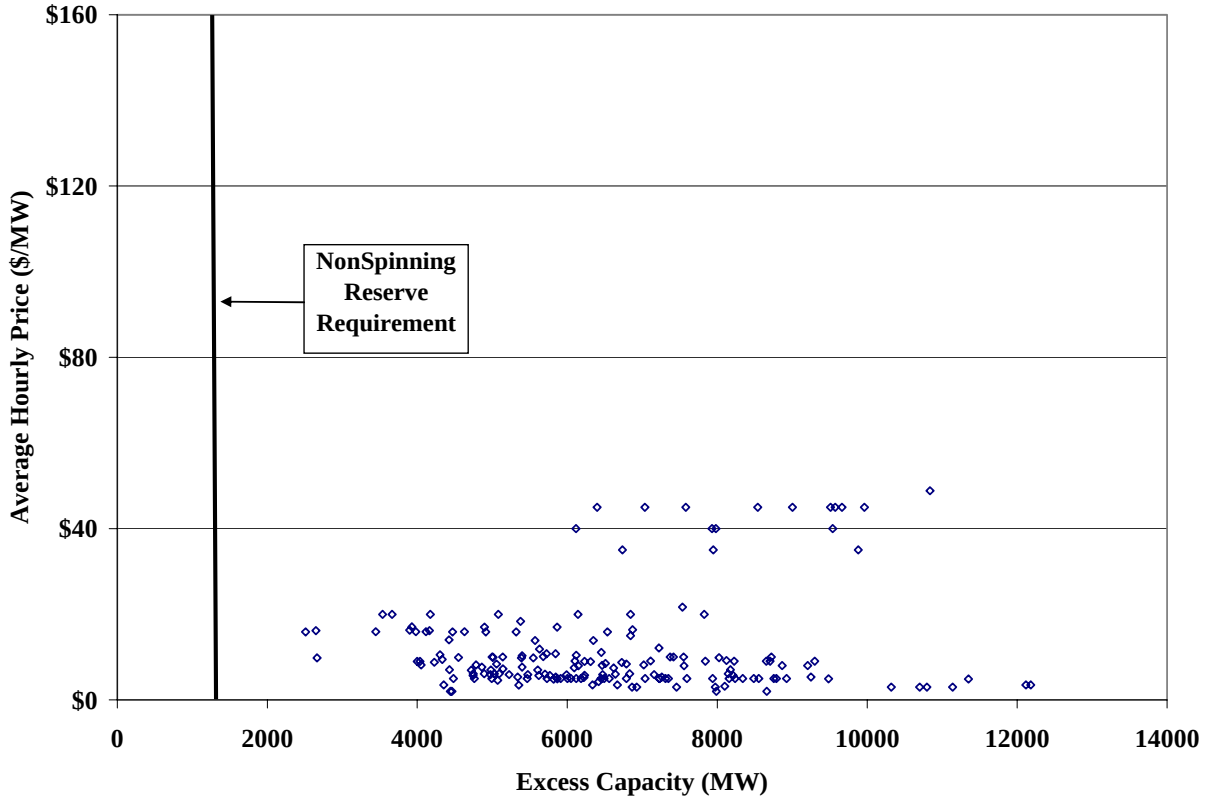


One factor affecting non-spinning reserve prices is that, prior to September 2006, the deployment of non-spinning reserves typically resulted in a significant reduction in the market clearing price of energy. Hence, units deployed for non-spinning reserves would often receive a price for the deployed energy that was significantly less than the operating cost of the unit. In September 2006, new pricing rules were implemented that provide for the recalculation of the energy clearing price when non-spinning reserves are deployed on an *ex post* basis by re-running the

market clearing engine under the assumption that the energy from the deployed non-spinning reserves was unavailable.[13]

Figure 26 shows the data as in Figure 25 for just the months of September though December, 2006. These results clearly show an overall reduction in the clearing price for non-spinning reserves after the implementation of the new rules, which was expected given that the new rules significantly reduce the financial operating risk to providers of non-spinning reserve.

Figure 26: Hourly Non-Spinning Reserves Capability vs. Market Clearing Price September - December 2006



Although the implementation of the new pricing rules associated with the deployment of non-spinning reserves have produced the expected results, ideally the pricing adjustments should be performed in real-time instead of after-the-fact to send accurate and timely price signals to both resources and loads. Further, the current re-pricing mechanism is rather extreme in that it effectively assumes that the energy from non-spinning reserve units is offered at the system-wide offer cap. It would be more reasonable to employ an *ex ante* proxy price that is a function of the

[13] These new rules were approved in Protocol Revision Request No. 650.

incremental costs of deploying an off-line gas turbine. However, because of limitations of the current systems, neither of these improvements is feasible under the current market design.

C. Replacement Reserve Service Market

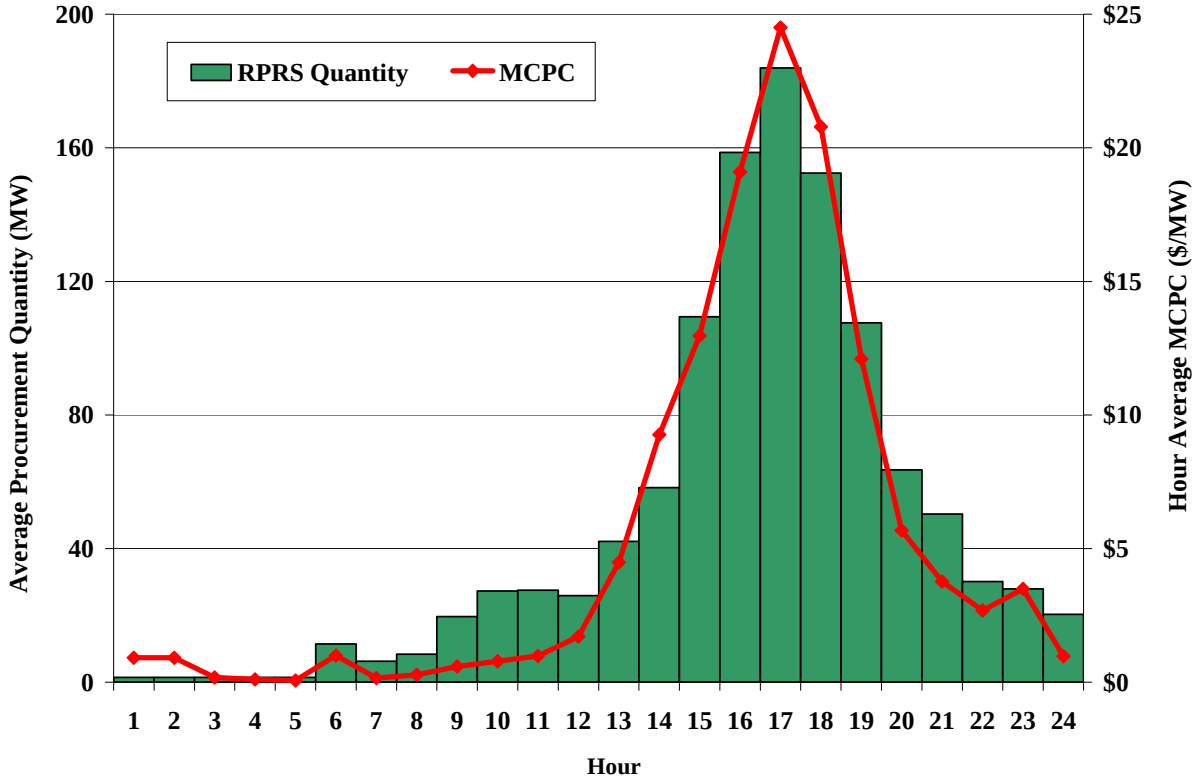
Unit commitment is the day-ahead process of determining which units will be online to meet the forecast demand for the next operating day while observing reliability requirements such as reserve requirements and transmission system limitations. In ERCOT, market participants self-commit units on a decentralized basis to accommodate bilateral energy schedules, ancillary services agreements and load requirements. Building off of this self-commitment, ERCOT conducts a centralized reliability unit commitment to secure additional units that may be necessary to ensure that the system capacity requirement is met and transmission congestion can be resolved in real time operations. Prior to late March 2006, ERCOT relied exclusively upon out-of-merit capacity (OOMC) for this purpose. However, beginning in April 2006, the Replacement Reserve Service (RPRS) market was implemented by ERCOT as the primary tool used to commit capacity in the day ahead to ensure system reliability. Unlike OOMC, RPRS allows ERCOT to optimize unit commitment considering economic and operational factors over all 24 hours of the next operating day.

The RPRS market uses a three step process to commit units and derive the market clearing prices for zonal replacement reserve services. In the first step, the units are selected to satisfy the system load requirement considering transmission limitations (*i.e.*, congestion). Pricing for units selected in step one is cost-based. In the second step, units will be committed when additional capacity is needed to satisfy the forecasted load and system ancillary services requirement. Unlike step one, pricing for units selected in step two is market-based and is a function of the replacement bids submitted by the market participants. Upon the completion of steps one and two, the RPRS market clearing engine generates the market clearing price for each hour for any unit selected in step two. The discussion in this section is limited to replacement reserve quantities procured in step two.

Figure 27 shows the hourly average replacement reserve prices in 2006. As shown in this chart, hour ending 1700 has the highest average market clearing price, which coincides with the typical occurrence of the daily peak load in the summer in the ERCOT market. The market clearing

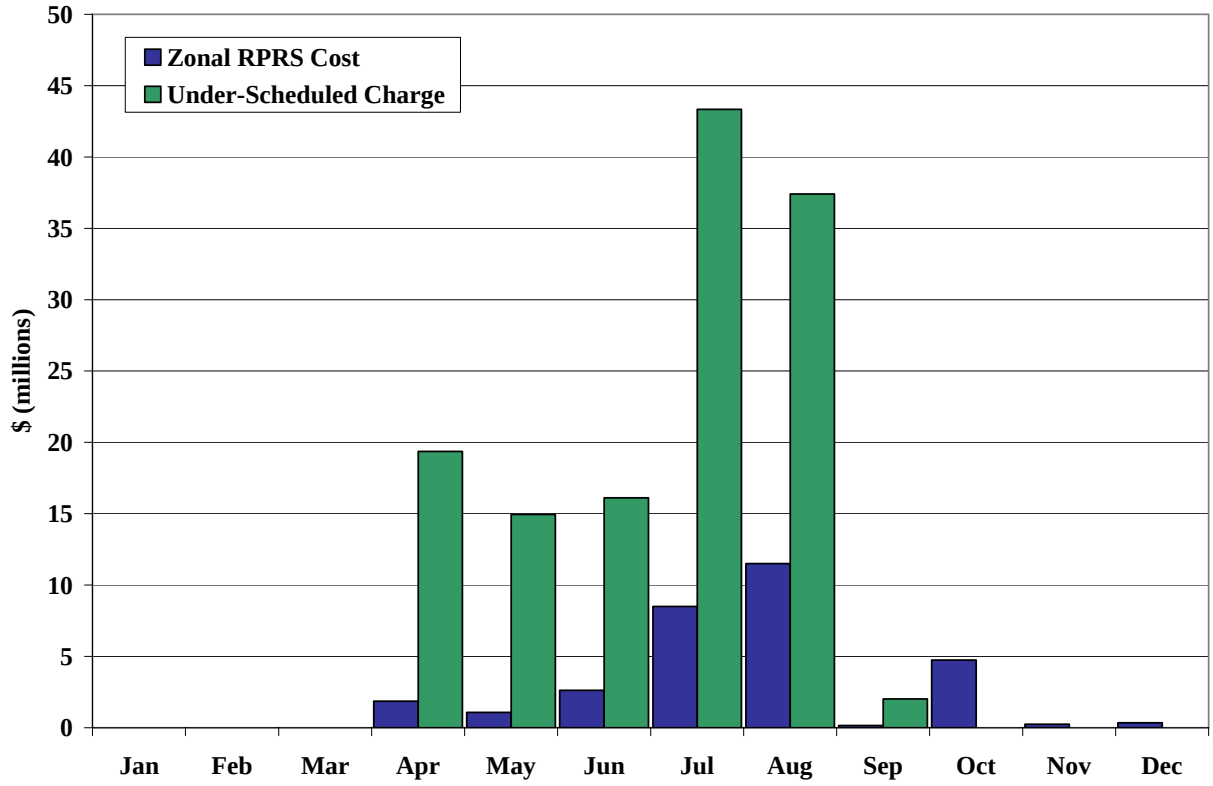
prices for the off peak hours are relatively low, which is consistent with the fact that ERCOT market usually has excess online capacity during off-peak hours.

Figure 27: Replacement Reserve Hourly Average MCPC & Capacity Procurement 2006



From late March through September 2006, costs associated with step two RPRS procurements by ERCOT were directly assigned to QSEs (under-scheduled charge) based upon the RPRS step two clearing price and the measured difference between the day ahead scheduled and actual load of the QSE. Figure 28 shows the zonal RPRS cost and the under-scheduled charge by month for 2006. The under-scheduled charge is greater than the RPRS cost in April through September because the under-scheduled quantity was greater than the quantity of RPRS procured.

Figure 28: Zonal RPRS Cost and Under-Scheduled Charge 2006



Due to concerns raised regarding the accuracy of the cost causation elements associated with the direct assignment provisions of RPRS, the direct assignment provisions were suspended by ERCOT effective October 1, 2006 pending consideration by the PUCT of an appeal relating to these matters.

Ultimately, the PUCT made permanent the suspension of the direct assignment of RPRS step two costs such that all RPRS costs are assigned to all QSEs on a load ratio share for the duration of the existence of the zonal market, noting that the implementation of the nodal market with a centralized day-ahead market and associated provisions related to unit commitment payment and cost allocation should largely resolve the issues associated with the RPRS market in 2006.

D. 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

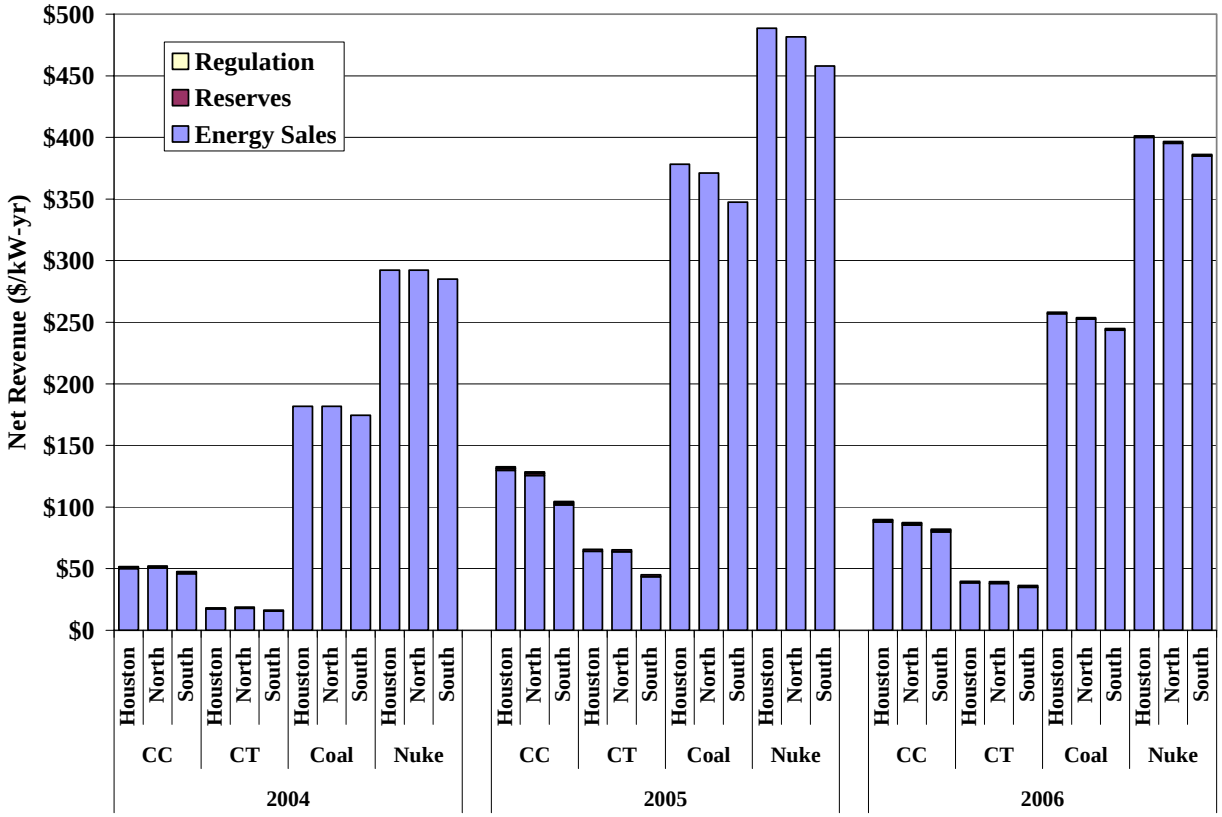
reserves, and regulation markets together provide the economic signals that govern suppliers' decisions to invest in new generation or retire existing generation. In a long-run equilibrium, the markets should provide sufficient net revenue to allow an investor to break-even on an investment in a new 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 2002 and 2006 by various types of generators in each zone.

Figure 29 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 incurring an 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 29: Estimated Net Revenue 2004 to 2006



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 Northeast and West zones fall within the range of the other three 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 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 29 shows that the estimated net revenue for all technologies grew significantly from 2002 to 2003 and again from 2004 to 2005. The net revenue fell in 2006 in each zone compared to 2005; however, net revenue remained higher in 2006 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 $60 to $85 per kW-year. The estimated net revenue for a new gas turbine in 2006 is approximately $40 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 $95 to $125 per kW-year. The estimated net revenue in 2006 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 2003 and 2004, but exceeded them in 2005 and 2006. Thus, it is not surprising that some market participants are expressing interest in

building new baseload facilities in ERCOT.[14] However, these results should be tempered by the fact that there are likely additional costs for these technologies that are not included in our generic cost estimates, including the costs associated with the nuclear waste disposal.

Although estimated net revenue grew considerably in 2005 and 2006 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 that ERCOT plans to implement by 2009 will have an effect on the profitably 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 30 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,[15] 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.[16]

---

14 NRG Energy announced plans to add 2,700 MW at the STP nuclear plant and 800 MW at the Limestone coal plant in a June 21, 2006 press release.

15 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.

16 The California ISO does not report capacity and ancillary services net revenue separately, so it is shown as a combined block in Figure 30. 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

Figure 30: Comparison of Net Revenue of Gas-Fired Generation between Markets 2004 to 2006

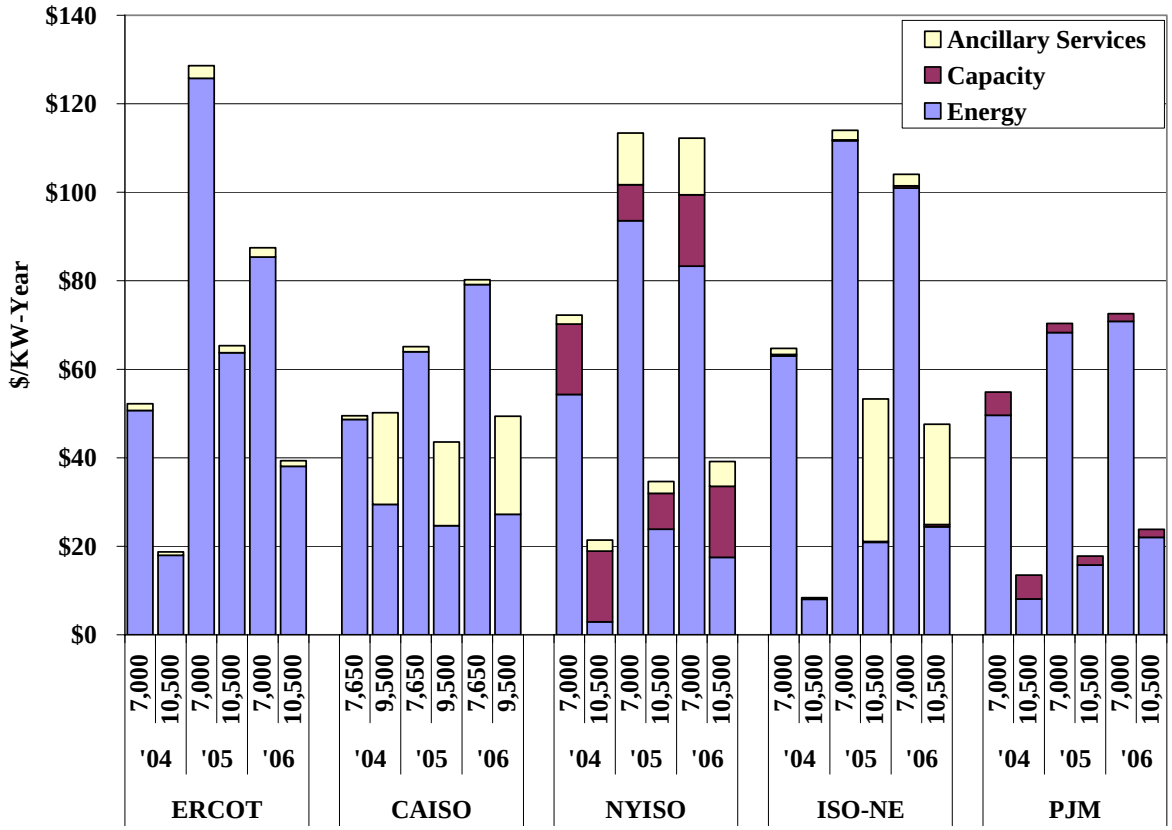


Figure 30 shows that net revenues increased slightly in California, New York and PJM from 2005 to 2006, and decreased in ERCOT and New England. These differences can be explained by several factors. First, ERCOT is much more dependent on natural gas than the other markets. The decrease in natural gas prices in the other regions does not translate as directly into lower electricity prices because natural gas units are displaced in many hours by other types of units. Second, many of the natural gas units in the Northeast are dual-fueled, allowing them to switch to oil when natural gas becomes relatively expensive. This causes the net revenue to fall for the hypothetical new units that can only burn natural gas. In 2006, the New York and New England markets exhibited net revenue in a range that might be sufficient to motivate investment in new gas-fired capacity, while net revenue in ERCOT, California and PJM likely would not likely be sufficient to support investment in new capacity. However, the costs of new investment can vary significantly by region due to widely varying costs of land, access to water and fuel, and other

Real-time and Day-ahead market, with the net revenues updated from 2004 to 2006.

regional factors, such as state and local tax and regulatory costs. 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 2006 were primarily affected by the following factors:

- Although continuing to decline relative to prior years, planning reserve margins in 2006 were approximately 16.5 percent, which is well 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 moderated in 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 Modified Competitive Solution Method ("MCSM") triggered price adjustments more frequently in 2006. MCSM is a PUCT-approved mechanism that was in effect in 2005 and through September 2006 that provided for an *ex post* reduction to the resulting market prices when all dispatchable balancing energy was exhausted. The average number of MCSM intervals per month almost doubled to over 26 per month in 2006 compared to less than 16 per month in 2005 for the months in which MCSM was in effect.
- The competitive performance of the ERCOT market improved in 2006.

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:

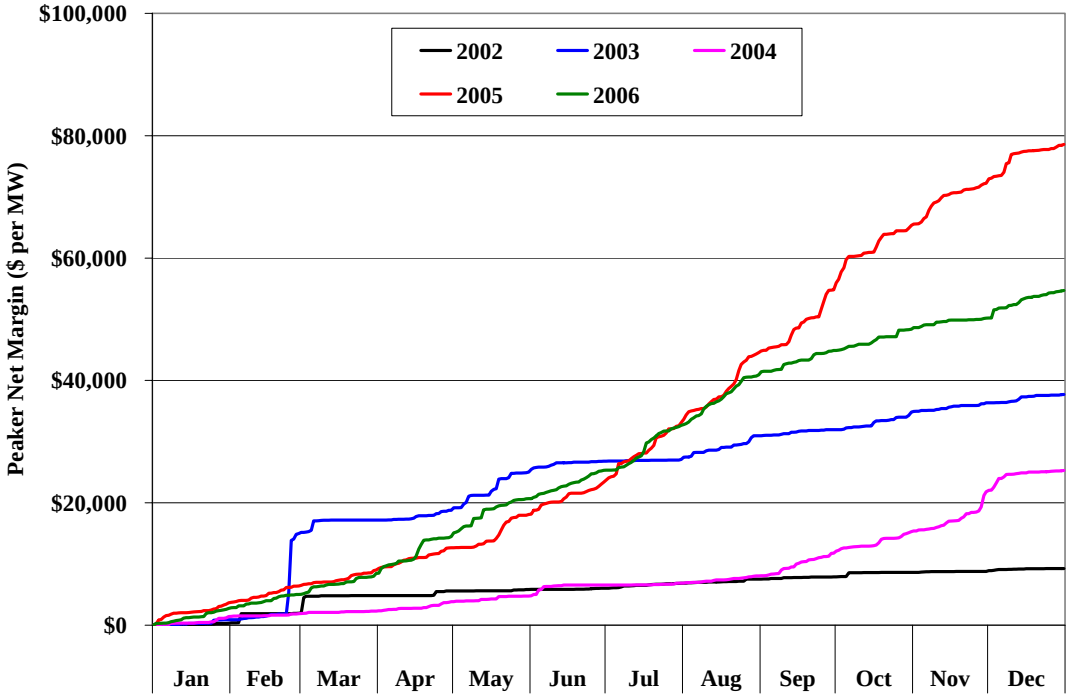
- A capacity market; and/or
- Shortage pricing provisions to ensure that prices rise appropriately in the energy and ancillary services markets to reflect the true costs of shortages when resources are insufficient to satisfy both the energy and ancillary services requirements.

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. The new rules also eliminated MCSM effective October 1, 2006.

The SPM also 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 31 shows the cumulative PNM that would have been produced for each year from 2002 through 2006.[17]

Figure 31: Peaker Net Margin
2002 to 2006



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

17 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.

to $85,000 per MW-year). Thus, as shown in Figure 31 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).

Unlike markets with a long-term capacity market, 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 for demand response and efficient incentives for new investment when required. During non-shortage conditions (*i.e.*, most of the time), the expectation of competitive 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 will remain 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).

Finally, the PUCT's energy-only market rule provides that the IMM may conduct an annual review of the effectiveness of the SPM. The IMM anticipates performing such a review in 2008 that will focus on the results of the first year of operation under the new rules, the outlook for future years, and potential modifications, if any, that may be required to ensure that the energy-only market achieves its intended objectives.

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. The results of this analysis lead us to make several recommendations to improve the operation of the current markets.

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. Finally, we analyze market participant resource plans to determine whether the information provided to ERCOT regarding generating units' projected commitment and output levels is affected by certain adverse incentives embodied in the ERCOT protocols.

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, QSEs may balance some or all of their 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 32 shows a scatter diagram that plots the ratio of the final load schedules to the actual load level during 2006. The ratio shown in the figure will be greater than 100 percent when the final load schedule is greater than the actual load.

Figure 32: Ratio of Final Load Schedules to Actual Load
All ERCOT – 2006

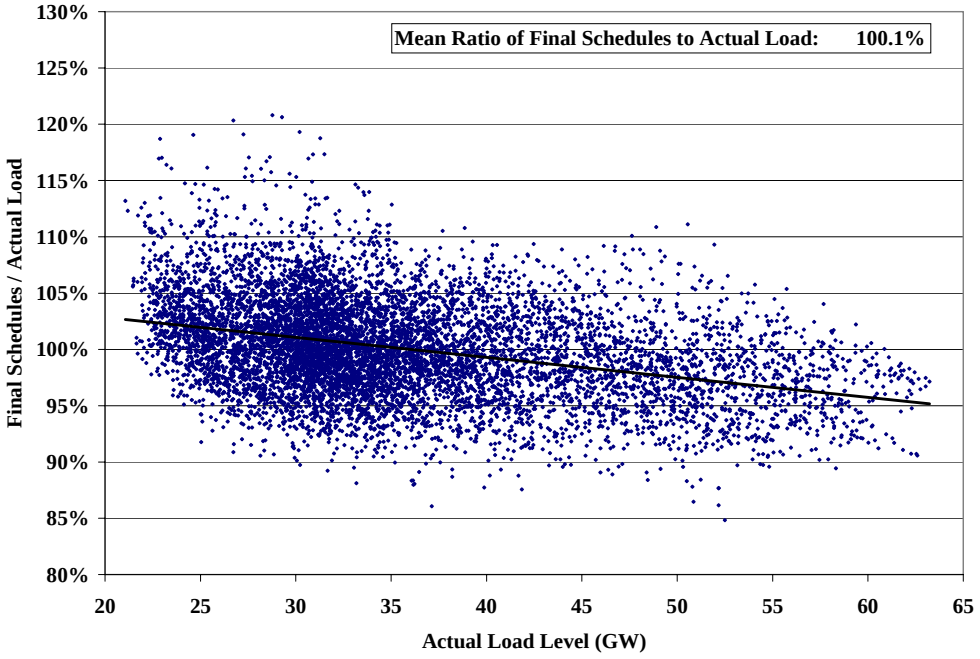


Figure 32 shows that final load schedules generally come very close to actual load in the aggregate, as indicated by an average ratio of the final load schedules to actual load of 100.1 percent. However, the figure also includes a trend line indicating that the ratio of final load schedules to actual load tends to decrease as load rises. In particular, the ratio given by the trend line is above 100 percent for loads under 37 GW and declines to 95 percent at higher load levels. The overall pattern shown in the figure above is similar to 2005, which exhibited the same

downward trend in final load schedules relative to actual load, although the average ratio was 101.2 percent.

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 33: Average Ratio of Final Load Schedules to Actual Load by Load Level All Zones – 2006

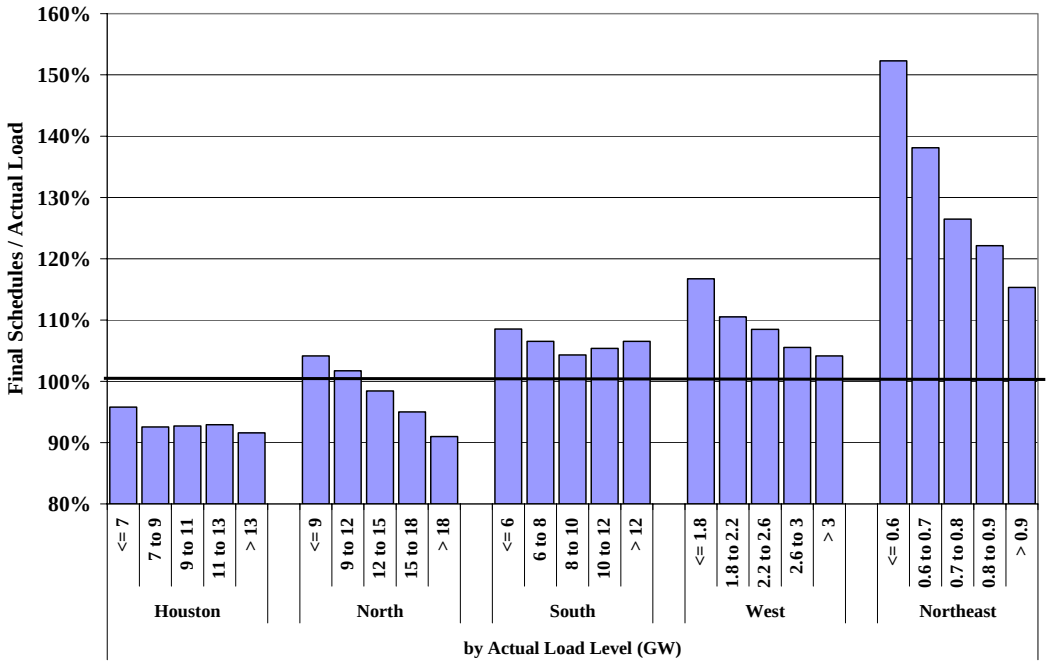


Figure 33 is a further analysis of final load schedules that shows the ratio of final load schedules to actual load evaluated at five different load levels for each of the ERCOT zones.

Figure 33 shows that:

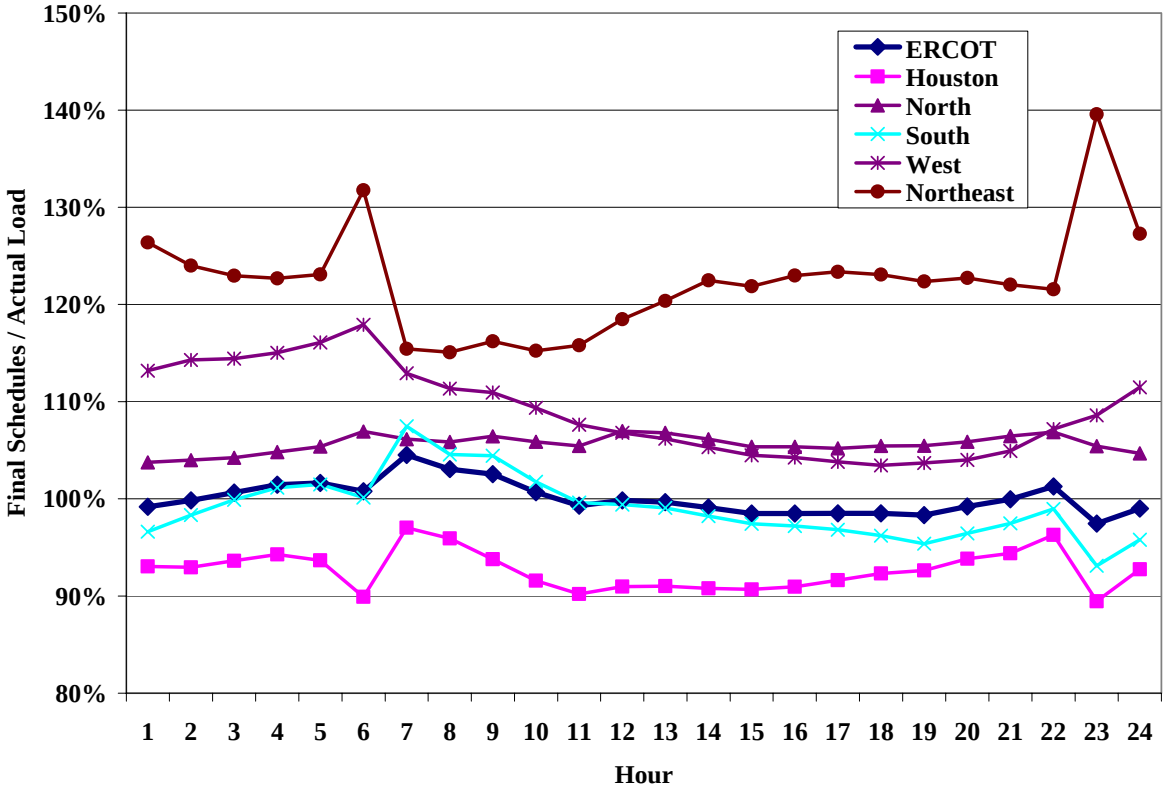
- The final schedule quantity decreases in each of the five zones as actual load increases, with the exception of the South zone which remains relatively flat as load changes.
- The South and West Zones are generally over-scheduled, although the ratios decline slightly as load increases.
- The Northeast Zone is consistently over-scheduled by a large margin. However, since the Northeast Zone accounts for less than 3 percent of ERCOT load, the total amount over-scheduled on average is about 190 MW.
- Houston is under-scheduled at most load levels, ranging from 4 percent at lower load levels up to 8 percent at high load levels.

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 Northeast Zone, and 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 34 shows the ratio of final load schedules to actual load by hour-of-day for each of the five zones in ERCOT as well as for ERCOT as a whole.

Figure 34: Average Ratio of Final Load Schedules to Actual Load All Zones by Hour of Day - 2006



This figure shows that on an ERCOT-wide basis, final schedules are close to actual load (between 99 percent and 102 percent) during hours ending 1 to 6. At hour ending 7, the ratio rises to 105 percent, the highest of any hour. By hour ending 10 through the remainder of the day, the ratio declines to a range between 97 percent and 101 percent.

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 34 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

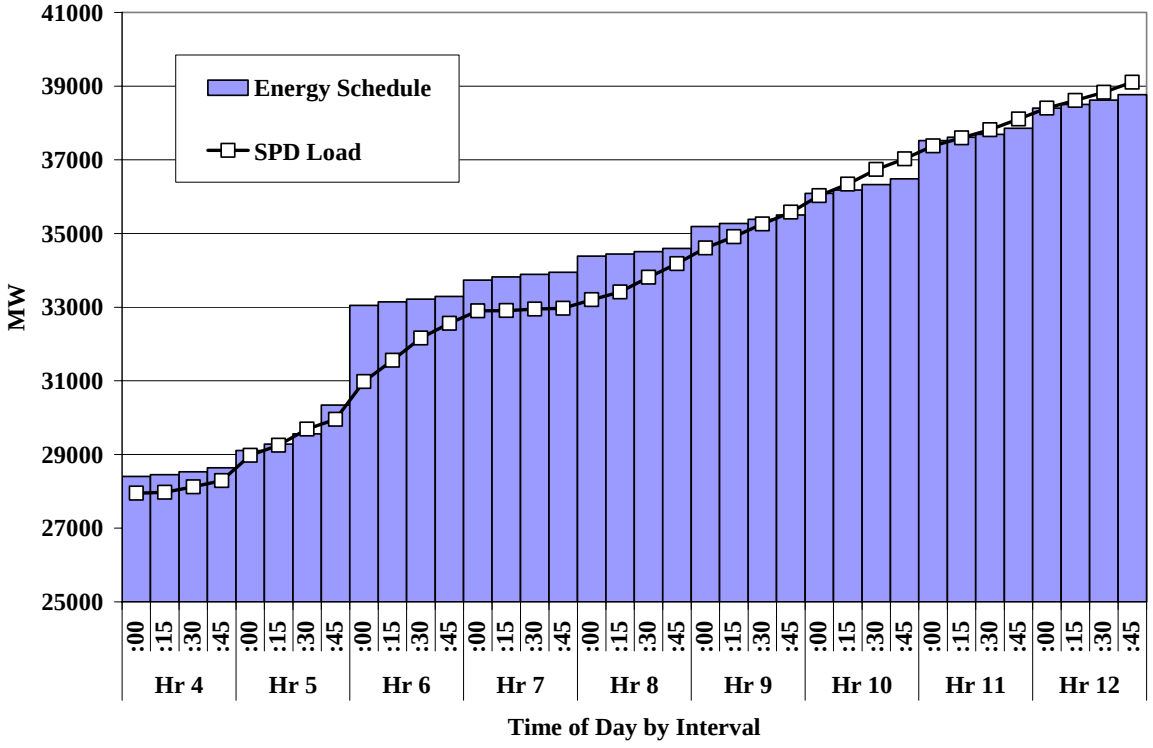
In the previous section, we analyzed balancing energy prices and load and found that while balancing energy prices are correlated to real-time load levels, other factors also have substantial

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 35 shows average energy schedules and actual load for each interval from 4 AM to 1 PM during 2006.

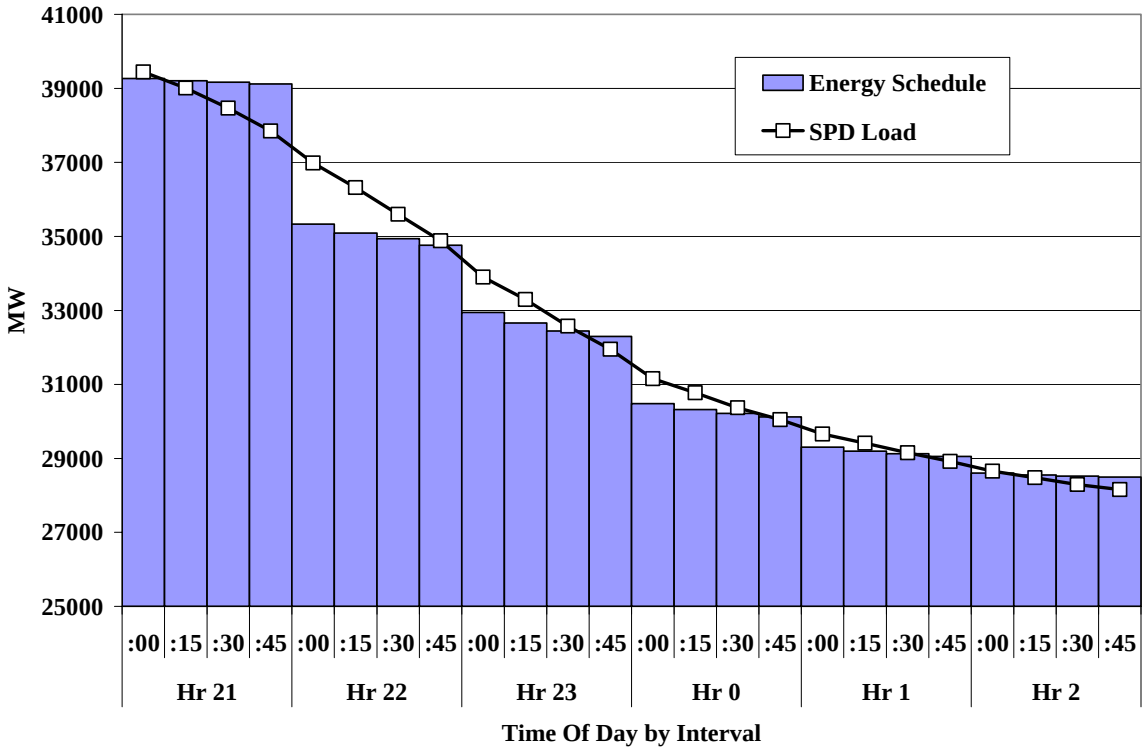
In general for ERCOT as a whole, energy schedules that are less than the actual load result in balancing energy purchases while energy schedules higher than actual load result in balancing energy sales. On average, load increases from approximately 28 GW to almost 39 GW in the nine hours shown in Figure 35. 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 35: Final Energy Schedules during Ramping-Up Hours 2006



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 2.7 GW from the last interval of the hour ending 6 AM to the interval beginning at 6 AM, while the average energy schedule increases by several hundred megawatts in the subsequent three intervals. The same scheduling patterns exist in the ramping-down hours. Figure 36 shows average energy schedules and load for each interval from 9 PM to 3 AM during 2006.

Figure 36: Final Energy Schedules during Ramping-Down Hours 2006



On average, load drops from approximately 39 GW to less than 29 GW in the six hours shown in Figure 36. 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-down hours, energy schedules decrease in relatively large steps at the top of each hour. For instance, the average energy schedule drops nearly 4 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 35 and Figure 36). This analysis is similar to that shown in Figure 16 and Figure 17, except instead of showing balancing energy prices relative to load, we show balancing energy prices relative to balancing energy deployments. Figure 37 shows the analysis for the ramping-up hours.

Figure 37: Balancing Energy Prices and Volumes
Ramping-Up Hours – 2006

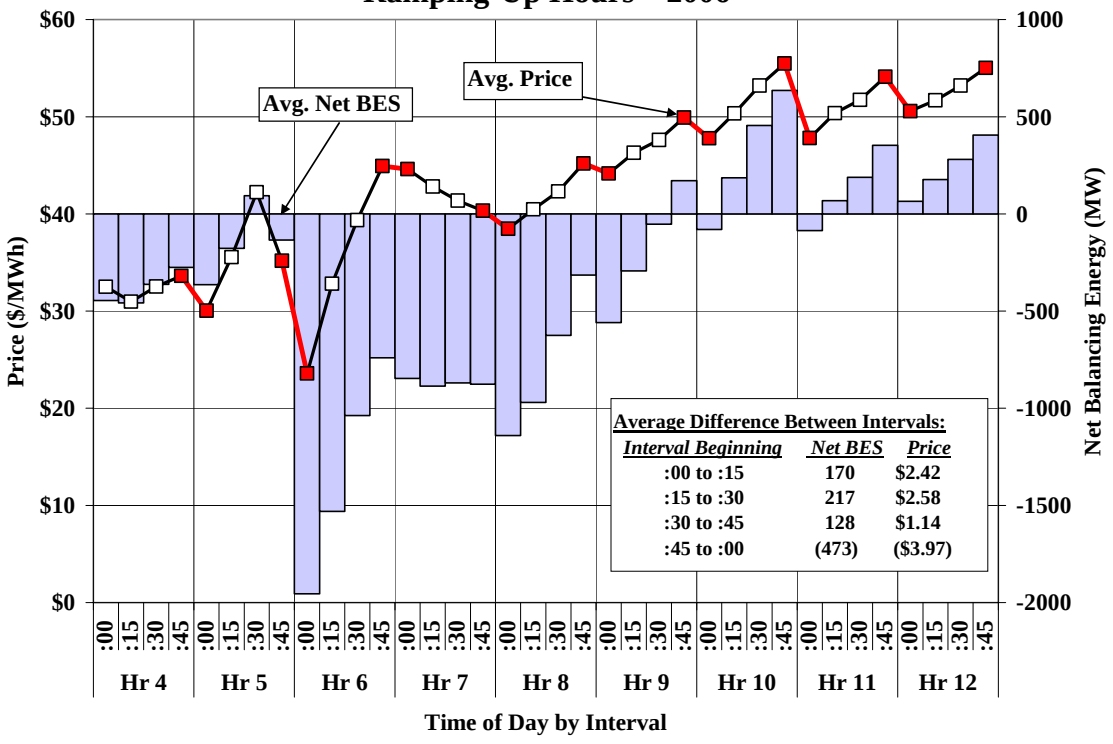
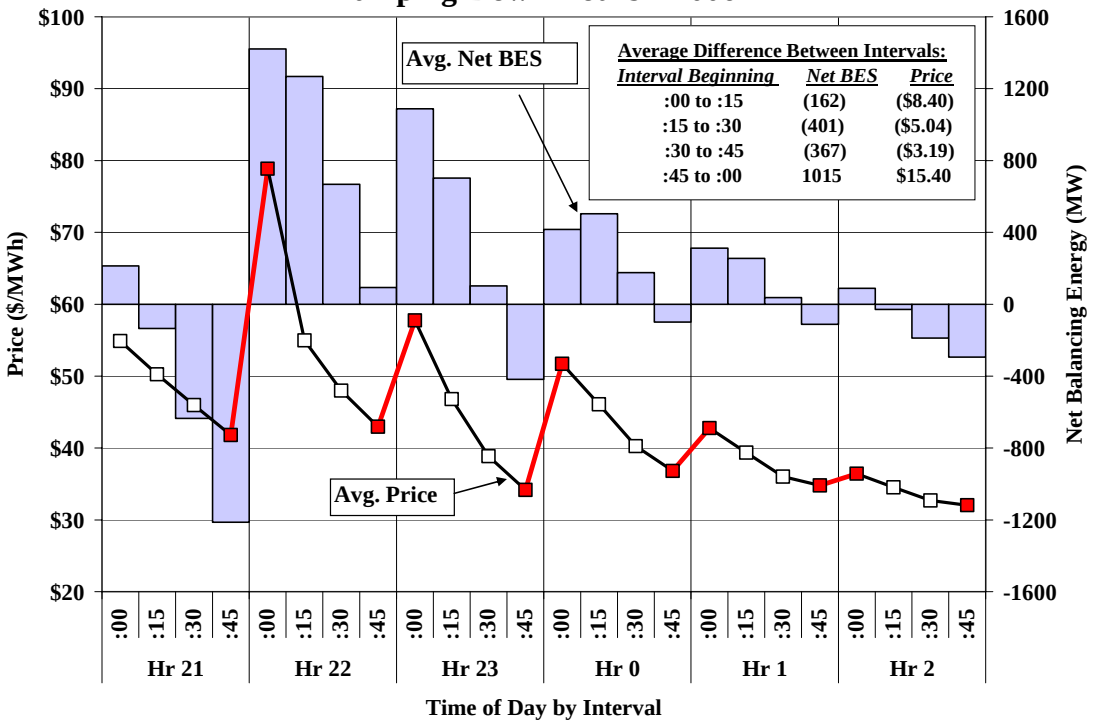


Figure 37 reveals two key aspects of the balancing energy market. First, as discussed above, balancing energy prices are highly correlated with balancing energy deployments. Second, with the exception of hour 7, there is a distinct pattern of increasing purchases during the hour. At the

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 38 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 by 2009 will largely resolve these inefficiencies.

Figure 38: Balancing Energy Prices and Volumes
Ramping-Down Hours – 2006



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 37 and Figure 38.

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 2006. 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.

C. Portfolio Ramp Limitations

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.[18] The operational implications associated with these issues continued in 2006 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.

D. Balancing Energy Market Offer Patterns

In this section, we evaluate balancing energy offer patterns by analyzing the rate at which capacity is offered. In Figure 39, we show the average amount of capacity offered to supply balancing up service relative to all available capacity. The analysis in this section differs from similar analyses in prior reports in the following important respect. In prior reports, un-offered capacity calculations included capacity that existed but was not offered. They did not attempt to quantify the amount of un-offered capacity that was actually available, and practicable to offer, given the ERCOT scheduling timelines, operating rules and conditions, and technical or commercial limitations that might limit a QSE's ability to offer capacity in the ERCOT market. In contrast, the approach used for the analysis of un-offered capacity in this section is focused on online, available capacity for which there is a reasonable expectation that the energy can be produced in light of the factors and considerations listed above. Specifically, the methodology for determining the quantities of un-offered capacity in this section are as follows:

Un-offered Capacity is equal to:

> Total Online Capacity plus qualified, off-line quick-start combustion turbines not providing non-spinning reserve;

[18] 2005 SOM Report at 68-76.

Less:

- Scheduled Generation;
- Up Balancing Energy Offers;
- Residual Reliability Must Run ("RMR") capacity;
- Residual Qualifying Facility ("QF") capacity from "non-bid" QF resources;
- Residual capacity from wind turbines;
- Scheduled energy (25 percent) from wind turbines that are not in wind-only QSEs;
- Generation Regulation Up obligation;
- Generation Responsive Reserve obligation; and
- Non-spinning Reserve obligation met by online resources.

The balancing energy offers are divided into that which is ramp-constrained, and would not actually be capable of supplying balancing energy in a single 15-minute interval, and that which is non-ramp-constrained, and thus would be available to supply balancing energy in a single 15-minute interval. Total capacity includes the maximum capacity of resources that are flagged as online in the final resource plan submitted by the QSE, as well as qualified, off-line quick start units that are not flagged as providing non-spinning reserve. Scheduled generation, regulation up, responsive reserve from generation resources and up balancing energy offers are deducted from the total capacity. Non-spinning reserve is deducted from the total online capacity for each QSE to the extent that the QSE has insufficient offline capability flagged as non-spinning reserve to meet it obligation. Residual RMR capacity is deducted from total capacity because, while such capacity could technically be offered, the financial incentives as set forth in the ERCOT Protocols are insufficient to provide a reasonable expectation that the residual RMR capacity would be offered. Capacity from a QF that is designated as "non-bid" is also deducted from the total online capacity. Under the ERCOT Protocols, QFs are allowed to specify capacity as "non-bid" for the purpose of local congestion management to reflect technical or commercial limitations associated with their specific operating requirements; therefore, such capacity is not reasonably expected to be offered as balancing energy. Residual wind capacity is deducted from the total online capacity to reflect the uncontrollable nature of wind turbines.

Finally, 25 percent of the scheduled wind generation from non-wind-only QSEs is deducted from total capacity to reflect the fact that, to the extent the wind does not produce as scheduled, the

portfolio balancing requirement for non-wind-only QSEs requires that sufficient capacity be reserved for this purpose. The final result of these deductions from the total online capacity plus qualified quick-start units that are not flagged as providing non-spinning reserve is the quantifiable un-offered capacity that could practicably and reasonably be expected to be offered, although, as discussed later in this section, there are several other structural impediments to offering even this capacity that are more difficult to quantify. The offered and quantifiable un-offered capacity data is shown for the peak hour of the day on a monthly average basis for 2006 in Figure 39.

Figure 39: Balancing Energy Offers Compared to Total Available Capacity
Daily Peak Load Hours – 2006

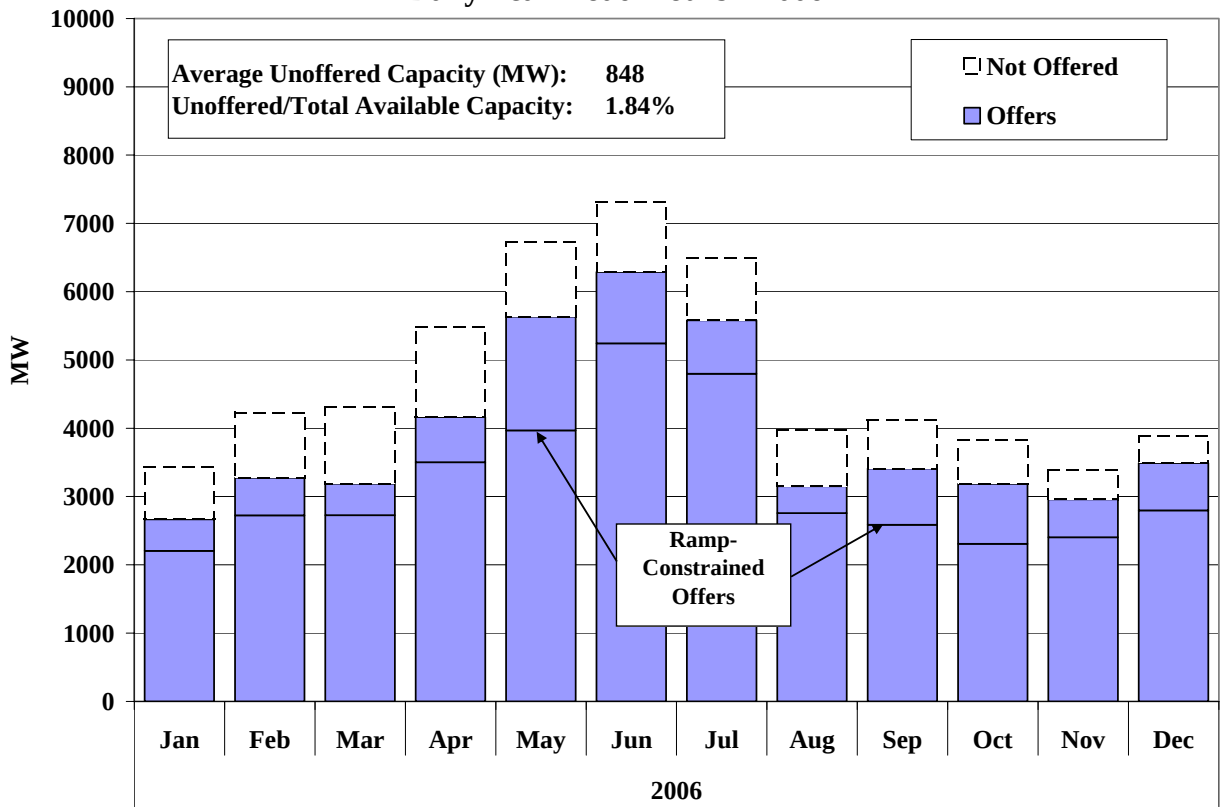


Figure 39 shows the trend in 2006 over time in quantities of energy available and offered to the balancing energy market. Up balancing offers are divided into the portion that is capable of being deployed in one interval and the portion which would take longer due to portfolio ramp rate offered by the 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 40 shows the same data as the previous figure, but arranged by load level for daily peak hours in 2006. 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 40: Balancing Energy Offers Compared to Total Available Capacity Daily Peak Load Hours – 2006

The figure indicates that in 2006, the average amount of capacity available to the balancing market increased gradually up to 55 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 40 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.

In regard to the residual un-offered capacity shown in the previous two figures, there are several possible explanations for the quantity of un-offered on-line and quick start capacity that was not quantifiable in the preceding analysis. First, issues related to ramp rates can affect the offer levels. Currently, a QSE is able to submit one up-balancing ramp rate for its portfolio per hour per zone, and the ramp capability tends to decrease as more of the offer is deployed. Thus, many QSEs may feel compelled to not offer slow ramping capability near the high sustainable limits of their resources. Moreover, to the extent that a supplier's portfolio includes slower-ramping low-cost resources, the supplier may not offer a significant share of its higher-cost resources. The supplier faces the risk that it will receive a balancing energy deployment that exceeds the ramp capability of the low-cost resources that would compel it to dispatch its high-cost resources at a loss.

Second, QSEs are subject to compliance measures in relation to their performance, which may include penalties. This may limit a QSE's willingness to adopt a very aggressive offer strategy in consideration of operational risks in real-time that may affect its ability to perform at the outer bounds of its rated capability. In aggregate, if such risks were managed by a conservative reduction of offered capacity of just one percent, this would represent 500 MW of un-offered capacity in an hour where 50,000 MW of rated generation capability was online.

Lastly, the duct firing ranges of combined cycle units and steam turbines can also be difficult to offer in to the balancing market for several reasons. A supplier may incur "start-up costs" associated with operating in the duct firing range. Typically, generators have slower ramp rates in their duct firing ranges and may incur losses if a brief price spike is followed by relatively low prices. Also, many generators cannot operate in the duct firing range and provide regulation simultaneously, so that dispatching in this range for energy could result in non-performance in the provision of ancillary services.

E. Resource Plan Changes

QSEs must have sufficient generation on-line to support their energy schedules and offers, and they are required to inform ERCOT about which resources they plan to use to satisfy their obligations. They do this by submitting resource plans at various points in the day-ahead and the operating day. While QSEs are expected to make their best effort to accurately forecast how they will operate their units, the resource plans are not financially binding and can be changed until shortly before real-time.[19] Resource plans are used by ERCOT in some of its reliability assessments before real-time and to make additional commitments to maintain reliability. Therefore, it is important for ERCOT to have accurate information in the resource plans that QSEs submit in order to avoid taking unnecessary and sometimes costly actions to maintain reliability.

It is important for QSEs to have the flexibility to incorporate new information prior to real time, such as demand forecast changes, generation and transmission outages, and other factors that suggest more or less resources will be needed in real-time. These factors can lead QSEs to significantly revise their resource plans after the day ahead. Under the current ERCOT market, however, there are other reasons why a participant may consistently provide unreliable information in its day-ahead resource plan, then revise the resource plan prior to real time when the balancing energy market is run. Participants could submit unreliable information as part of a gaming strategy, or they might unintentionally submit unreliable information.

This section of the report analyzes the changes in the resource plans between the day ahead and real time and differences between the real-time resource plan and actual operation. Specifically, we evaluate units that are frequently committed out-of-merit or frequently dispatched out-of-merit and receive substantial out-of-merit payments. Such units receive additional payments from ERCOT and we investigate whether market participants may engage in strategies to increase the probability of receiving these payments.

We first analyze the behavior of suppliers that are the primary recipients of payments by ERCOT for out-of-merit capacity or replacement reserves. OOMC or RPRS occurs when ERCOT

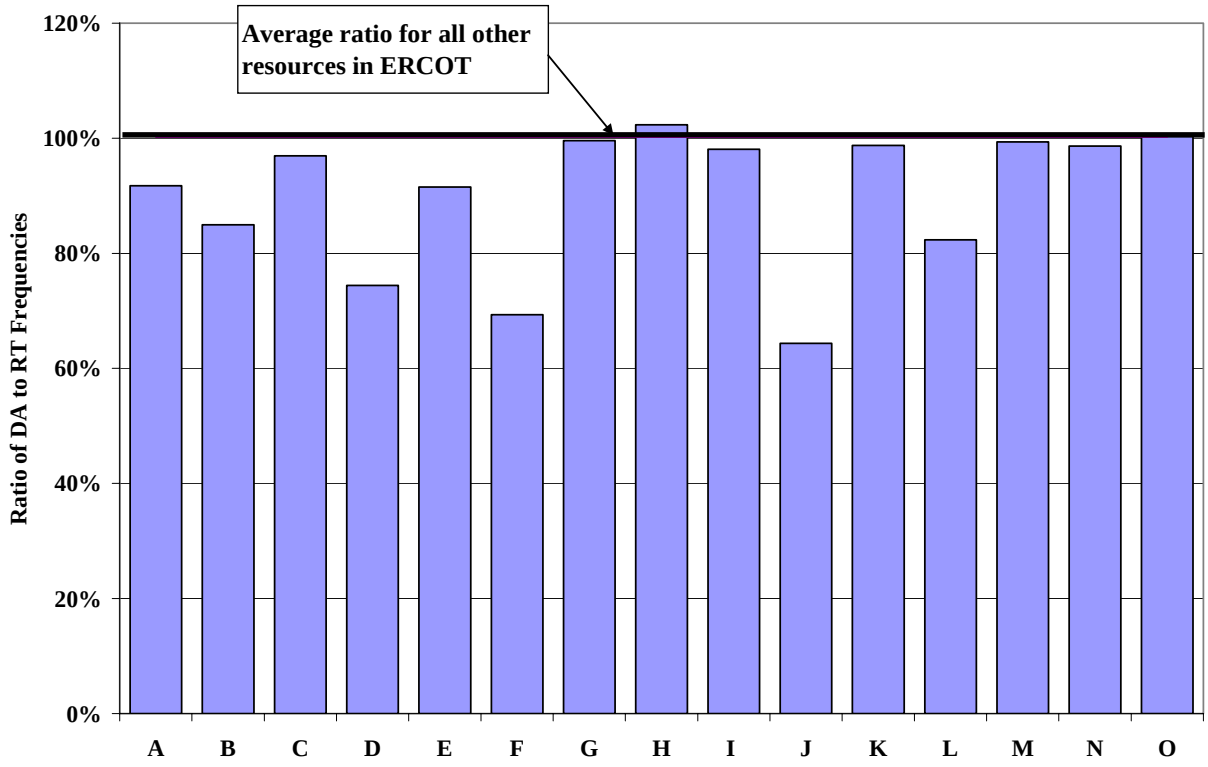
[19] While resource plans are not financially binding, the real-time planned generation is used in the OOME payment formulas to determine the amount of megawatts deployed by the OOME instruction.

instructs a unit that is not committed in the QSE's day-ahead resource plan to start in order to ensure sufficient capacity in real time to meet forecasted load and manage transmission constraints. When suppliers receive OOMC or RPRS instructions, they receive payments from ERCOT that are designed to cover an estimate of the cost of starting the unit plus the cost of running at the minimum level. However, the unit retains any profits from sales above the minimum level into the balancing market. Thus, for units with significant commitment costs that are frequently committed out of merit, a supplier has the financial incentive to show the unit as uncommitted in the day-ahead resource plan to compel ERCOT to commit the unit. This supplier can subsequently commit the unit before real time if it is not called upon by ERCOT.

Because of the incentives presented by the OOMC and RPRS processes, we would expect suppliers that anticipate having units committed out-of-merit and that would benefit from the resulting payments to avoid showing the units as committed until after the out-of-merit commitments are announced. We examined the patterns of commitment for units that receive substantial OOMC and RPRS payments. Figure 41 shows the ratio of day-ahead resource plan commitments to actual real-time commitments during 2006 for the 15 generating plant sites receiving the largest OOMC and RPRS payments.[20] The generating plants in the figure are sequenced from highest to lowest payment from left to right, and the total payments for all generating plants in the figure constitute two-thirds of the total OOMC and RPRS payments in 2006. Hours when the resources are under OOMC, RPRS or OOME instructions are not included in order to assess systematic changes made voluntarily by market participants. The units are shown in decreasing order of payments received from ERCOT. To show how the commitment of these units compares to all other units in ERCOT, the figure also shows the capacity-weighted average ratio of day-ahead to real-time resource plan commitments for all units.

[20] For the purpose of this analysis, all generating units at the same electrical location are group into a single generating plant.

Figure 41: Ratio of Day-Ahead to Real-Time Resource Plan Commitments* Frequent OOMC Resources – 2006



* Excluding hours when resources were under OOMC instructions or dispatched out-of-merit.

While most of the generating plants shown in Figure 41 have ratios that are comparable to the market as a whole that reflects consistency between the day-ahead and real-time resource plans, a minority of the generating plants have ratios less than 80 percent. The results shown in this figure are consistent with the concern that some QSEs may wait until after the OOMC and RPRS process to commit units that are necessary for reliability.

For the generating plants shown in Figure 41, uplift payments for OOMC and RPRS commitments are substantial enough to provide significant incentives to behave in ways that maximize the likelihood of receiving them. Figure 41 suggests that some QSEs with resources that frequently receive OOMC or RPRS instructions may delay the decision to commit those units until after ERCOT determines which resources to select for OOMC or RPRS. This approach to address capacity insufficiency in the Protocols has several deleterious effects on the market. First, ERCOT incurs OOMC and RPRS costs to commit resources that are otherwise economic and that should be committed voluntarily without supplemental payments. Second,

when resources are committed out-of-merit, some other resources committed in day-ahead resource plans will no longer be economic. This can result in over-commitment of the system. However, the QSE generally has the opportunity to modify its other commitments after it receives the OOMC or RPRS instruction and often does so. Third, this conduct tends to undermine the accuracy of the information that ERCOT depends on to manage reliability. Ultimately, this can cause ERCOT to take a variety of costly actions, including making out-of-merit commitments that should not be necessary. These problems stem from the de-centralized process for unit commitment under the current market design, and underscore the reliability and efficiency benefits of the centralized commitment process that will be implemented with the nodal market re-design.

In our next analysis, we evaluate incentive issues associated with out-of-merit dispatch in real-time. In order to resolve intrazonal congestion in real-time, ERCOT will increase or decrease a unit's output (out-of-merit energy or "OOME") to reduce the flow on a constrained transmission facility within a zone. When the unit is dispatched up in this manner (*i.e.*, OOME Up), it receives payments corresponding to the higher of the estimated running cost of the out-of-merit portion of the unit (plus a margin), or the balancing energy price. Although the potential profits are limited by the formula used to calculate the OOME payment, the system can still provide incentives to schedule resources strategically.

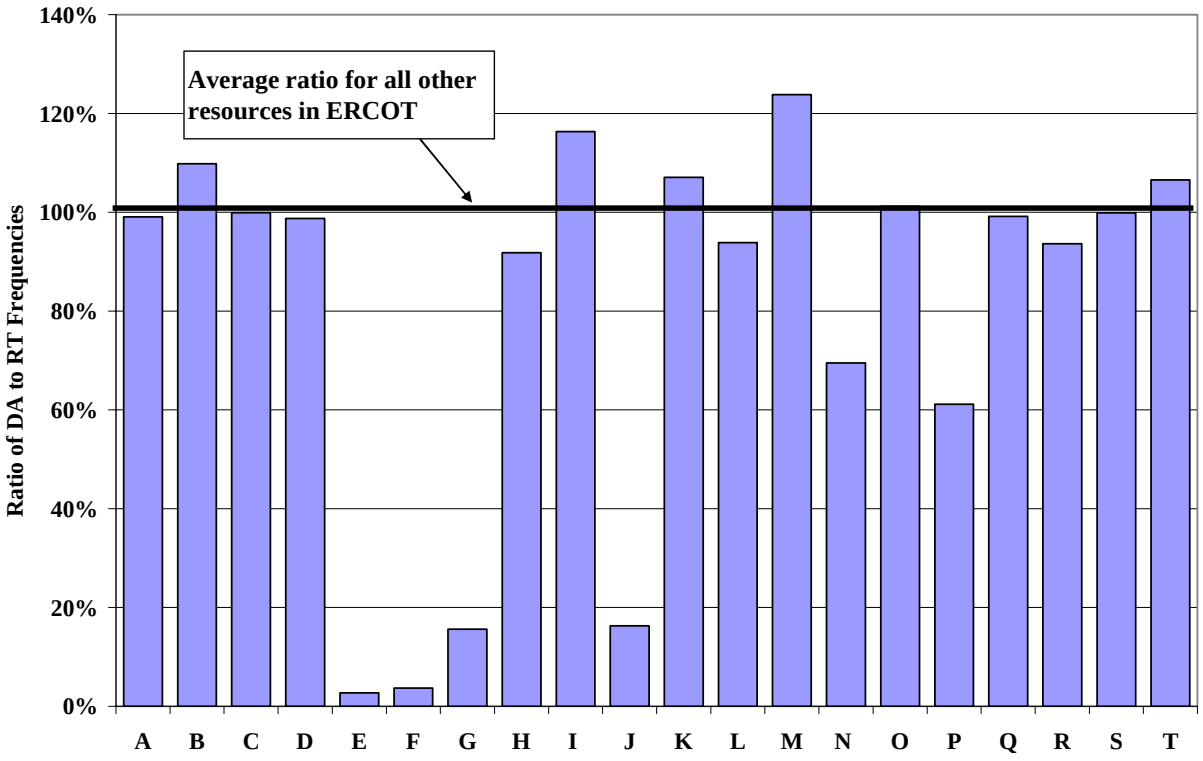
If a supplier is able to predict which of its units may be dispatched out-of-merit, it may under-schedule those units and over-schedule other units in its portfolio.[21] Although this resource plan output may not be efficient, it can be effective at compelling an OOME instruction and the associated uplift payment. Following the OOME instruction, the supplier can adjust its over-scheduled units to restore an economic dispatch pattern. If the supplier can accurately predict when the units will be called out-of-merit, this strategy can generate significant uplift payments. When the unit is not called for out of merit dispatch, the supplier can adjust the output levels of the units in its portfolio to correct the inefficient schedule.

Under this type of strategy, one would expect that units often needed to resolve congestion would be frequently under-scheduled. To test for this strategy, Figure 42 shows the ratio of real-

[21] "Scheduling" in this context refers to the unit-specific planned generation in the QSEs' resource plans.

time resource plan scheduled output to actual generation for the 20 units that received the highest average payments for OOME Up per MWh of generation across all hours of 2006.[22]

Figure 42: Ratio of Real-Time Planned Generation to Actual Generation* Frequent OOME-Up Resources – 2006



* Excluding hours when resources were under OOMC instructions or dispatched out-of-merit.

To include only the scheduling and dispatch decisions made solely by the supplier, the ratio does not include hours when the resource was under OOMC or OOME instructions. The 20 resources shown in Figure 42 are presented in decreasing order of average payments, from $3.24 per MWh of generation across all hours for the unit on the far left to $0.22 per MWh for the unit on the far right. The generation-weighted average ratio of real-time resource plan output to actual generation for the whole ERCOT market is also shown for reference.

Of the 20 resources shown in Figure 42, 4 have ratios of less than 50 percent while ten have ratios between 50 and 100 percent. The other units in ERCOT had a weighted average ratio of

[22] To focus on the most significant units, the analysis excludes resources received OOME up instruction less than 5 times within the year as well as resources that operated in-merit for fewer than 10 hours.

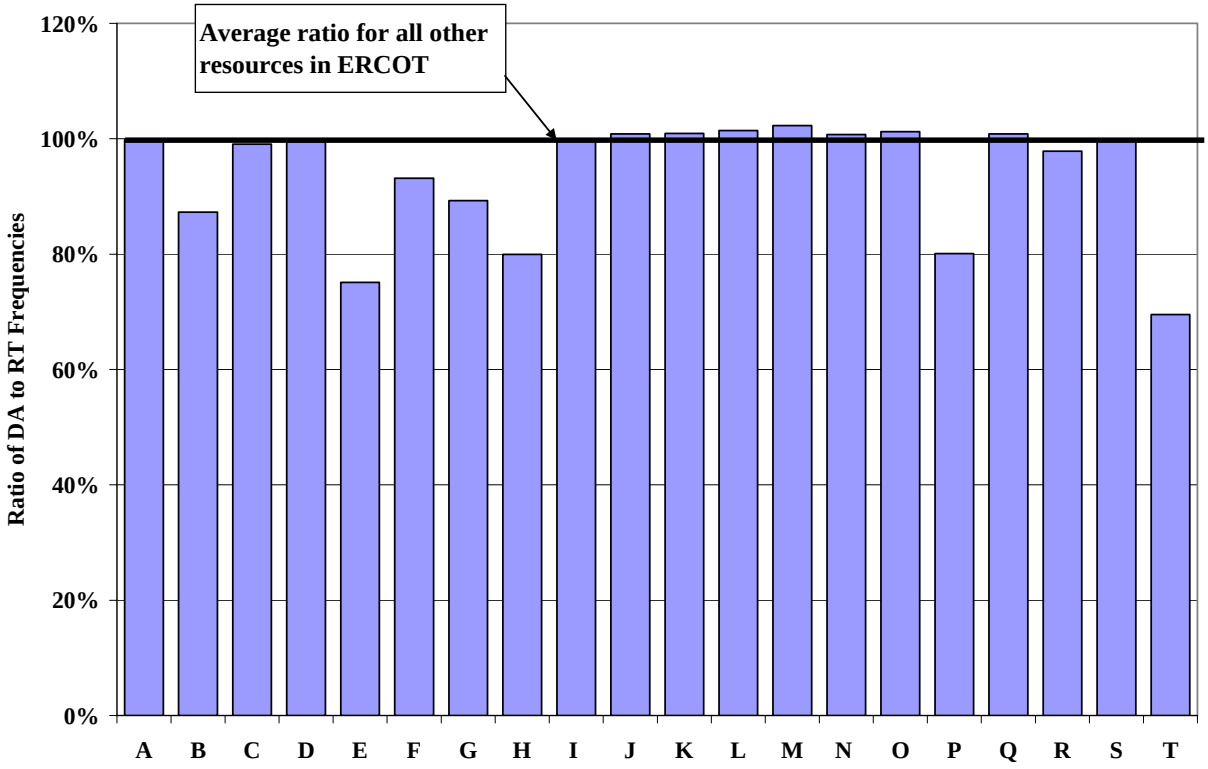
101 percent during the period, reflecting consistency between the scheduled output and actual generation. The data suggests that resources frequently providing OOME up are sometimes included by the QSEs in the real-time resource plans at output levels that are lower than their actual output. This is consistent with the hypothesis that the OOME procedures may provide inefficient incentives that lead QSEs to submit inaccurate resource plans.

We next evaluate the incentives associated with providing OOME down. The incentives associated with rules for OOME down payments are the reverse of the incentives for OOME up payments. Since ERCOT pays units to reduce output from the real-time resource plan output levels, a supplier able to foresee the need for an OOME down instruction can over-schedule the unit to compel the OOME down action by ERCOT. If the OOME down settlement rules provide strong incentives to engage in this conduct, the units that frequently receive OOME down instructions should be consistently over-scheduled. However, we would note before presenting our analysis that the magnitude of payments for OOME down is far lower than the magnitude of uplift payments for OOME up.

Figure 43 shows the ratio of real-time resource plan output to actual generation for the twenty resources that earned the highest average payments for providing OOME down (on per MWh basis) in 2006.[23] The figure shows units that received the highest OOME down payments for their total production. The resources are shown in decreasing order of the average OOME down payments received per MWh of output, ranging from $0.38 per MWh on the far left to $0.04 per MWh on the far right. For comparison purposes, the figure also shows the generation-weighted average ratio of real-time resource plan output to actual generation for all other units.

[23] This analysis excludes resources with received OOME down instruction less than 5 times within the year.

Figure 43: Ratio of Real-Time Planned Generation to Actual Generation* Frequent OOME-Down Resources – 2006



* Excluding hours when resources were under OOMC instructions or dispatched out-of-merit.

None of the twenty resources shown in Figure 43 had a ratio that was significantly above 100 percent. The figure above reflects good consistency between the planned output level and actual generation for OOME down units. Thus, there is no indication that frequent OOME down units have systematically over-scheduled their resources to earn more OOME uplift.

III. DEMAND AND RESOURCE ADEQUACY

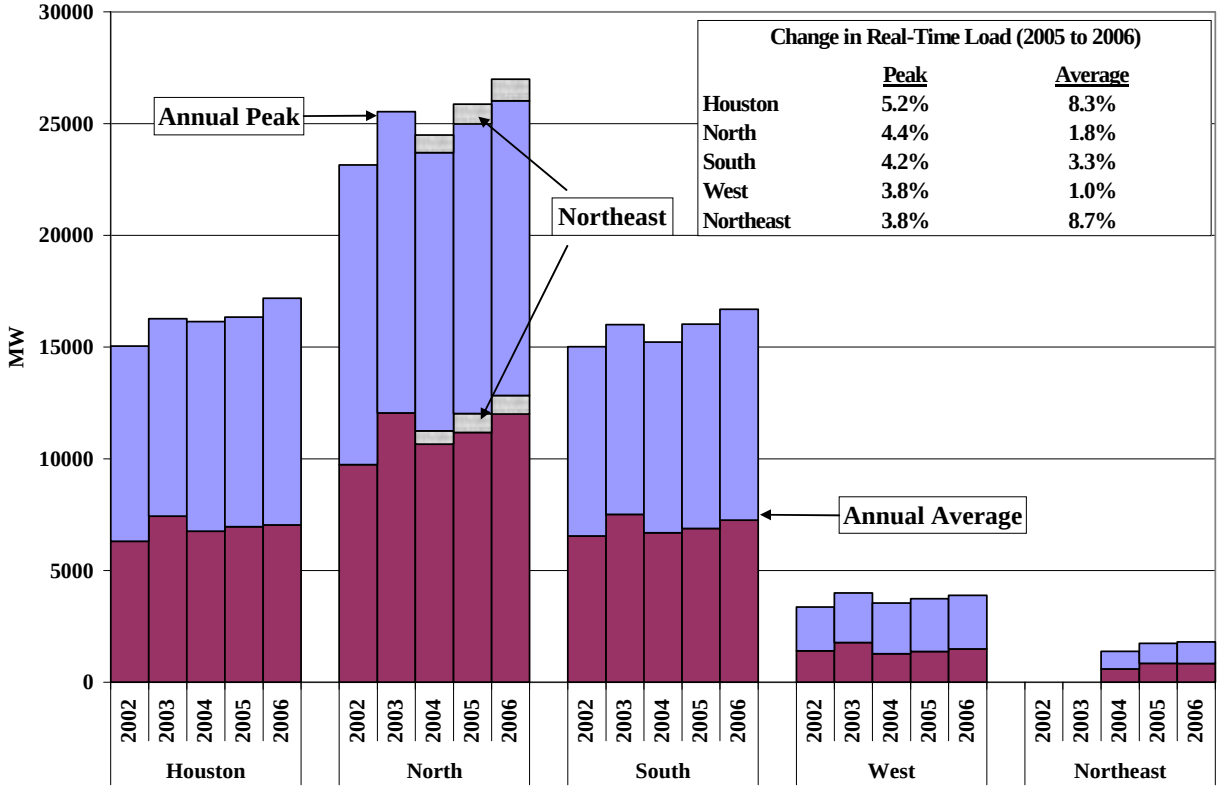
The prior sections of this report reviewed the market outcomes and provided analyses of a variety of factors that have influenced the market outcomes. This section reviews and analyzes the load patterns during 2006 and the existing generating capacity available to satisfy the load and operating reserve requirements.

A. ERCOT Loads in 2006

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 2006 are examined in this subsection and summarized in Figure 44.

This figure shows peak load and average load in each of the ERCOT zones from 2002 to 2006. 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 37 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 and Northeast Zone are the smallest (with about 7 percent and 3 percent of the total ERCOT load). Figure 44 shows the annual non-coincident peak load for each zone. This is the highest load that occurred in a particular zone for one hour during the year; however, the peak can occur in different hours for different zones. As a result, the sum of the non-coincident peaks for the five zones was greater than the annual ERCOT peak load.

**Figure 44: Annual Load Statistics by Zone*
2002 to 2006**



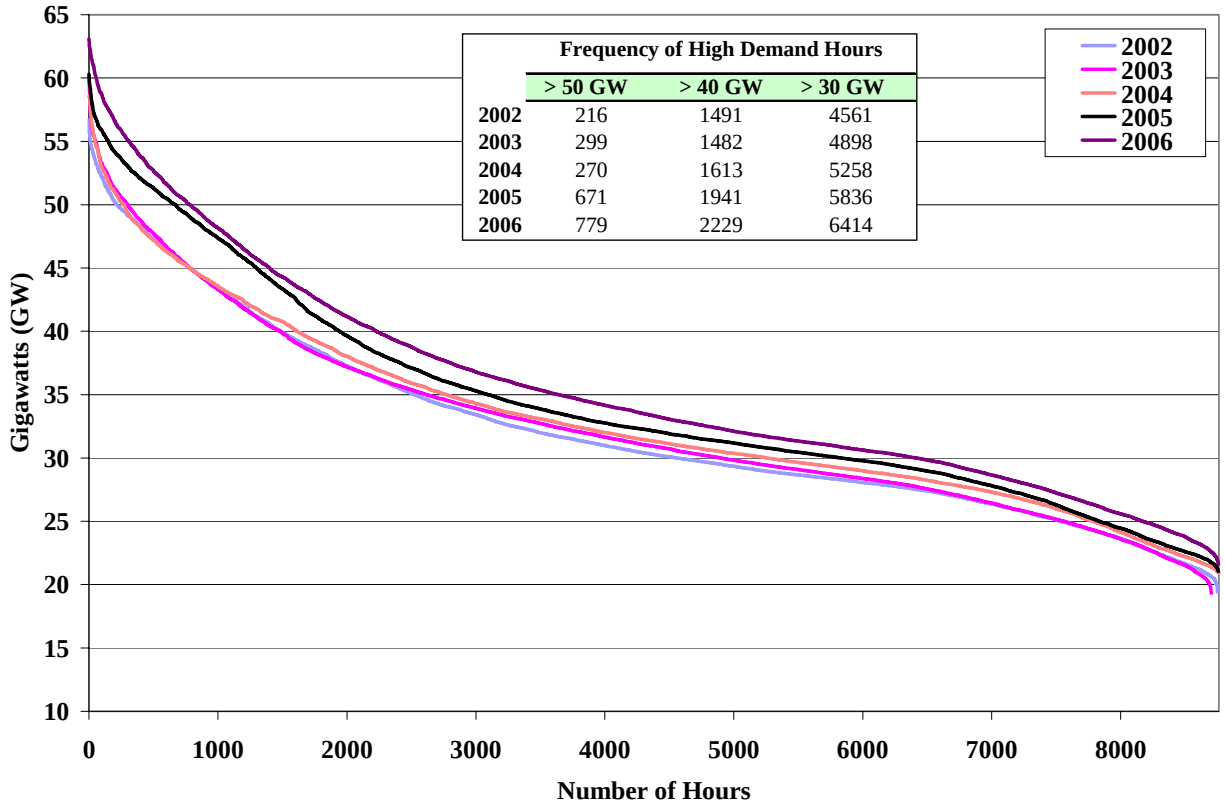
* The figure above is based on the load that SPD uses to schedule supply in the balancing energy market. This can differ from actual load in individual intervals.

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 45 compares load duration curves for each year from 2002 to 2006. A load duration curve shows the number of hours (shown on the x-axis) that load exceeds a particular level (shown on the y-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. The highest load hours occur in the summer months, and ERCOT dispatched generation to meet a record peak demand of 63 GW in August 2006.[24]

[24] This value is the total load to be served in real-time as represented in ERCOT's Scheduling, Pricing and Dispatch software (including transmission and distribution losses), and may differ from settlement values.

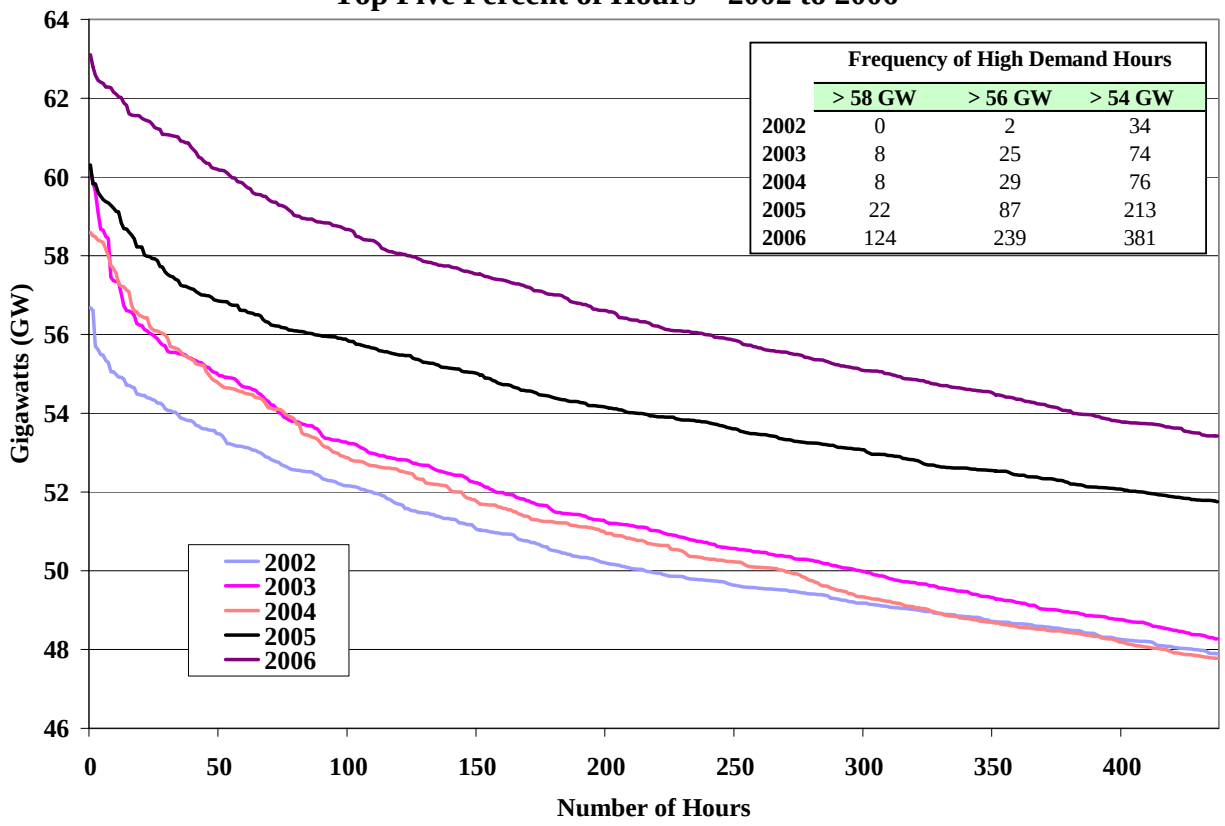
Figure 45: ERCOT Load Duration Curve
All Hours – 2002 to 2006



As shown in Figure 45 , the load duration curve for 2006 lies above the curves for the previous four years. Load increased more from 2005 to 2006 than it did in the previous three years on average. In 2006, there were 15 percent more hours when load exceeded 40 GW than in 2005.

To better show the differences in the highest-demand periods between years, Figure 51 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 2002 to 2005, the increase from 2005 to 2006 was much larger during the peak hours. Load exceeded 58 GW in 124 hours in 2006, 22 hours in 2005 and eight hours in 2003 and 2004. In 2002, demand was not higher than 58 GW in any hour. The same pattern prevailed at lower load levels with 2006 demand being considerably higher than in previous years.

Figure 46: ERCOT Load Duration Curve
Top Five Percent of Hours – 2002 to 2006



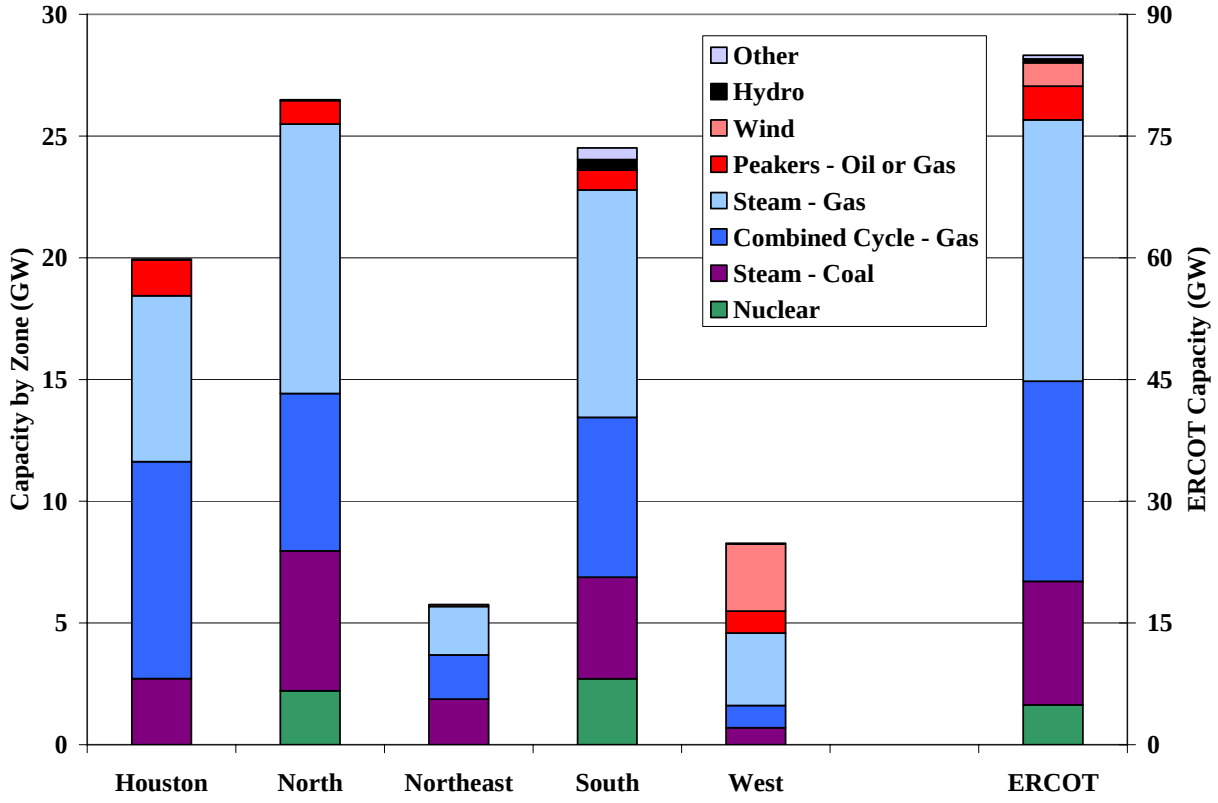
	Frequency of High Demand Hours		
	> 58 GW	> 56 GW	> 54 GW
2002	0	2	34
2003	8	25	74
2004	8	29	76
2005	22	87	213
2006	124	239	381

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 63 GW while the 95th percentile was lower than 54 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 9 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 47 shows the installed generating capacity by type in each of the ERCOT zones.

Figure 47: Installed Capacity by Technology for each Zone 2006



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 76 percent of generation capacity in ERCOT as a whole, and 86 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 31 percent of capacity, the South Zone 29 percent, the Houston Zone 23 percent, the West Zone 10 percent, and the Northeast Zone 7 percent. The North Zone and Houston are typically importers of power, while the Northeast Zone exports significant quantities because it has over two times more generation than its peak zonal load. Because large amounts of power flow out of the South Zone into the North Zone and Houston, the South-to-North CSC and the South-to-Houston CSC experienced the greatest amounts of congestion during 2006.

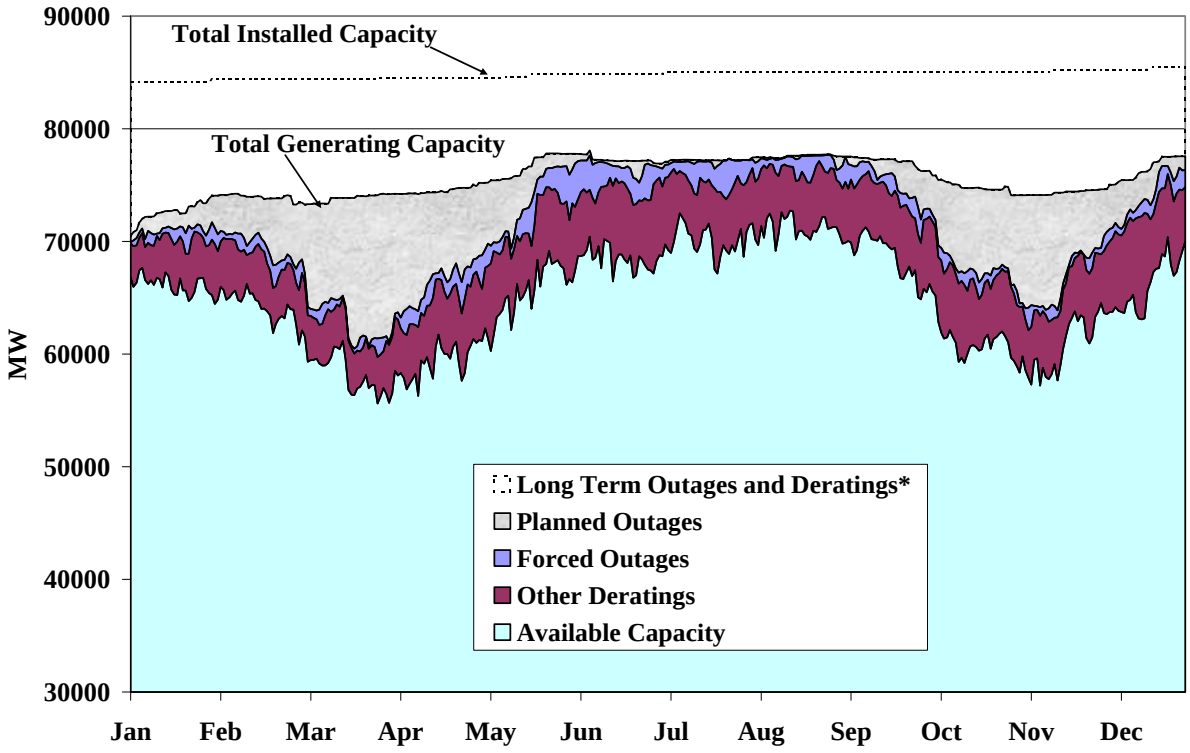
1. Generation Outages and Deratings

Figure 47 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 48 below shows a breakdown of total installed capability for ERCOT on a daily basis during 2006. 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.

The long-term deratings category includes any outages and deratings lasting for 60 days or longer while the remaining outages and deratings are included in the short-term categories. We generally separate the long-term outages because it provides an indication of the generating capacity that is generally not available to the market, which typically exceeds 10 GW. Long-term deratings can occur for several reasons. First, some of this capacity may be out-of-service for extended periods due to maintenance requirements. Second, if their owners predict that wholesale market prices will not be sufficiently high to justify the periodic costs required to keep them available, some units may go out-of-service temporarily. Third, the owners of some cogeneration plants routinely use steam output to support their processes rather than generate electricity. However, a large share of these deratings reflect output ranges on generating units that are not capable of producing up to the full installed capability level.

Figure 48: Short and Long-Term Deratings of Installed Capability**
2006



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

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

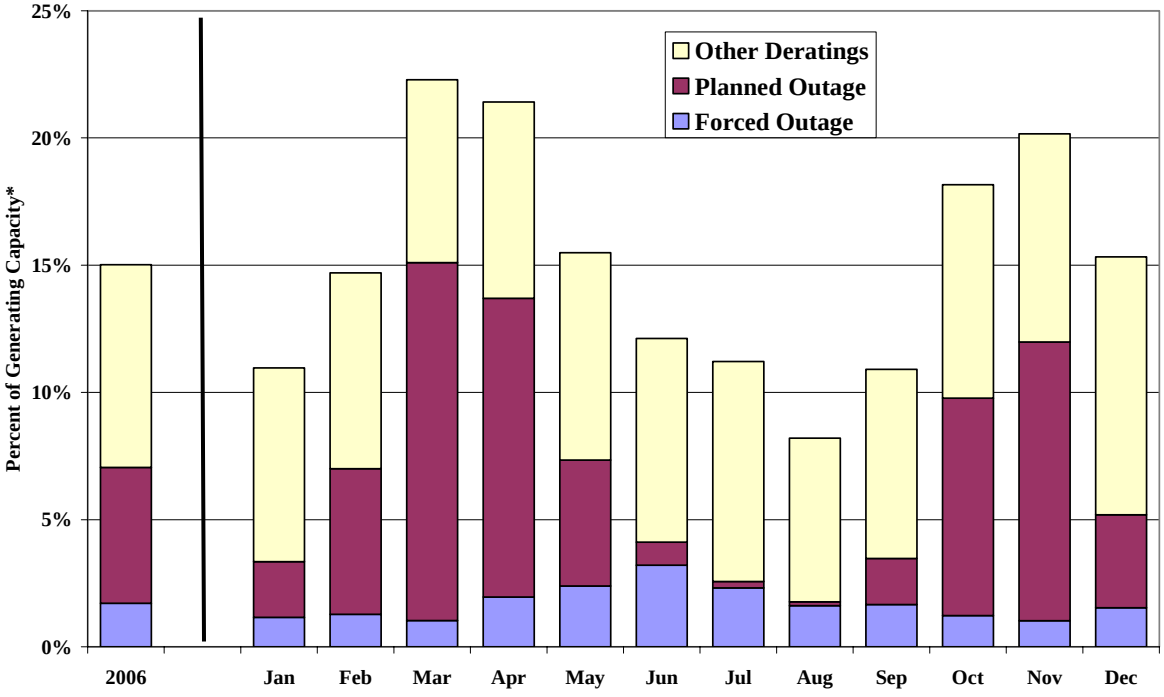
Figure 48 shows that installed capacity, including mothballed and switchable capacity, rose from 84 GW at the beginning of 2006 to 85 GW at the end of 2006. This increase is due to several

new generators coming on-line although it was diminished by several retirements. There were 1,028 MWs of new wind capacity coming on line from May through December 2006. The figure shows that the long-term outages and deratings fluctuated between 6 GW and 13 GW. The long-term outages and deratings also include over 8 GW of mothballed capacity. [25] These classes of capacity can be made available if market conditions become tighter as load rises.

As expected, short-term planned outages are relatively large in the spring and fall, decreasing to close to zero during the summer. Available in-service capacity fluctuated between 51 GW in March and 69 GW in August. The peak hour for the year required just over 63 GW to satisfy ERCOT's energy requirements plus approximately 4 GW for operating reserves and regulation-up requirements, resulting in surplus capacity of approximately 2 GW on that day.

The next analysis focuses specifically on the short-term forced outages and other short-term deratings. Figure 49 shows the average magnitude of the outages and deratings lasting less than 60 days for the year and for each month during 2006.

Figure 49: Short-Term Outages and Deratings*
2006



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

25 See "Report on the Capacity, Demand, and Reserves in the ERCOT Region," June 2006.

Figure 49 shows that total short-term deratings and outages were as large as 23 percent of installed capacity in the spring and fall, and dropped below 15 percent for the summer. Most of this fluctuation was due to anticipated planned outages, which ranged as high as 10 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 1 percent and 4 percent of total capacity on a monthly average basis during 2006. 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 49 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, and other factors described above. Furthermore, suppliers may delay maintenance on components such as boiler tubes, resulting in reduced capability. Because these deratings can fluctuate day to day or seasonally, some of the deratings are included in the "long-term outages and deratings" category while the others are included in this category. The other deratings were approximately 7 percent on average during the summer in 2006 and as high as 12 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.

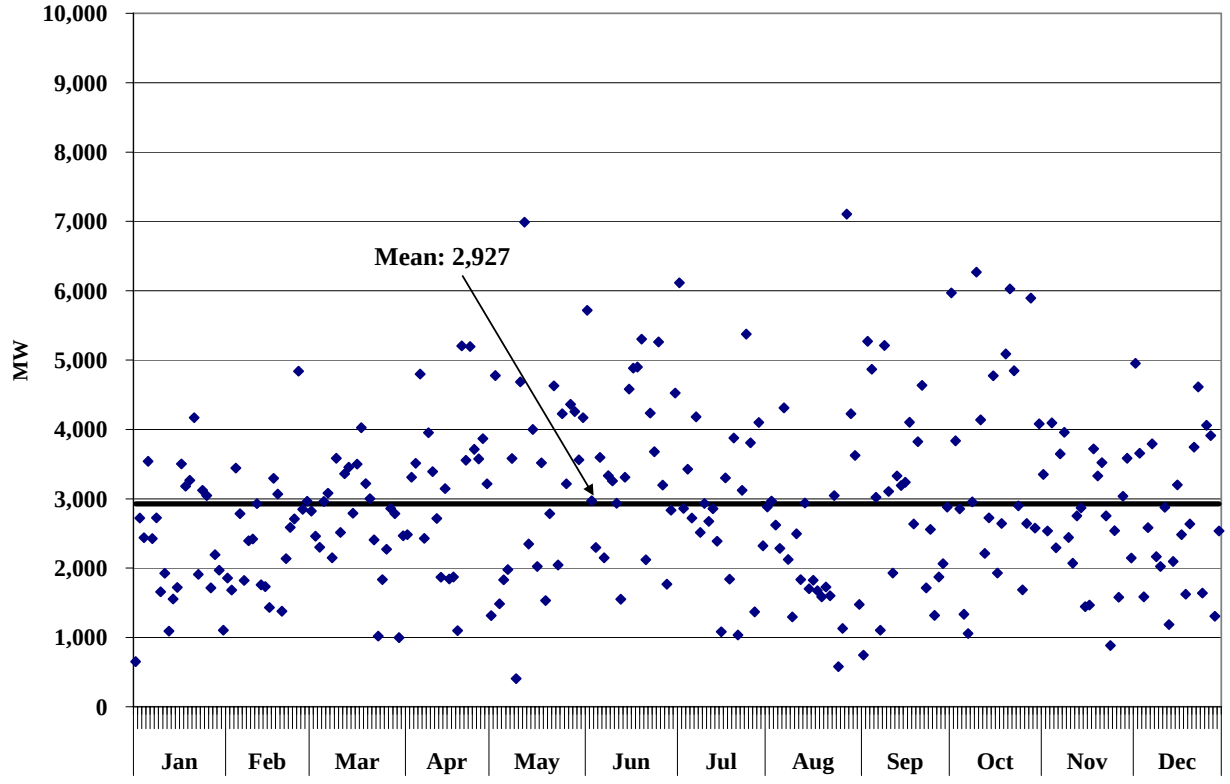
2. Daily Generator Commitments

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[26] 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 instead.

To evaluate the commitment of resources in ERCOT, Figure 50 plots the excess capacity in ERCOT during 2006.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 50: Excess On-Line and Quick Start Capacity During Daily Peaks on Weekdays -- 2006



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

Figure 50 shows that the excess on-line capacity during daily peak hours on weekdays averaged 2,927 MW in 2006, which is approximately 8 percent of the average load in ERCOT. This is a significant decrease from the average of 4,313 MW in 2005 and 6,627 MW in 2004. These decreases can be attributed in part to the continued increase in ERCOT load with a relatively static available supply, fewer quick-start gas turbines that were qualified to provide balancing energy, and a continuation of the trend from previous years of ERCOT committing fewer units via OOMC instructions and RMR.

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 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 planned for implementation by 2009 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").

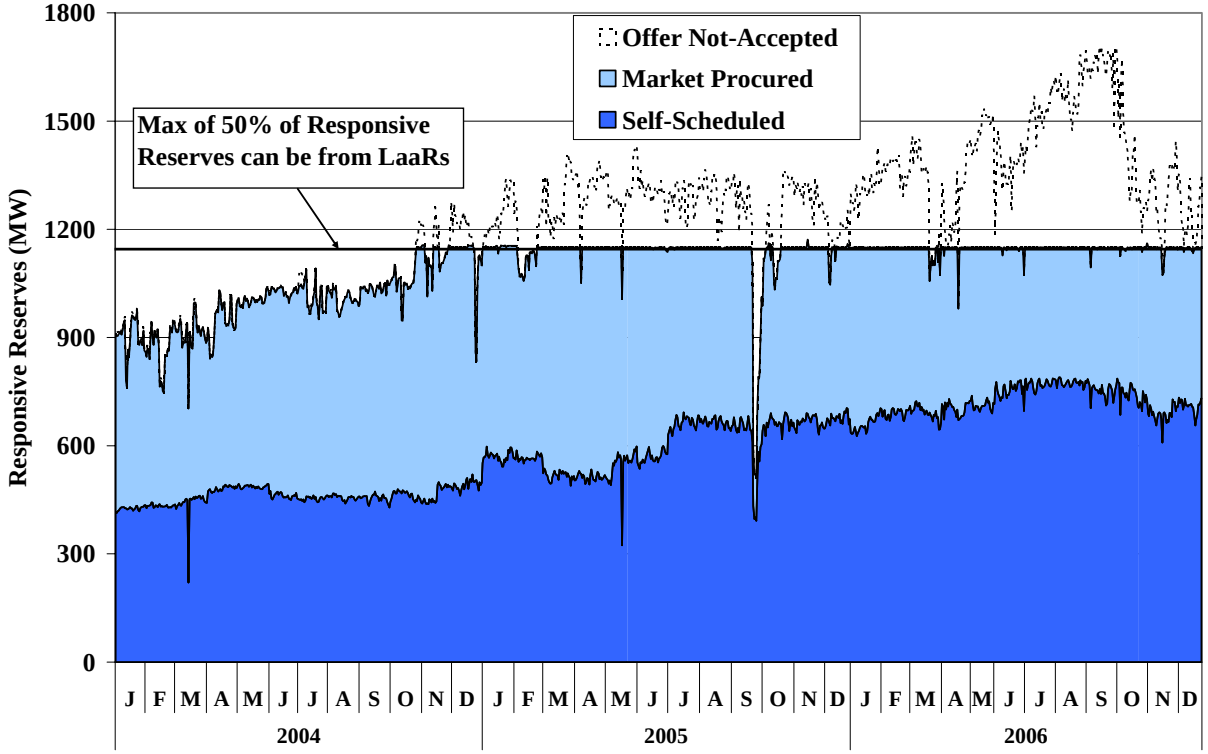
ERCOT allows qualified LaaRs to offer responsive reserves and non-spinning reserves into the day-ahead ancillary services markets. Qualified LaaRs can also offer blocks of energy in the

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 2006, 1,985 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. There were no BULs registered with ERCOT in 2006. Figure 51 shows the amount of responsive reserves provided from LaaRs on a daily basis in 2006.

Figure 51: Provision of Responsive Reserves by LaaRs
Daily Average – 2006



The high level of participation by demand response sets ERCOT apart from other operating electricity markets. Figure 51 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 1,147 MW in 2006. The majority of this increase was procured through self-provision and bilateral agreements rather than the ERCOT administered auction. Currently, 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.[27] The auction goes up the offer stack until it reaches the 2,300 MW required quantity of 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.

---

27 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 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.

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. In order 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 reserves pricing and incentives for suppliers, we recommend that ERCOT set separate prices for the two types of responsive reserves. 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 averaged only 14 MW in 2006. 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. Finally, prices in the balancing energy market have not been high enough to attract active load participation in that market. 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.

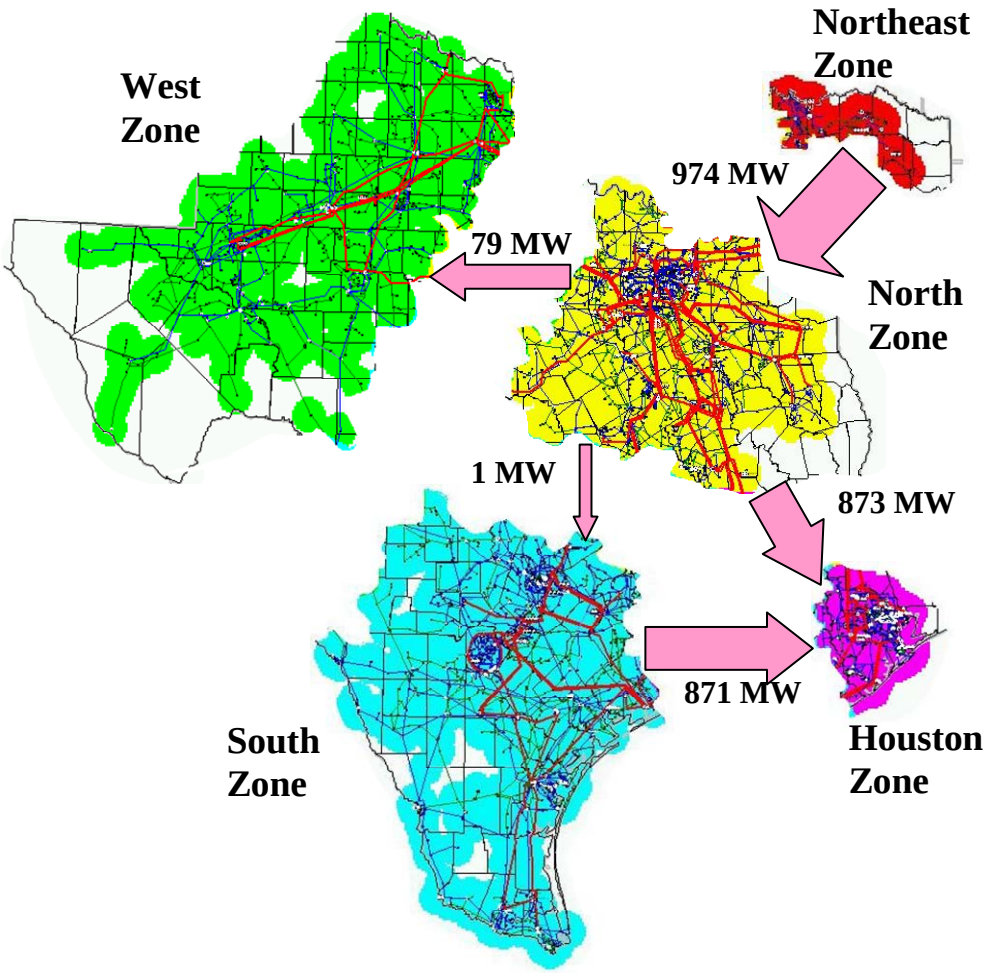
A. Electricity Flows between Zones

In 2006, there were five commercial pricing zones in ERCOT: (a) the North Zone, (b) the West Zone, (c) the South Zone, (d) the Houston Zone, and (e) the Northeast Zone, which was created in 2004 by dividing the North 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 2005 to 2006. ERCOT operators use the SPD software to dispatch balancing energy in each zone in order 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 five zone-based locations and six transmission interfaces. These six 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 2006.

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 52 shows the average SPD-modeled flows over CSCs between zones during 2006. A single arrow is shown for the modeled flows of both the North to West and West to North CSCs.

Figure 52: Average SPD-Modeled Flows on Commercially Significant Constraints During All Intervals in 2006



Note: In the figure above, CSC flows are averaged taking the direction into account. So one arrow shows the average flow for the for the North-to-West CSC was 79 MW, which is equivalent to saying that the average for the West-to-North CSC was *negative* 79 MW.

Figure 52 shows the five ERCOT geographic zones as well as the six 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 Northeast to North interface, (e) the North to Houston interface, and (f) the North to West interface. Based on SPD modeled flows, Houston is a significant importer while the Northeast Zone and the South Zone export significant amounts of power.

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. The most important simplifying assumption is that all generators in a zone have the same effect on the flows over the CSC, or the same generation shift factor ("GSF")[28] 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. Table 2 shows this analysis, which is based upon 2006 data but would not be significantly different for 2005. 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 – 2006

CSC 2006	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	-79	-80	*-1*	-162	*-82*
South-North	-1	32	*33*	26	*-7*
South-Houston	871	874	*3*	1211	*337*
North-Houston	873	845	*-28*	661	*-184*
NorthEast-North	974	970	*-4*	942	*-28*

[28] 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.

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 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, North to Houston, and Northeast to North CSCs, 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 337 MW and reduced the calculated flows on the North to Houston CSC by 184 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 2006, the six 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 2006.

Table 3: Actual Net Imports vs. SPD-Calculated Flows on CSCs 2003 to 2006

Year	Zone	Actual Net Imports	SPD Flows on CSCs
2003	**Houston**	1796	565
	North	-507	191
	South	-1213	-702
	West	-76	-54
2004	**Houston**	2479	1265
	North	867	264
	NorthEast	-2116	-858
	South	-1531	-800
	West	304	129
2005	**Houston**	2596	1247
	North	660	164
	NorthEast	-2138	-845
	South	-1501	-728
	West	386	162
2006	**Houston**	3434	1744
	North	462	20
	NorthEast	-2334	-974
	South	-1741	-870
	West	180	79

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 2006. 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.[29] The West Zone shifted from being a net exporter in 2003 to importing substantial quantities in 2004 and in 2005, with the average import levels dropping by about 50 percent in 2006 compared to 2005. From 2003 to 2006, net exports increased from the South Zone as well as the combined area of the North and Northeast zones. In every case, the SPD-calculated flows on CSCs were significantly less than the actual interchange.

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

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

were binding. Although this excludes most intervals, it is in these constrained intervals that the performance of the market is most critical.

Figure 53 shows the average SPD-calculated flows between the five 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 53: Average SPD-Modeled Flows on Commercially Significant Constraints During Transmission Constrained Intervals in 2006

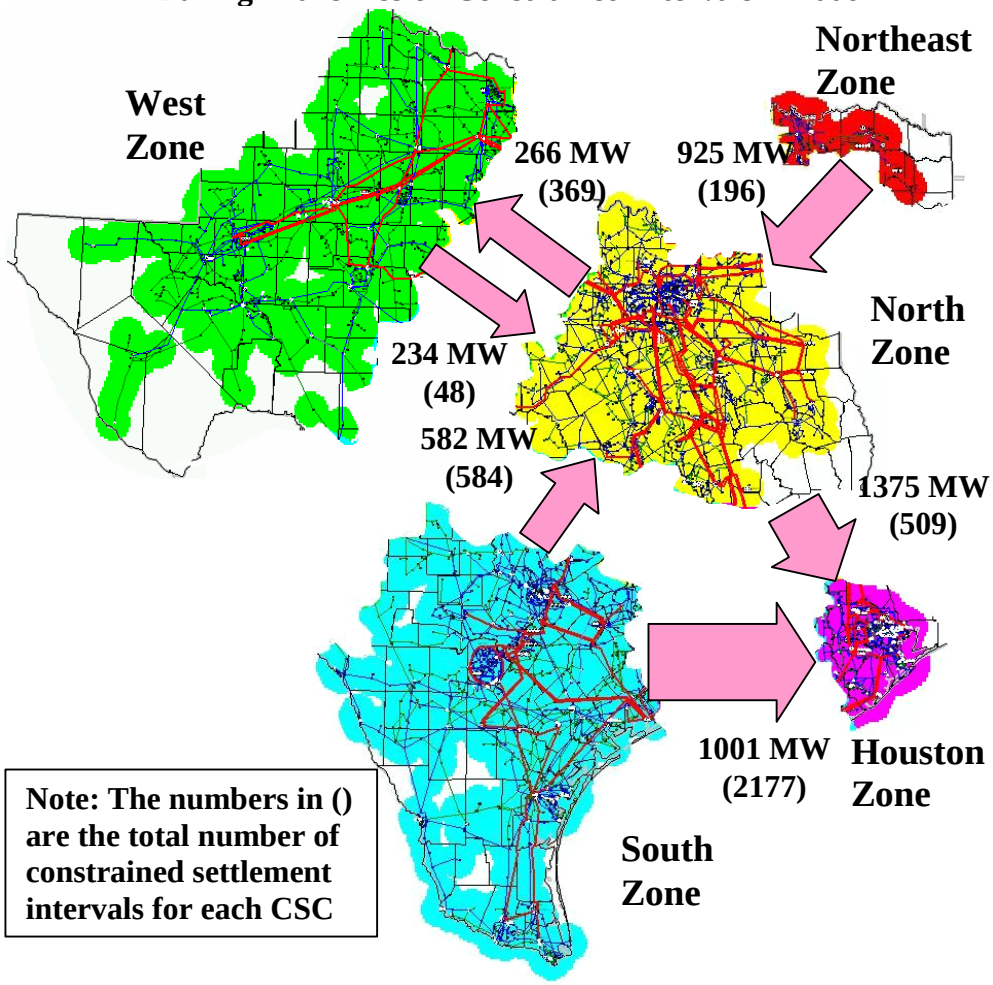


Figure 53 shows that inter-zonal congestion was most significant on the South to Houston CSC which exhibited SPD-calculated flows averaging 1,001 MW during 2,177 constrained intervals in 2006. Congestion was also significant on the South to North and North to Houston CSCs. The North to West CSC experienced much more congestion than the West to North CSC which

was congested for just 48 intervals during 2006. The Northeast to North CSC was constrained more frequently during 2006 than in 2005, although the majority of this congestion was related to transmission construction and the Northeast to North CSC was eliminated in 2007.

1. Congestion Rights in 2006

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 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 2006. Figure 54 shows the average number of TCRs and PCRs that were allocated for each of the CSCs in 2006, as well as the average SPD-modeled flows during the constrained intervals.

Figure 54: Transmission Rights vs. Real-Time SPD-Calculated Flows Constrained Intervals – 2006

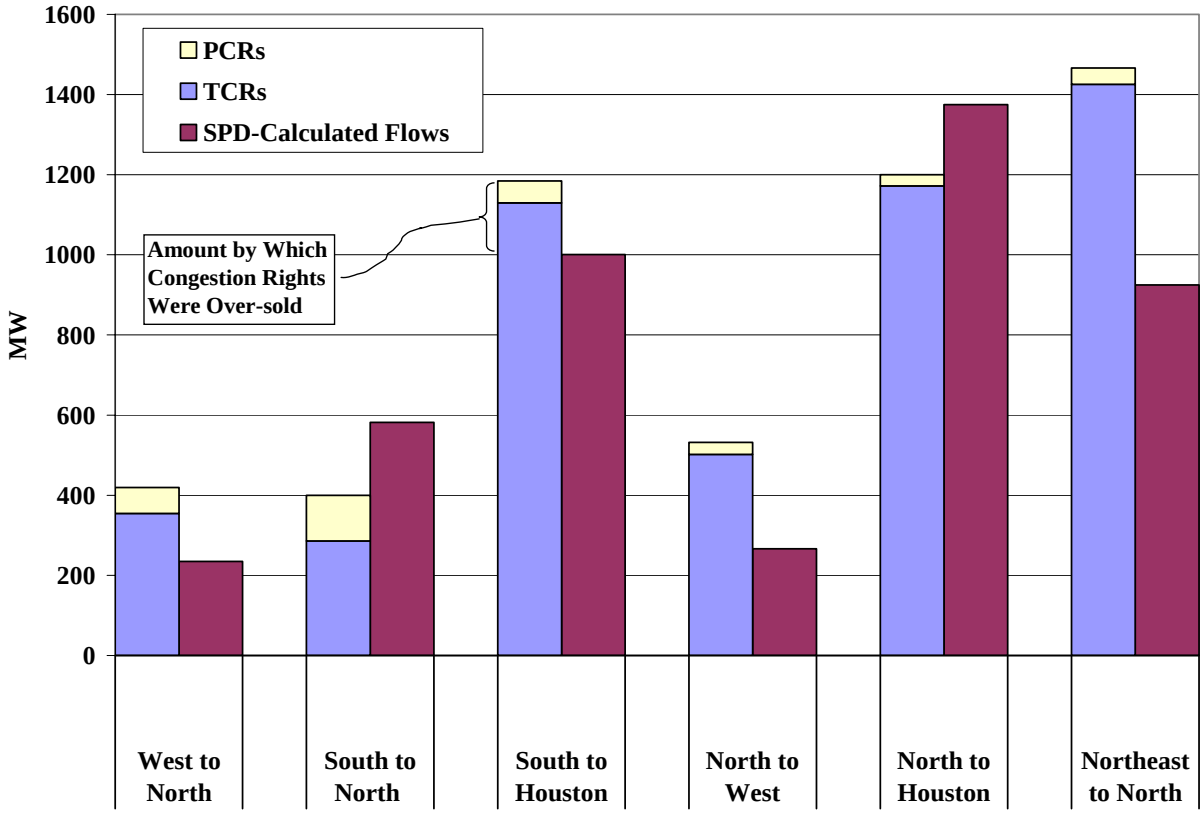


Figure 54 shows that total congestion rights (the sum of PCRs and TCRs) on the West to North, North to West, South to Houston, Northeast to North interfaces exceeded the average real-time SPD-calculated flows during constrained intervals while the congestion rights on the South to North and North to Houston CSCs were less than SPD calculated flows. 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 South to Houston CSC were oversold by an average of 184 MW.

The largest divergence between the SPD-calculated limits and the limits implied by the congestion rights was on the Northeast to North CSC where 1,466 MW of congestion rights were allocated, but the average SPD-calculated flow during constrained intervals was 925 MW. Hence, the congestion rights that determine ERCOT's total obligation to make congestion payments exceeded the modeled flow over the CSC by an average of 541 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 54, we compare the SPD-calculated flows and congestion rights quantities for each of the constrained intervals by CSC.

2. Congestion on South to North CSC

Figure 55 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 2006. Because only congested intervals are shown, some months will have significantly more observations than other months. Although some congestion occurred in every month, the three months from June to August accounted for 71 percent of all constrained intervals during 2006.

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 55 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 55: Congestion Rights Allocated vs. SPD Flows during Constrained Intervals South to North – 2006

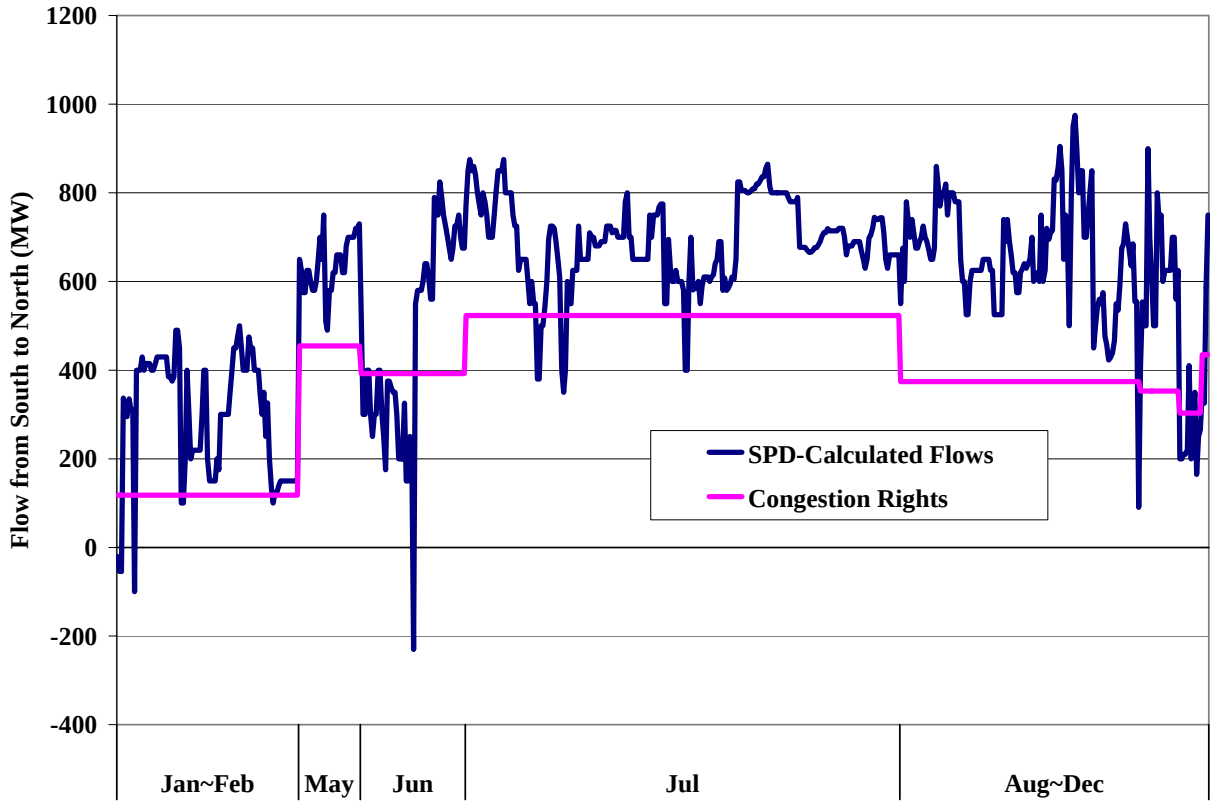


Figure 55 indicates that the quantity of outstanding congestion rights fluctuated considerably during 2006. From January to February, fewer than 200 MW of rights were allocated for the South to North CSC, whereas for May and July, more than 400 MW of congestion rights were allocated for the South to North CSC in 2006. 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.

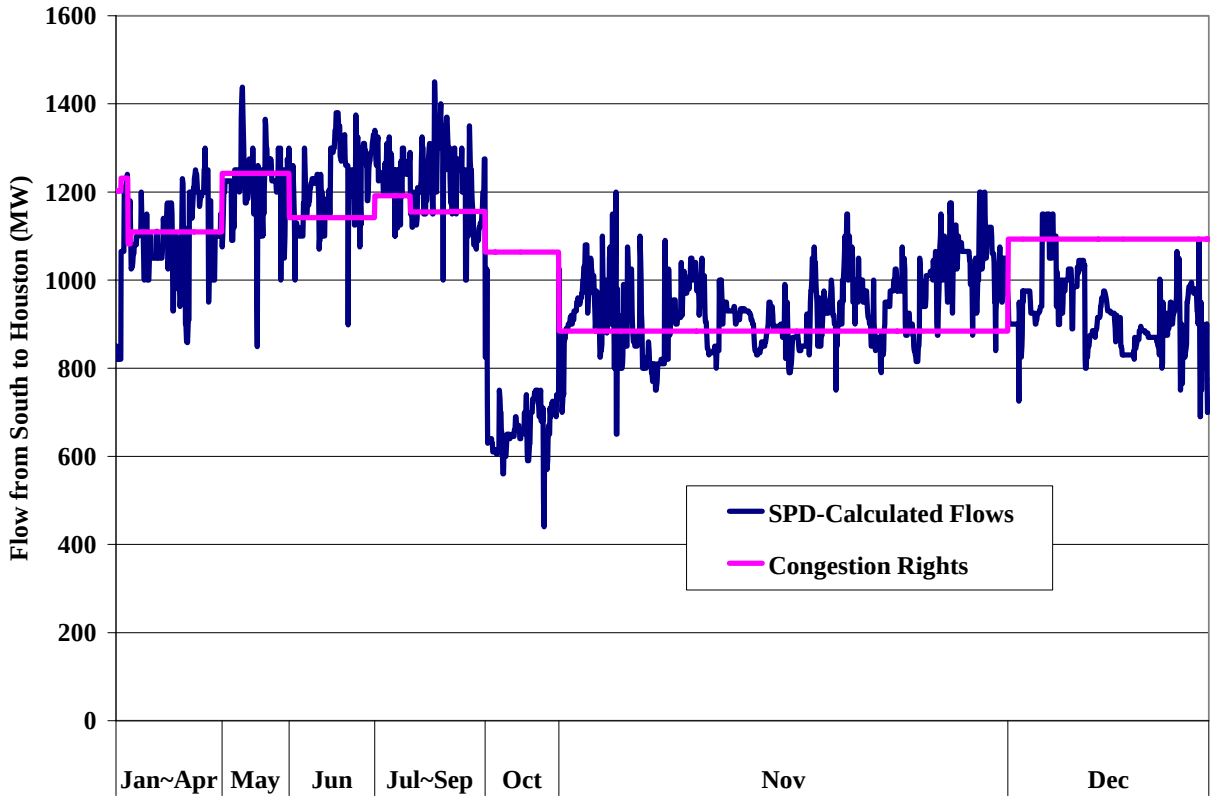
For the South to North CSC, the congestion rights were nearly always below SPD flows for the congested intervals in 2006. The figure shows five constrained intervals when the SPD-calculated flows were *negative* at times during January, February, and June.

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.

3. Congestion on South to Houston CSC

With 2,177 constrained intervals, this interface experienced the most frequent congestion of any CSCs during 2006. The most congestion occurred in November and December. In the months with significant congestion, SPD flows averaged between 940 and 924 MW. However, there was significant variation in the number of congestion rights allocated for this CSC by month, with as little as 884 MW in November and 1,241 MW in May. Figure 56 shows the comparison between actual flow and the congestion rights quantities.

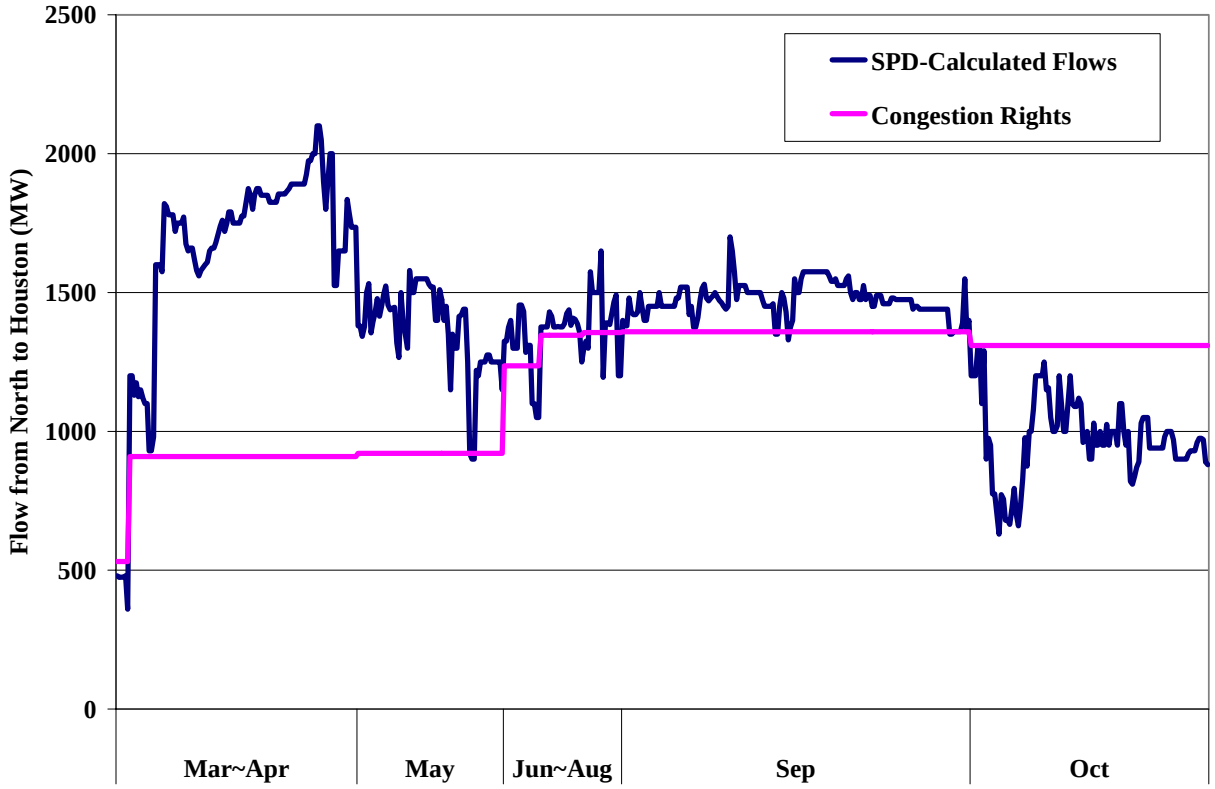
Figure 56: Congestion Rights Allocated vs. SPD Flows during Constrained Intervals South to Houston – 2006



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. Prior to May 2006, ERCOT generally allocated between 530 and 920 MW of congestion rights for this CSC. After May 2006, however, the number of congestion rights was increased to 1,235 in June and further to above 1,300 MW in subsequent months. From March to May, the rights were significantly under-sold while in October, the rights were significantly over-sold. Frequency of transmission constraints rose dramatically in September and October in conjunction with the increase of rights allocated.

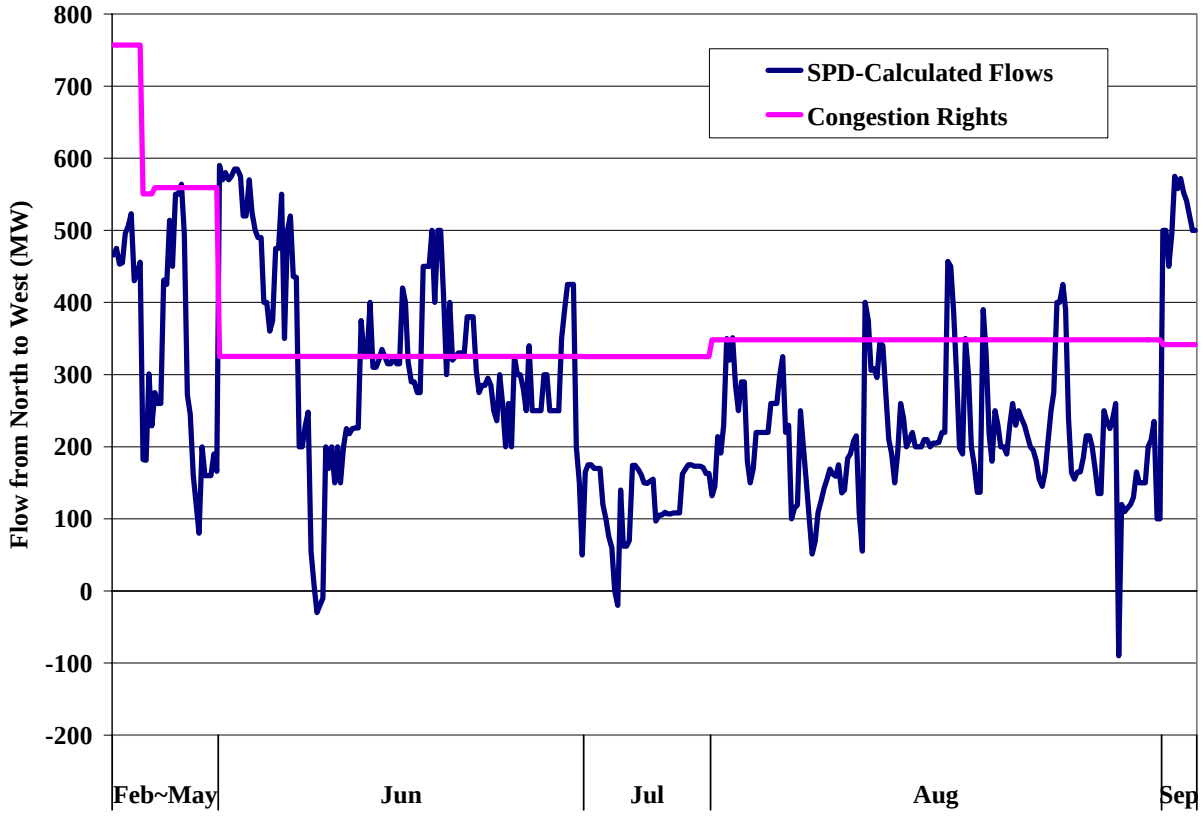
Figure 57: Congestion Rights Allocated vs. SPD Flows during Constrained Intervals North to Houston – 2006



5. Congestion on North to West CSC

This CSC was congested primarily during the summer months with approximately 87 percent of constrained intervals in June, July and August. Although the number of congestion rights allocated for this interface varied from 325 to 757 MW over the year, the SPD flows averaged just 266 MW during constrained intervals.

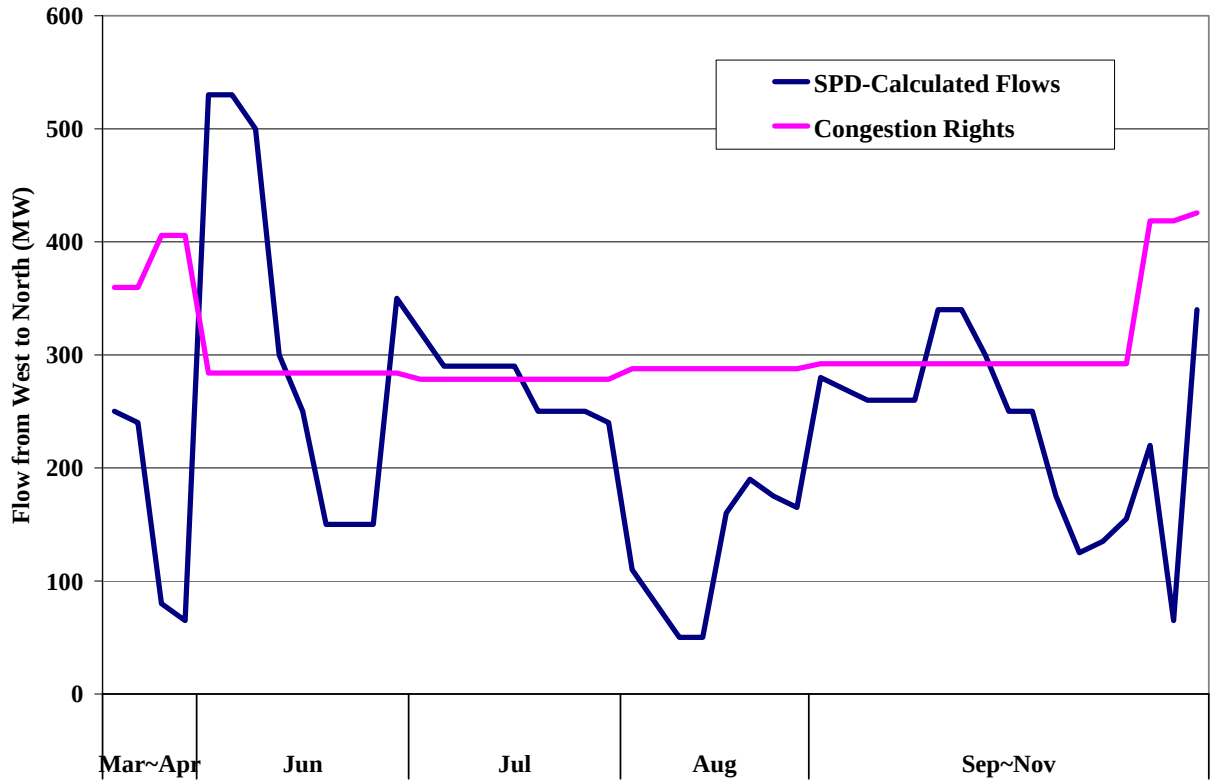
Figure 58: Congestion Rights Allocated vs. SPD Flows during Constrained Intervals North to West – 2006



6. Congestion on West to North CSC

This CSC was the most infrequently congested CSC. During the year of 2006, the West to North CSC congested for only 48 intervals. This CSC was congested primarily during the summer months, June to September. Although the number of congestion rights allocated for this interface varied from 278 to 425 MW over the year, the SPD flows averaged just 234 MW during constrained intervals. As can be seen in Figure 59, in most of the months, the congestion rights were over sold.

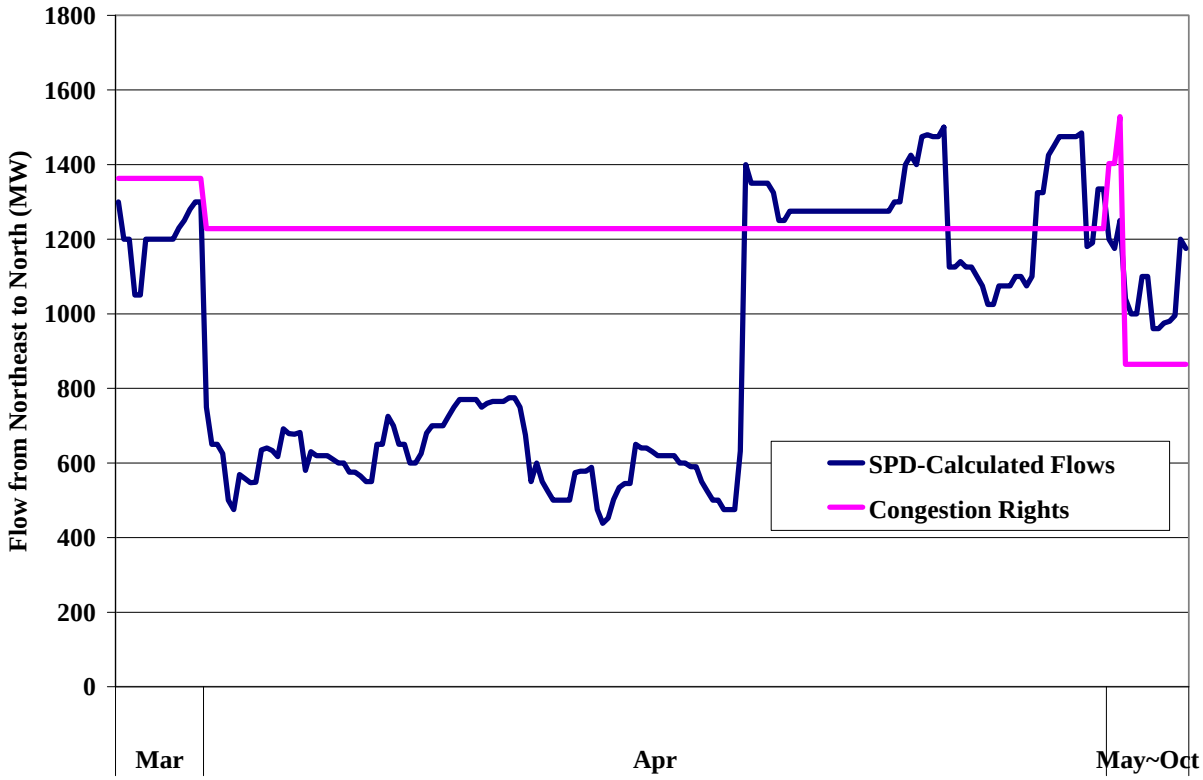
Figure 59: Congestion Rights Allocated vs. SPD Flows during Constrained Intervals West to North – 2006



7. Congestion on Northeast to North CSC

The Northeast to North CSC was created in 2004. However, during the entire year of 2005, it was never congested. In 2006, the Northeast to North CSC congested 196 times with the actual flow between 438 to 1500 MW. During the months of March and April, the congestion rights exceed actual flow in most of the congested periods. Figure 65 shows the monthly comparison between actual flow and the number of congestion rights sold for the Northeast to North CSC. Most of the congestion on the Northeast to North CSC was associated with transmission construction related to increased transfer capability, and this CSC was eliminated in 2007.

Figure 60: Congestion Rights Allocated vs. SPD Flows during Constrained Intervals Northeast to North – 2006



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.[30] Since the monthly studies tend to more accurately reflect conditions that will prevail in

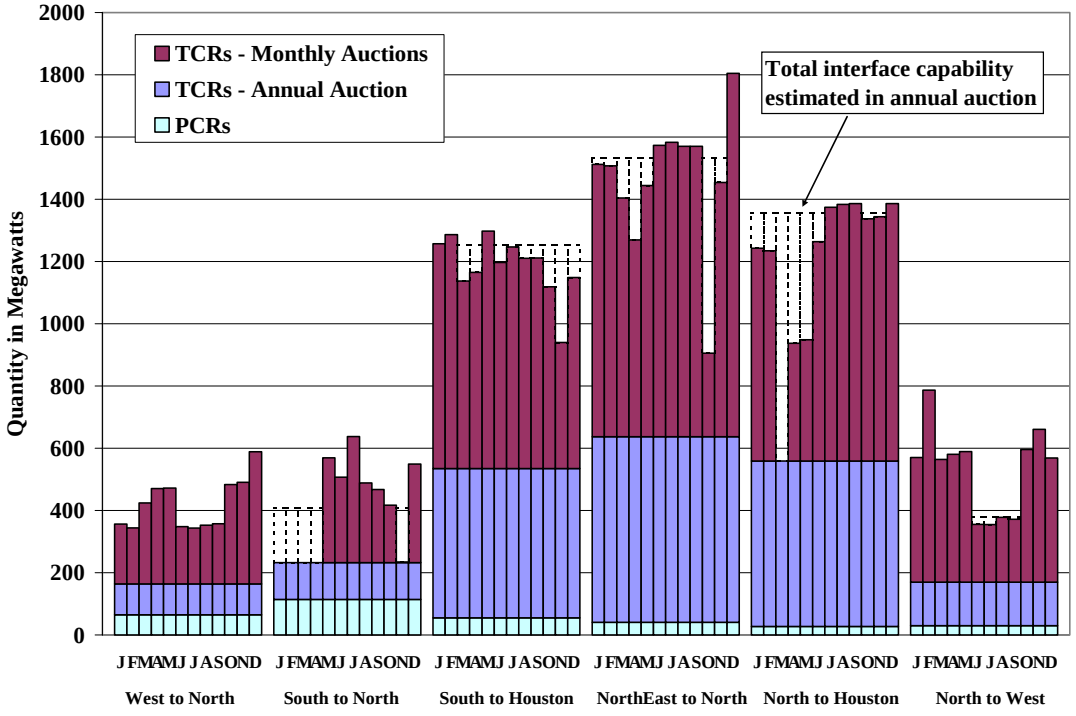
[30] Prior to 2005, 60 percent of estimated capability (after accounting for Pre-assigned Congestion Rights which are assigned to NOIEs) was sold in the annual auction. The remaining 40 percent was sold in the monthly auctions. This was changed because there were instances when the capability estimated 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.

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 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. 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 in order 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 61 shows the quantity of each category of congestion rights for each month during 2006. The quantities of PCRs and annual TCRs are constant across months and were determined before the beginning of 2006, while monthly TCR quantities can be adjusted monthly.

Figure 61: Quantity of Congestion Rights Sold by Type
2006



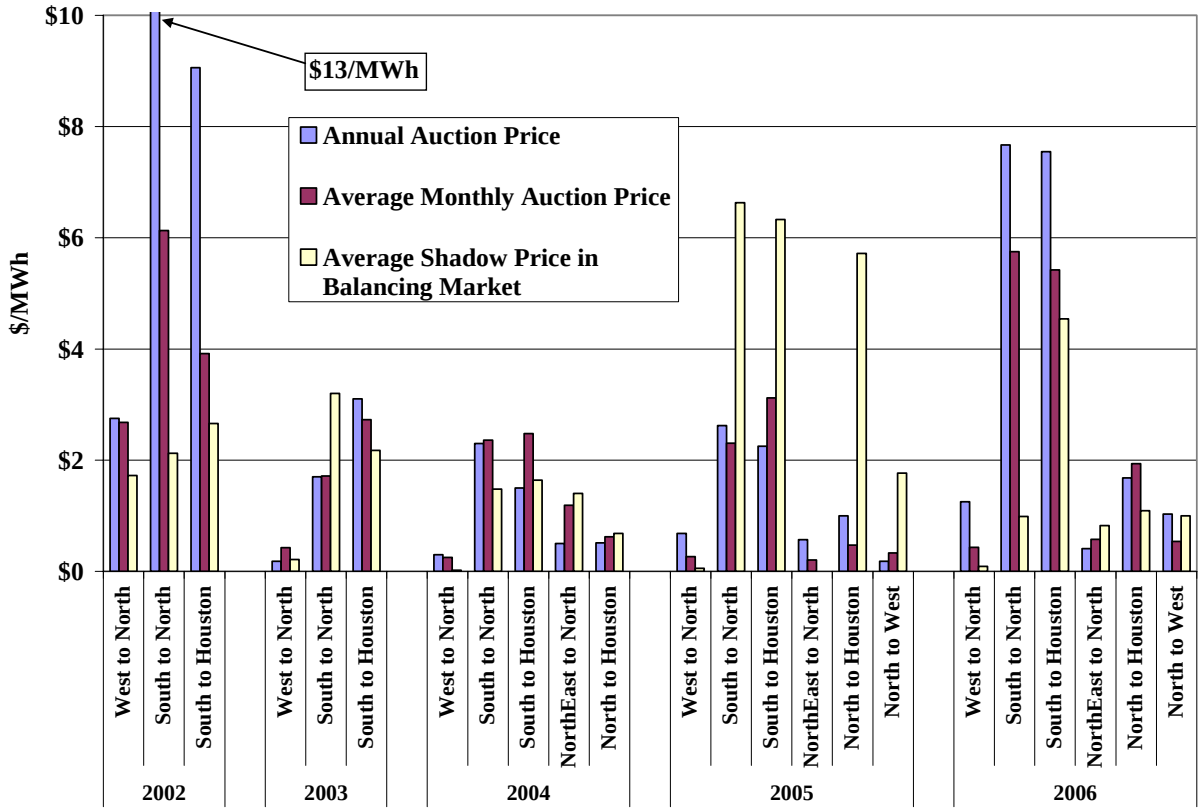
When the monthly planning studies indicate changes from the summer study, revisions are often made to the estimated transmission capability. Therefore, the auctioned congestion rights may increase or decrease relative to the amount estimated in the summer study. The shadow boxes in the figure represent the capability estimated in the summer study that is not ultimately sold in the monthly auction. When there is no shadow box in Figure 61, the total quantity of PCRs and TCRs sold in the annual and monthly auctions equaled or exceeded the summer estimate and therefore no excess capability is shown.

The South to North and North to Houston interfaces experienced the largest fluctuations in the estimates of transmission capacity from the annual auction to the monthly auction. In fact, South to North TCRs were not even auctioned during four of the monthly auctions. There were also several instances when no congestion rights were available to be sold for the North to Houston CSC in 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 62 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 62 indicates that in 2002, the annual auction for the TCRs resulted in prices that substantially over-valued the congestion rights, particularly on the South to North and South to Houston interfaces. Monthly TCR prices for these interfaces were roughly one-half of the prices from the annual auctions, but were still significantly higher than the ultimate congestion payments to the TCR holders. In the West to North interface, the annual and monthly TCR auction prices were close in magnitude and were both much closer to the true value of the congestion rights.

Figure 62: TCR Auction Prices versus Balancing Market Congestion Prices 2002 to 2006



In 2003, the TCR prices for all of the interfaces decreased considerably, causing the prices to converge more closely with the actual value of the congestion rights. It is noteworthy that the TCRs for the South to North and South to Houston interfaces settled at prices in 2004 that were closer to the previous year's value than in 2003. This indicates that participants improved in their ability to forecast interzonal congestion and to value the TCRs, in part by observing historical outcomes. This improvement was likely facilitated by the simplified zonal representation of the ERCOT network embedded in the balancing energy market.

In 2004, TCR auction prices for the West to North, South to North, and South to Houston interfaces were similar to the previous year. Since congestion tends to be consistent across time, the auction prices for 2004 were reasonable predictors of real-time congestion. In 2004, there were two new products in the TCR auctions for the new Northeast to North and North to Houston CSCs. In both cases, the annual TCR price was below the monthly average TCR price, which was slightly below the average value of congestion, but the divergence between auction prices and actual congestion values was not as significant as in 2002. This reflects cautiousness

on the part of market participants when purchasing a TCR for a CSC that did not exist before 2004.

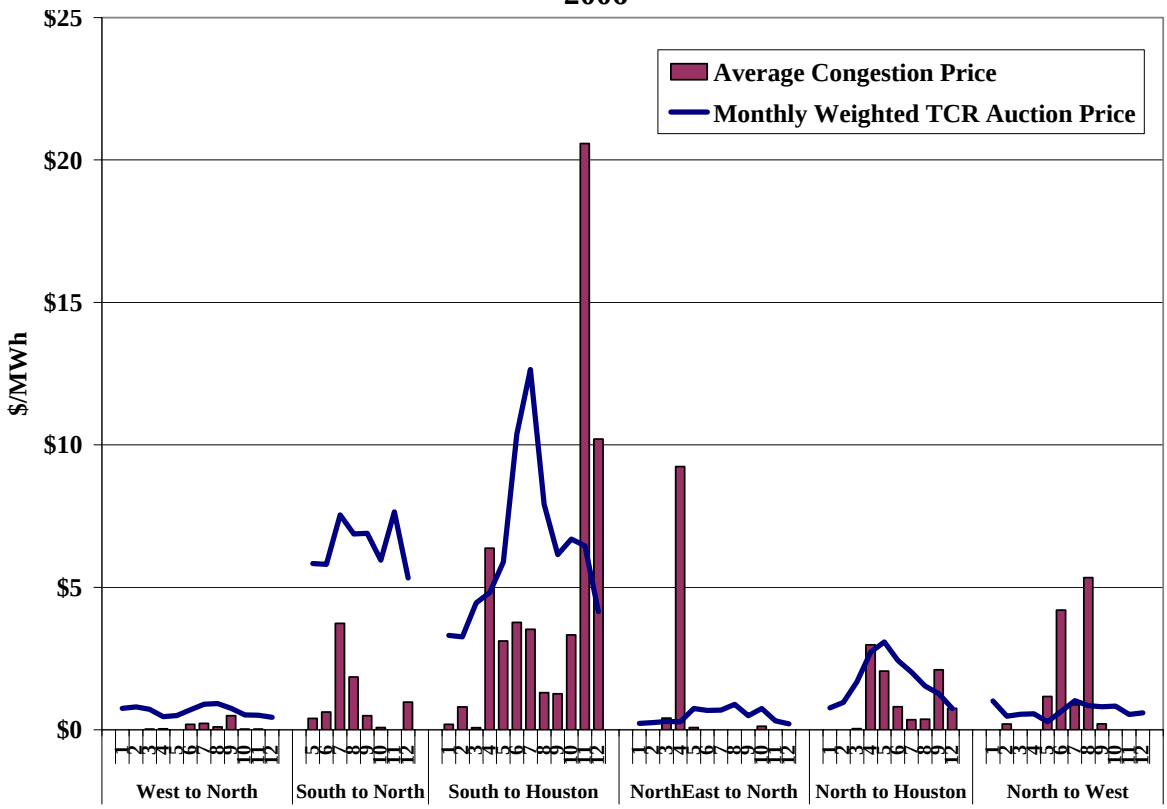
In 2005, market participants substantially under-estimated the value of congestion on the CSCs. The annual and monthly TCR prices in 2005 were generally in line with the TCR prices and the levels of balancing market shadow prices that prevailed in 2004. However, the actual volume and prices of congestion were substantially greater than in 2004, particularly on the South to Houston, South to North, and North to Houston CSCs. The North to West CSC was also substantially under-valued in the TCR auctions, although this is understandable given the lack of experience that market participants have with a newly created CSC.

In contrast to 2005, market participants over-estimated the annual value of congestion on the South to North, South to Houston, and North to Houston CSCs in 2006. The annual auction price for the North to West CSC converged well with the congestion value. The West to North CSC congestion value was again over-estimated as was the case from 2002 to 2005. Although South to North TCR auction concluded with the highest auction price among all the other CSCs, the South to North CSC actual congestion value decreased significantly from 2005.

Figure 63 compares monthly TCR auction prices with monthly average real-time CSC shadow prices from SPD for 2006. The TCR auction prices are expressed in dollars per MWh. In months when the monthly auction did not occur (*i.e.*, when the annual auction designated sufficient congestion rights for that month) no data is presented. This explains the missing months for the South-North CSC and the North-Houston CSC.[31]

[31] Notice that these missing months correspond to the missing monthly auction values in Figure 61.

Figure 63: Monthly TCR Auction Price and Average Congestion Value 2006



The TCR price trends for South to North and North to Houston CSCs, correlated well with the actual congestion prices, although the TCR prices for the South to North CSC far exceeded the congestion prices. Overall, market participants did a poor job predicting fluctuations in congestion during 2006, particularly on the South to Houston and Northeast to North interfaces. For both of these interfaces, there were several months when balancing market congestion spiked, far exceeding the TCR prices in those months. However, based on the TCR prices, there is little sign that market participants expected an increase in congestion in those months relative to other months.

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.[32] 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 64 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.

[32] 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 64: TCR Auction Revenues, Credit Payments, and Congestion Rent[33] 2002 to 2006

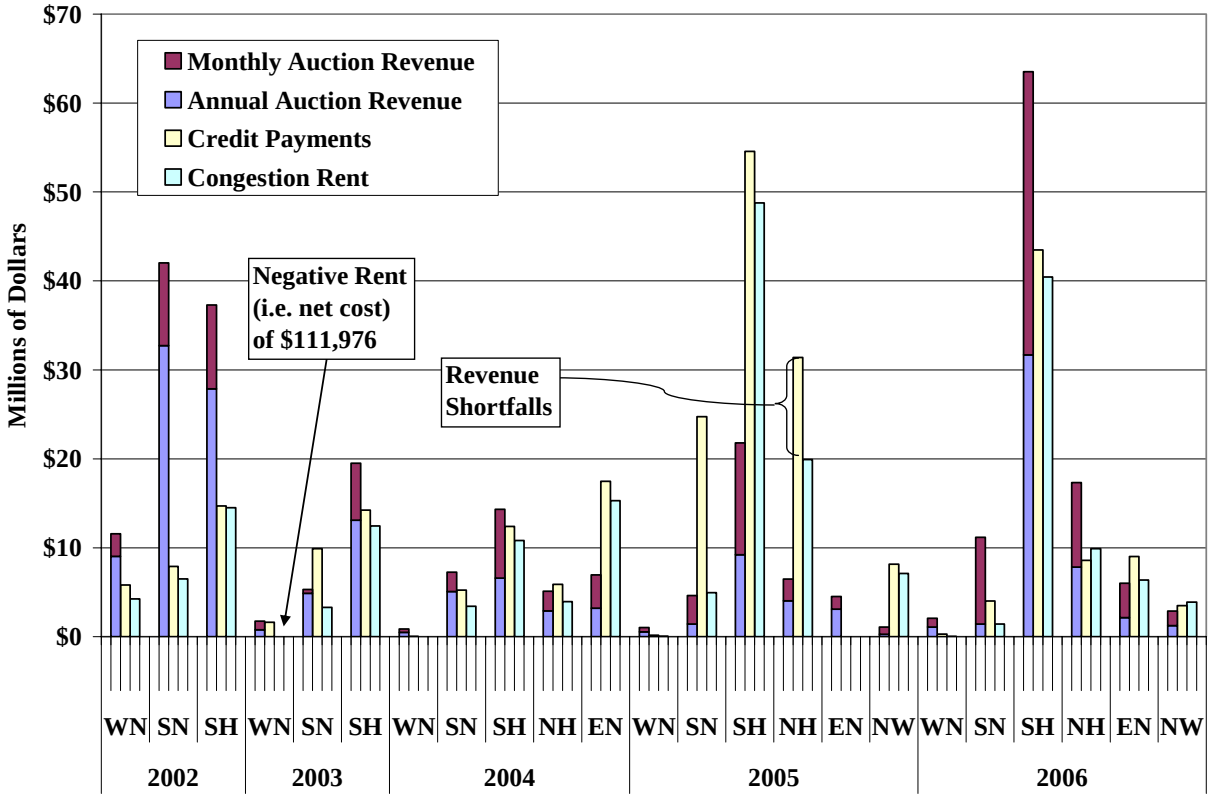


Figure 64 shows that in 2002, the total auction revenues were far greater than credit payments to TCR holders. This is the result of the auction prices being much greater than the average shadow prices that occurred in the balancing energy market (as was shown in Figure 63 above). The figure also shows that from 2002 to 2003, there was a significant reduction in auction revenues (a reduction of 71 percent). Auction revenues were reduced in 2003 because both annual and monthly auction prices decreased significantly due to improvements in the ability of market participants to forecast congestion on CSCs.

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 prior years. 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

33 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.

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 balancing market 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 63, 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.

Figure 64 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 2002 when congestion rents covered 93 percent of payments to TCR holders and in 2004 when they covered 81 percent. 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. 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.[34] These shortfalls are included in the Balancing Energy Neutrality Adjustment charge and assessed to load ERCOT-wide. Collecting substantial portions of the congestion costs for the market through such uplift charges reduces the transparency and efficiency of the market. It also increases the risks of transacting and serving load in ERCOT because uplift costs cannot be hedged.

D. Local Congestion and Local 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.

---

[34] 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 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.

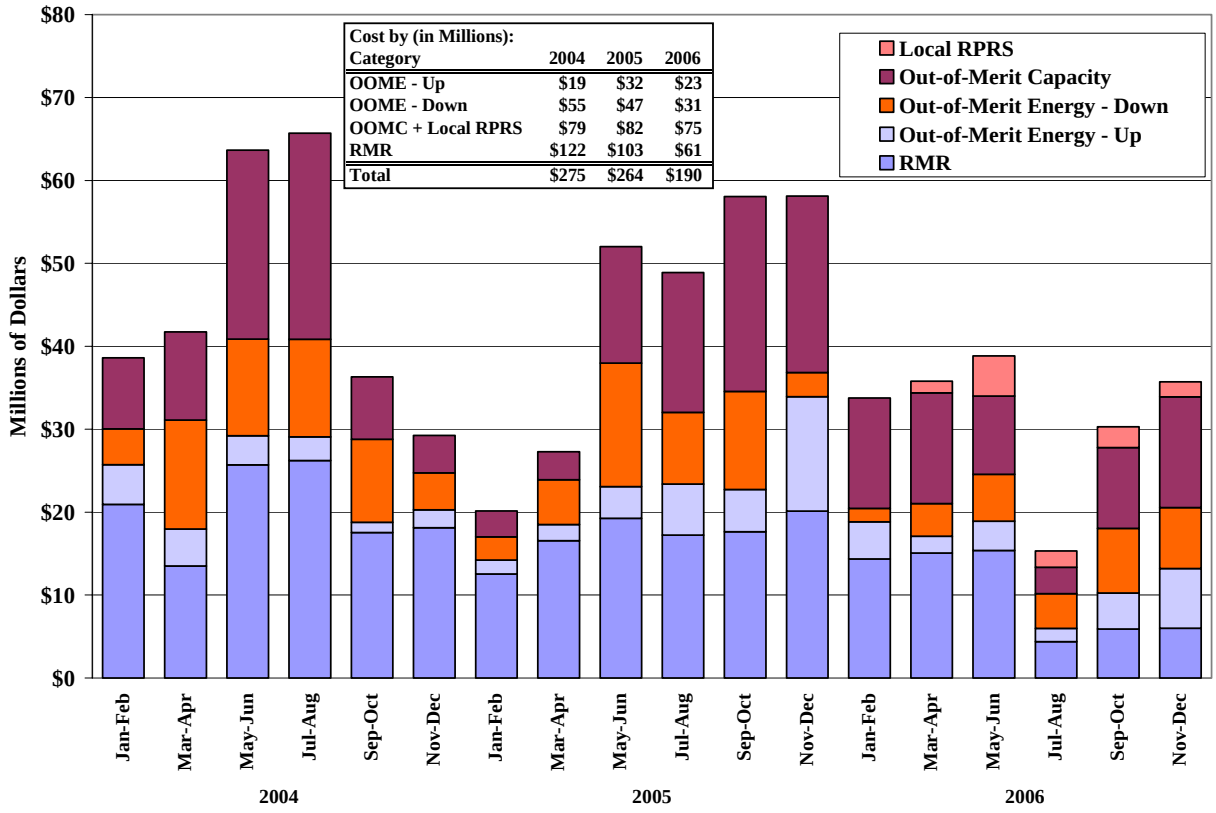
When a unit is dispatched out of merit (OOME up or OOME down), the unit is paid for a quantity equal to the difference between the scheduled output based on the unit's resource plan and the actual output resulting from the OOME instruction from ERCOT. The payment per MWh for OOME is a pre-determined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

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

A unit providing capacity under an OOMC 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. Units contracted to provide RMR service to ERCOT are compensated for start-up costs, energy costs, and are also paid a standby fee. Figure 65 shows each of the four categories of uplift costs from 2004 to 2006.

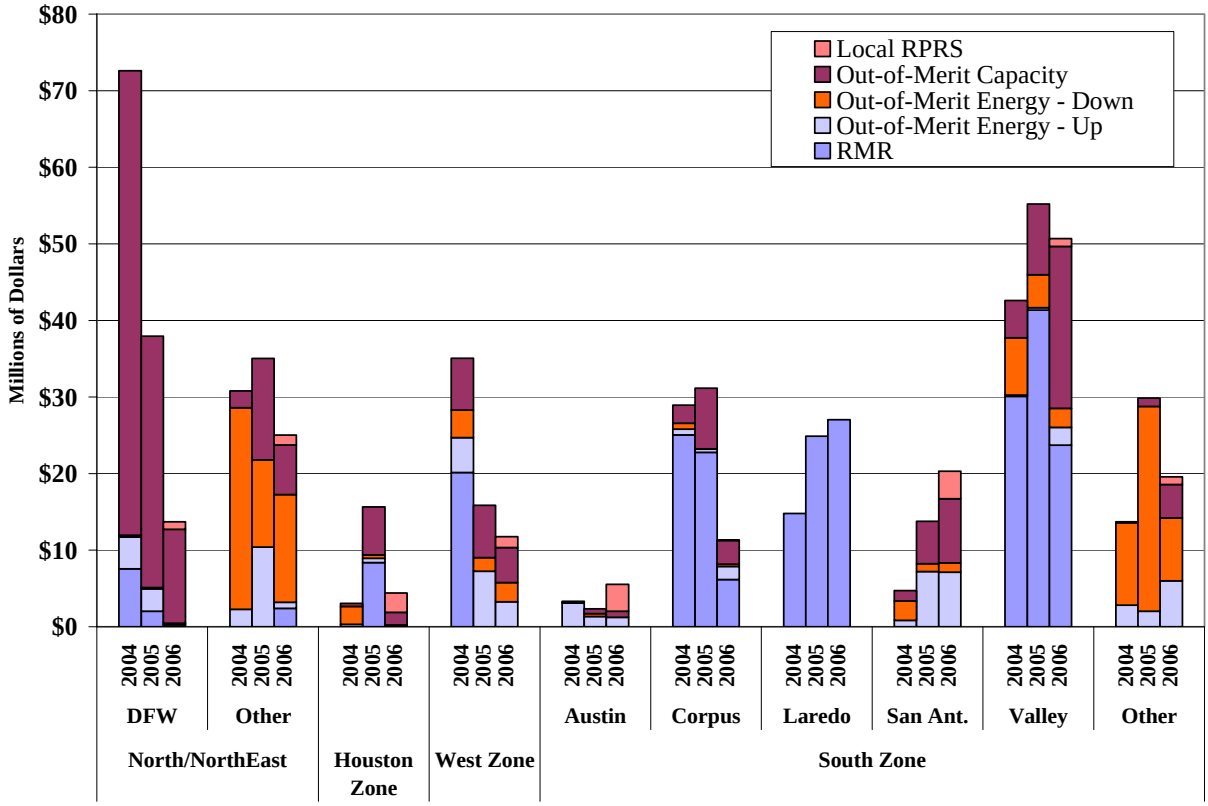
Figure 65: Expenses for Out-of-Merit Capacity and Energy 2004 to 2006



The results in Figure 65 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 $190 million, a reduction of 28 percent. As previously noted, there were substantial reductions to RMR cost due to the expiration of RMR agreements in 2006, which accounts for $42 million of the $74 million decrease from 2005 to 2006. Total OOME Up and OOME Down costs also decreased from $79 million in 2005 to $54 million in 2006, a reduction of 32 percent. This reduction is likely due to the continued improvements to the ERCOT transmission system resulting in less frequent local congestion, and the introduction of an enhanced replacement reserve procurement process by ERCOT in 2006.

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 2006. Figure 66 shows these payments by location.

Figure 66: Expenses for OOME, OOMC and RMR by Region 2004 – 2006



Uplift costs decreased dramatically from 2004 to 2006 in the Dallas/Ft. Worth ("DFW") area, in the West zone and in the South zone Corpus Christi 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 $6 million in 2006. RMR costs in the Laredo area increased from 2004 to 2006 due to increased fuel costs, as the number of RMR units in that area remained constant during this time period. The most significant increases in uplift costs associated with local reliability actions from 2004 to 2006 was in the San Antonio area, increasing fourfold from around $5 million in 2004 to approximately $20 million in 2006.

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 2006. 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 2006 than in 2005. 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 was improved in 2006 relative to 2005.

A. Structural Market Power Indicators

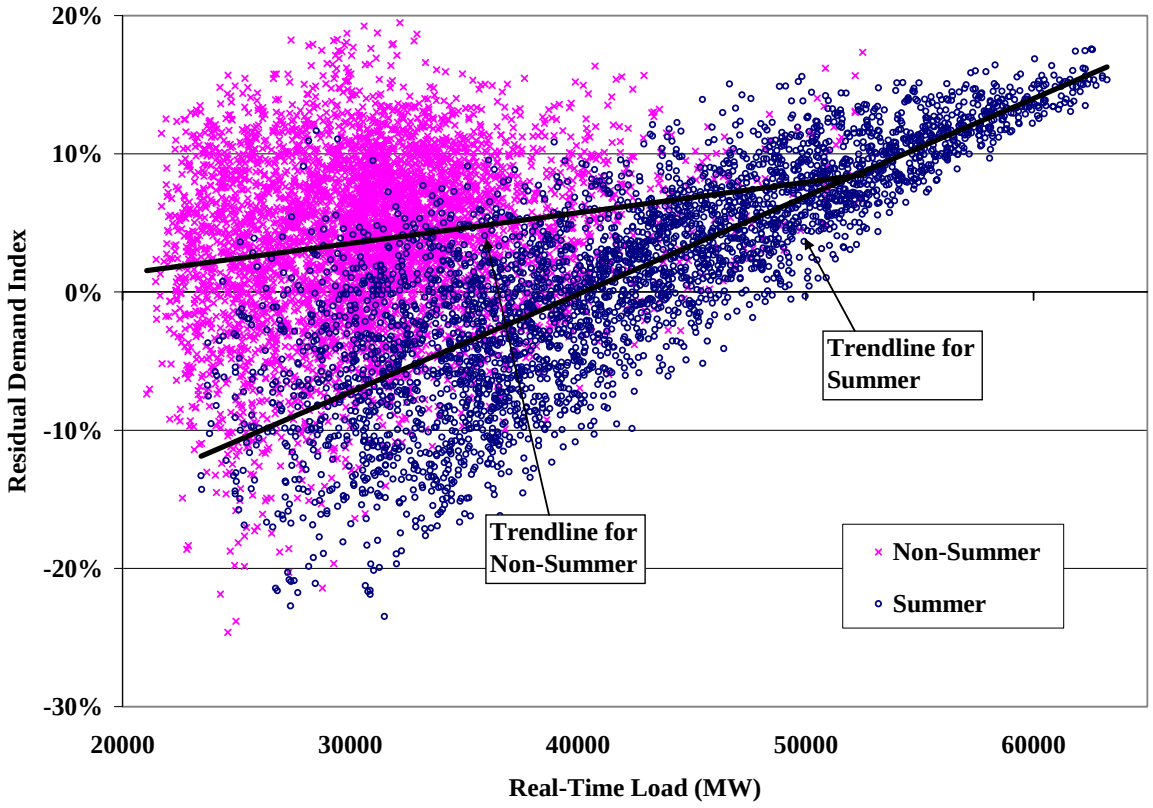
We analyze market structure using the Residual Demand Index ("RDI"), a statistic that measures the percentage of load that could not be satisfied without the resources of the largest supplier. When the RDI is greater than zero, the largest supplier is pivotal (*i.e.*, its resources are needed to satisfy the market demand). When the RDI is less than zero, no single supplier's resources are required in order 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, indicating 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 67 shows the RDI relative to load on an hourly basis in 2006. 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 67: Residual Demand Index
2006



The figure shows that the RDI for the summer (i.e. May to September) was usually positive in hours when load exceeded 40,000 MW. During the summer, the RDI was greater than zero in approximately 58 percent of hours. During the non-summer period, the RDI was generally positive under all load conditions. 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 75 percent of hours during that period. In addition to being higher on average, the non-summer trend line exhibits a flatter slope than the trend line for the summer period. The flatter slope of the non-summer trend line indicates a weaker relationship between the RDI and demand level in the non-summer months. 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 68 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[35] owned by other suppliers. Figure 68 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 69 shows the same data as in Figure 68 except that the balancing energy offers are limited by portfolio ramp constraints in each interval.

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

Figure 68: Balancing Energy Market RDI vs. Actual Load
2006

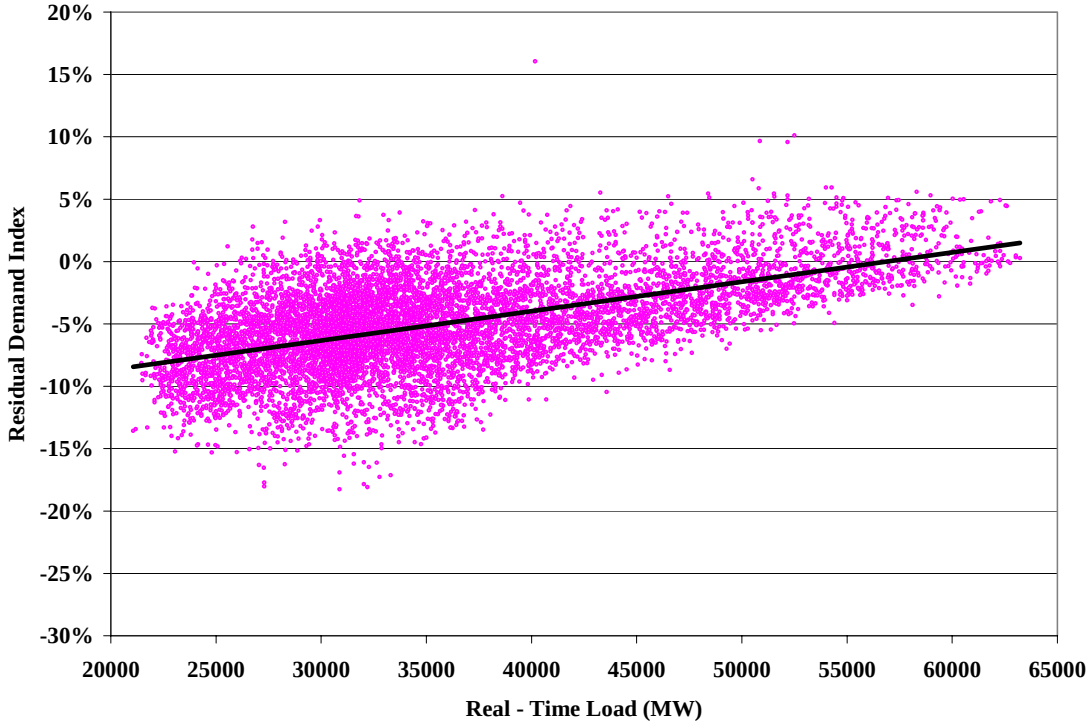
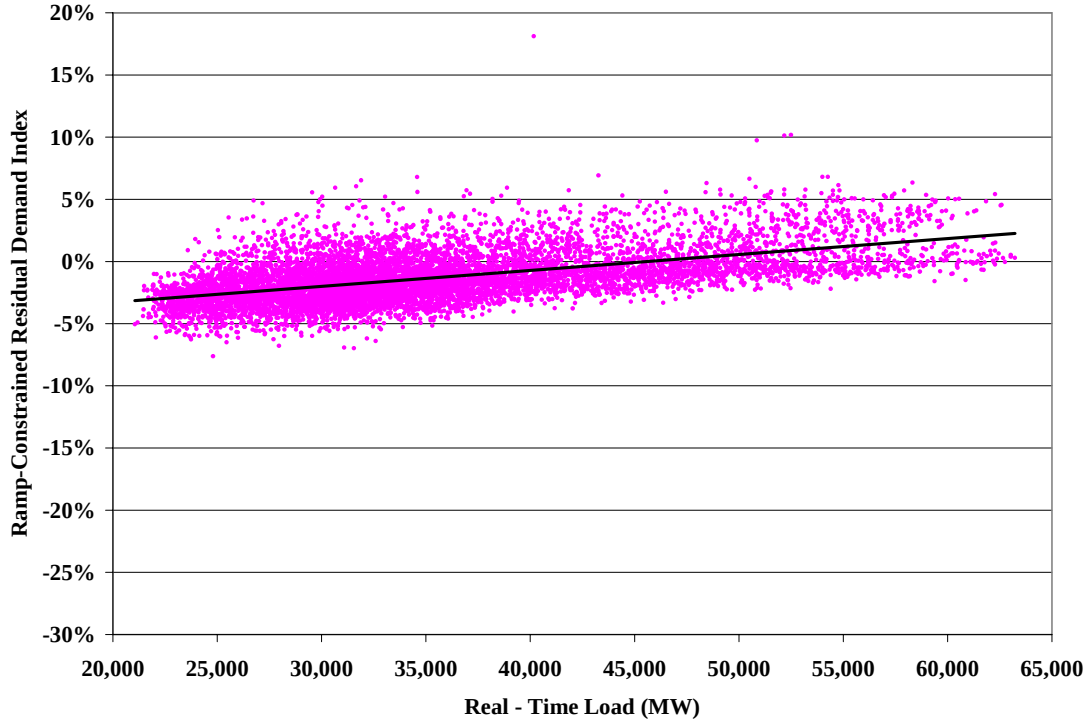
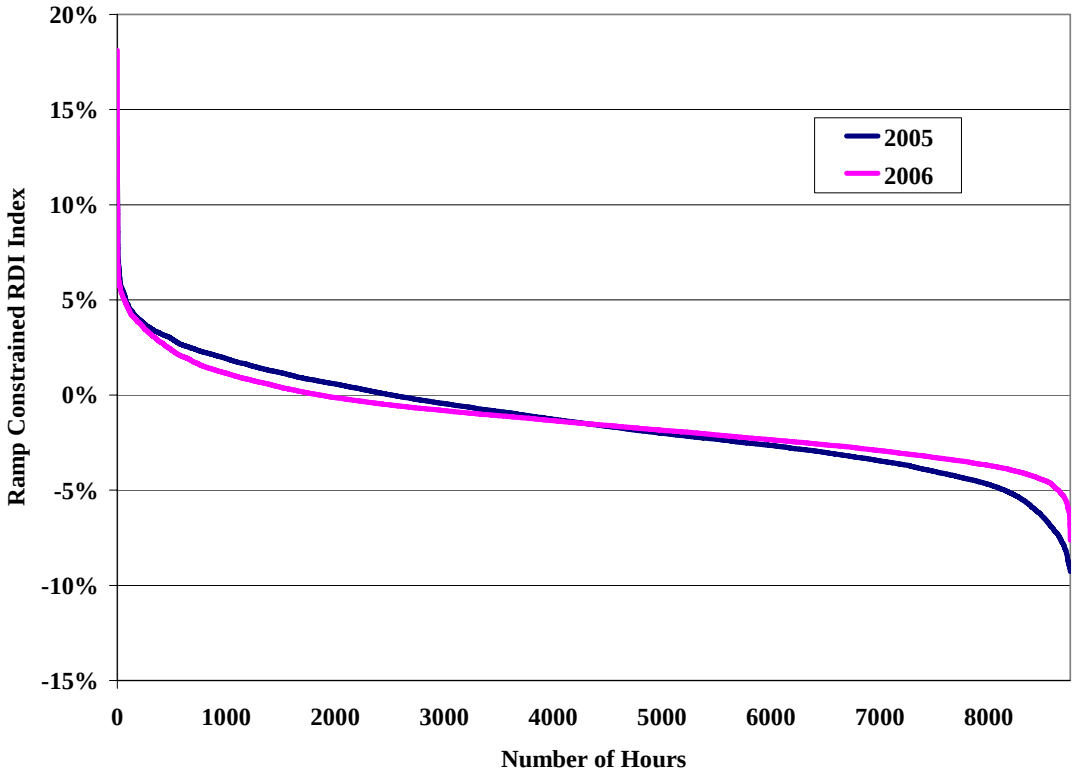


Figure 69: Ramp-Constrained Balancing Energy Market RDI vs. Actual Load
2006



In 2006, the instances when the RDI was positive occurred over a wide range of load levels, from 25 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 2006 are similar in shape to 2005, although the frequency of data points that are positive is significantly lower in 2006 than in 2005. This difference is highlighted in Figure 70 which compares the balancing energy market RDI duration curves for 2005 and 2006.

Figure 70: Ramp-Constrained Balancing Energy Market RDI Duration Curve 2005 & 2006



In 2006, there were 1,861 hours (21.2 percent) when the balancing energy market RDI was greater than zero, which means a supplier was pivotal in the balancing energy market 21.2 percent of the time in 2006. In contrast, there were 2,525 hours (28.8 percent) when the balancing energy market RDI was positive in 2005. Hence, the frequency with which a supplier was pivotal in the balancing energy market decreased 26 percent in 2006 indicating that the overall competitiveness of the balancing energy market improved in 2006. Among other factors, this decrease can be attributed to an average reduction in up balancing energy deployments in 2006, which was influenced by the existence of the under-scheduled charges associated with the

replacement reserve market. Figure 71 examines how the balancing energy market RDIs are correlated with balancing energy market prices as adjusted for gas prices in 2006, and Figure 72 shows the same data for 2005.

Figure 71: Ramp-Constrained Balancing Energy Market RDI vs. Balancing Energy Price Adjusted for Fuel Price 2006

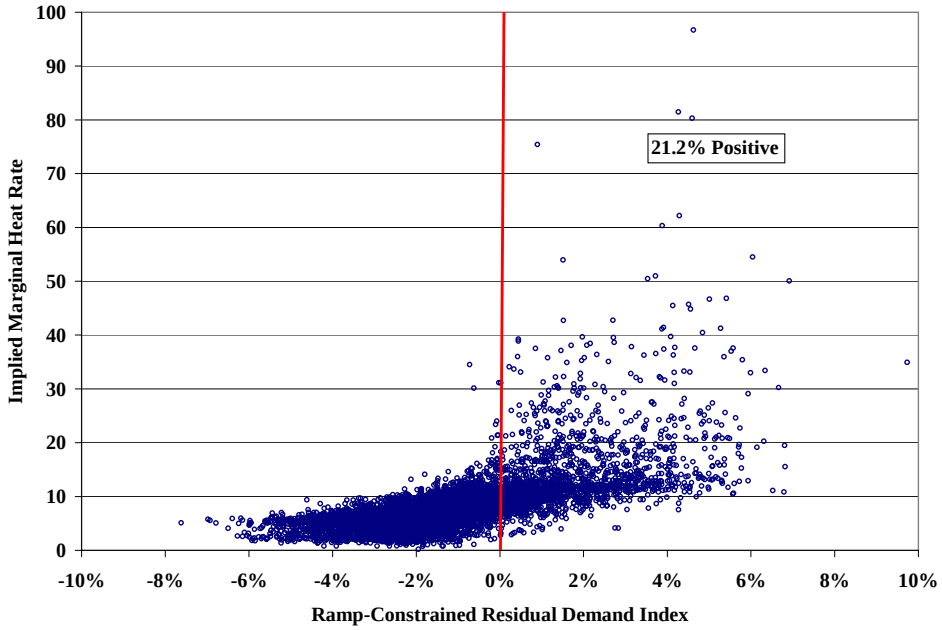
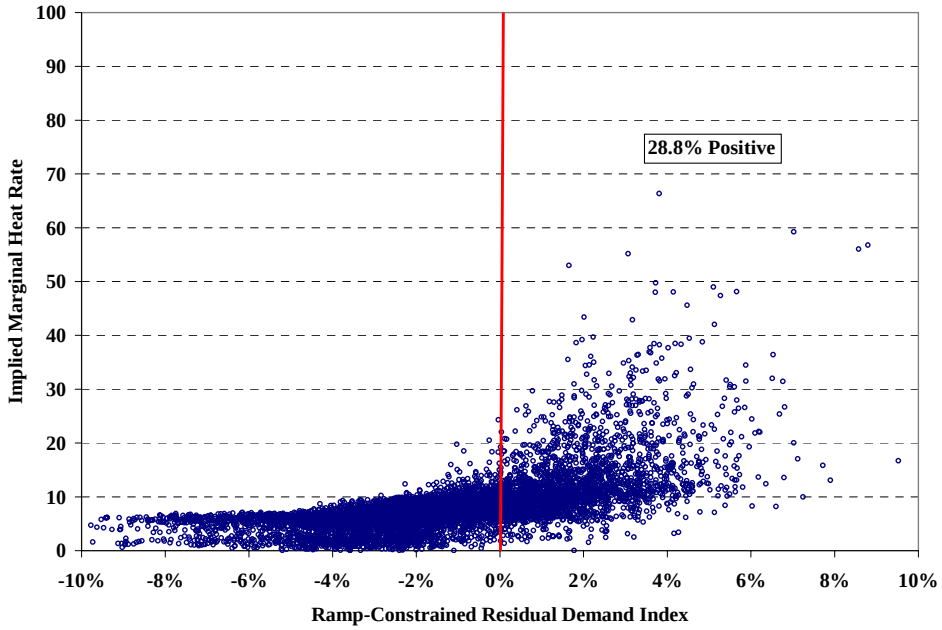


Figure 72: Ramp-Constrained Balancing Energy Market RDI vs. Balancing Energy Price Adjusted for Fuel Price 2005



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 2005 and 2006, with the rate of change becoming exponentially larger as the balancing energy market RDI enters the positive range. However, Figure 70 reveals that the number of data points with positive ramp-constrained balancing energy market RDIs is over 26 percent less in 2006 than in 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 67 through Figure 72 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 73 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 73: Short-Term Deratings by Load Level and Participant Size
June to August, 2006

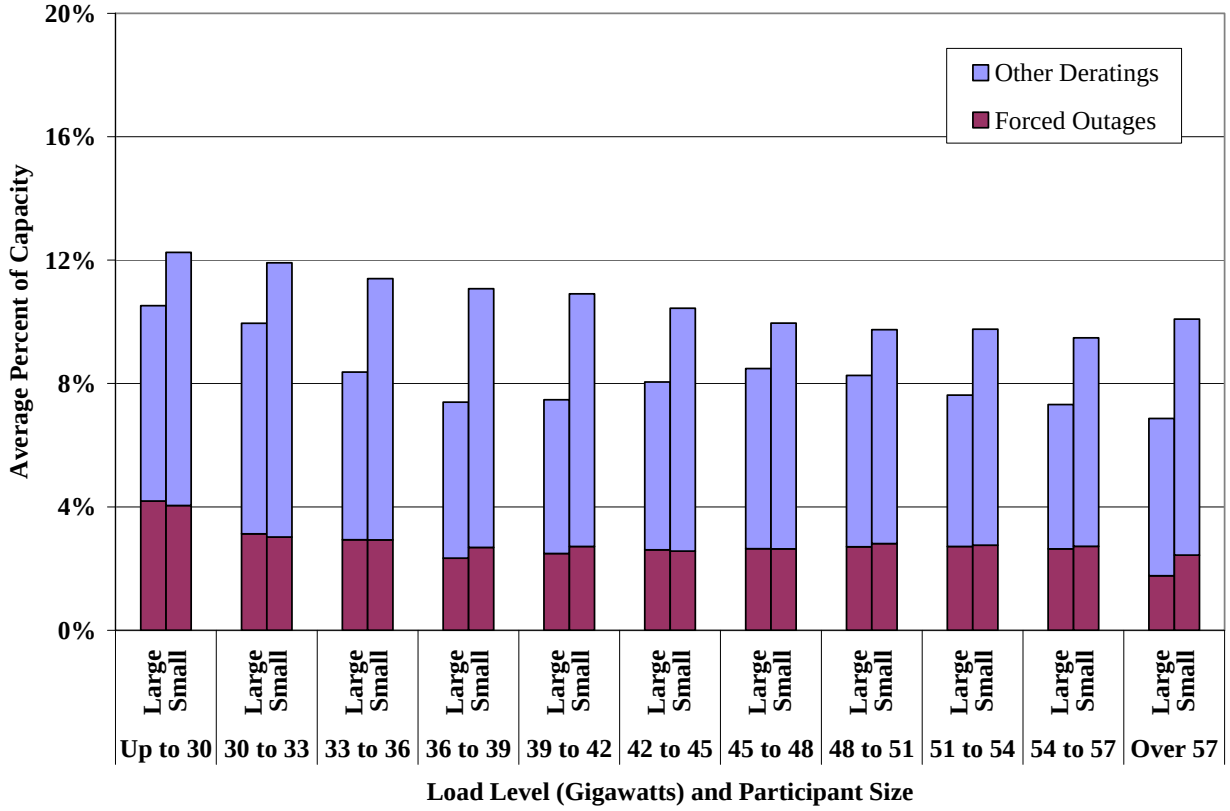


Figure 73 suggests that as electricity demand increases, both large and small market participants tend to make more capacity available to the market. For large and small suppliers, the short-term derating or forced outage rates decreased from approximately 11 to 12 percent at low demand levels to about 7 to 10 percent at load levels above 51 GW.

Large suppliers have derating rates that are lower than those of small suppliers across the range of load levels. Furthermore, large suppliers' deratings and outages generally decline as load levels increase. Given that the market is more vulnerable to market power at the highest load levels, these derating patterns do not provide evidence of physical withholding by the large suppliers. Although these data do not provide evidence of physical withholding by the large suppliers, the average derating rates for large and small suppliers are approximately 2 and 5 percent higher, respectively, than in 2005 at load levels greater than 51 GW. One possible explanation for this is the heightened awareness and importance that has been placed on the submission and updating of accurate resource plans following the rolling blackouts in April 2006

which may result in QSEs more accurately reflecting the physical unit capabilities in the resource plans that they submit to ERCOT.

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 74 shows the relationship between the output gap from committed resources and real-time load for all hours during 2006.

---

[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 74: Output Gap from Committed Resources vs. Actual Load 2006

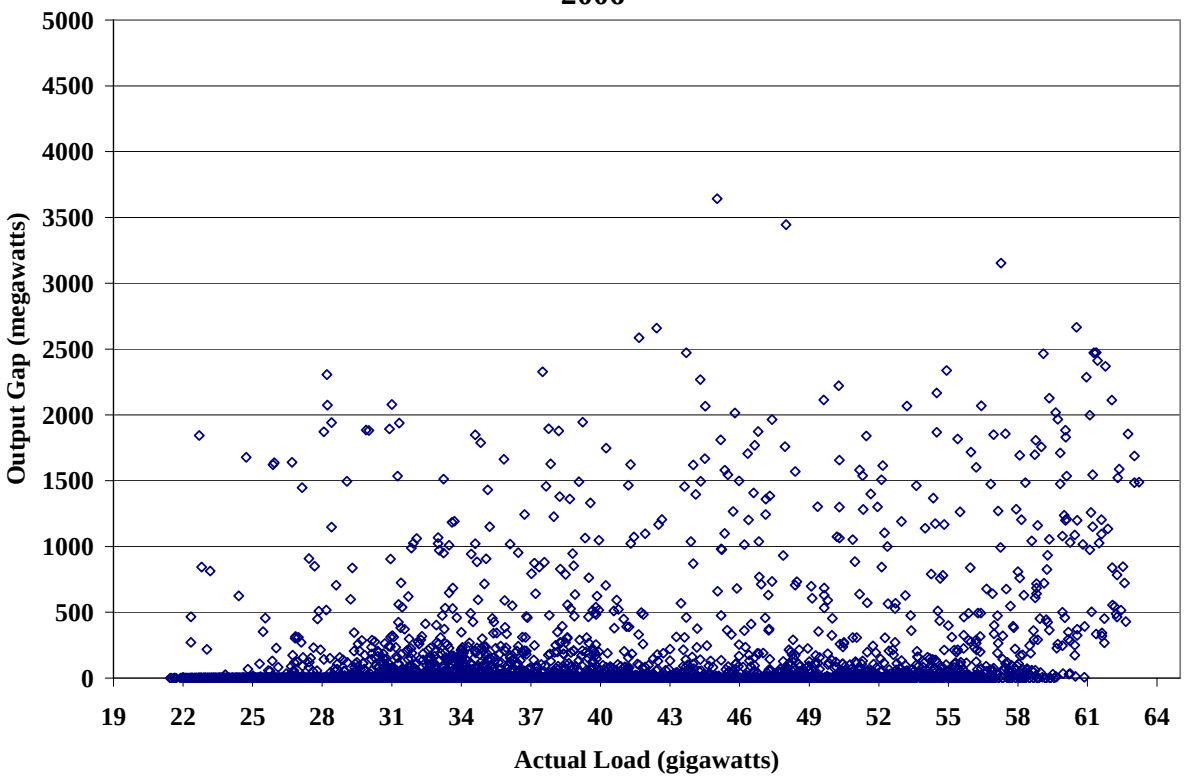


Figure 74 shows that the output gap from committed resources ranged from zero in most hours to a maximum of around 3,500 MW during 2006. As more clearly shown in Figure 75, the average output gap from committed resources rises slightly with real-time demand. This is not surprising given that clearing 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 75 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 75: Output Gap by Load Level and Participant Size
2006

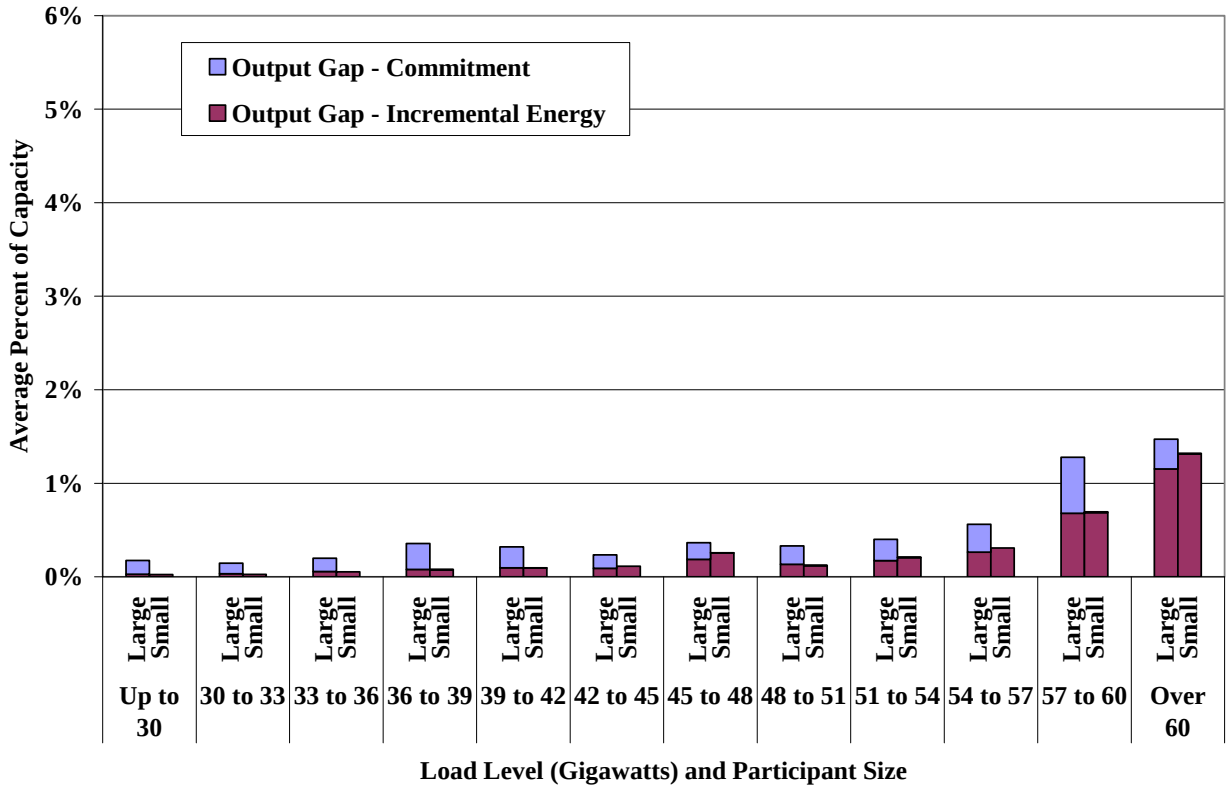


Figure 75 shows that the output gap quantities for incremental energy of large and small suppliers were comparable across all load levels, but that large suppliers had substantially higher output gaps for commitment across all load levels. The greater output gaps for large suppliers were driven primarily by the failure to commit economic resources, as this measure was close to zero over all load levels for small suppliers. However, the overall output gap for both large and small suppliers was reduced considerably in 2006 as compared to 2005. Overall, based upon the analyses in this section, we find that the competitive performance of the market improved in 2006.