POTOMAC ECONOMICS

2004 STATE OF THE MARKET REPORT
FOR THE
ERCOT WHOLESALE ELECTRICITY MARKETS

POTOMAC ECONOMICS, LTD.

Advisor to Wholesale Market Oversight
Public Utility Commission of Texas

July 2005

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

We wish to acknowledge the helpful input and numerous comments provided by the staff of the Wholesale Market Oversight of the Public Utility Commission of Texas, including Parviz Adib, Richard Greffe, Danielle Jaussaud, Julie Gauldin, Sam Zhou, and David Hurlbut. We are also grateful for the assistance of ERCOT in providing the data used in this report and in responding to our inquiries regarding the operation of the market.

EXECUTIVE SUMMARY

This report reviews and evaluates the outcomes of the ERCOT wholesale electricity markets in 2004. 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. 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 two previous reports we have issued regarding the ERCOT electricity markets.[1] These reports have included a number of recommendations designed to improve the performance of the current ERCOT markets. Many of these recommendations have been considered by ERCOT working groups and some have been embodied in protocol revision requests ("PRR"). We make reference to proposed changes that will address some of the issues and reiterate key recommendations for which no action has been taken. However, many of the issues identified in this report could be effectively addressed by the introduction of an alternative wholesale market design, which is currently being considered by participants and regulators Texas.

A. Review of Market Outcomes

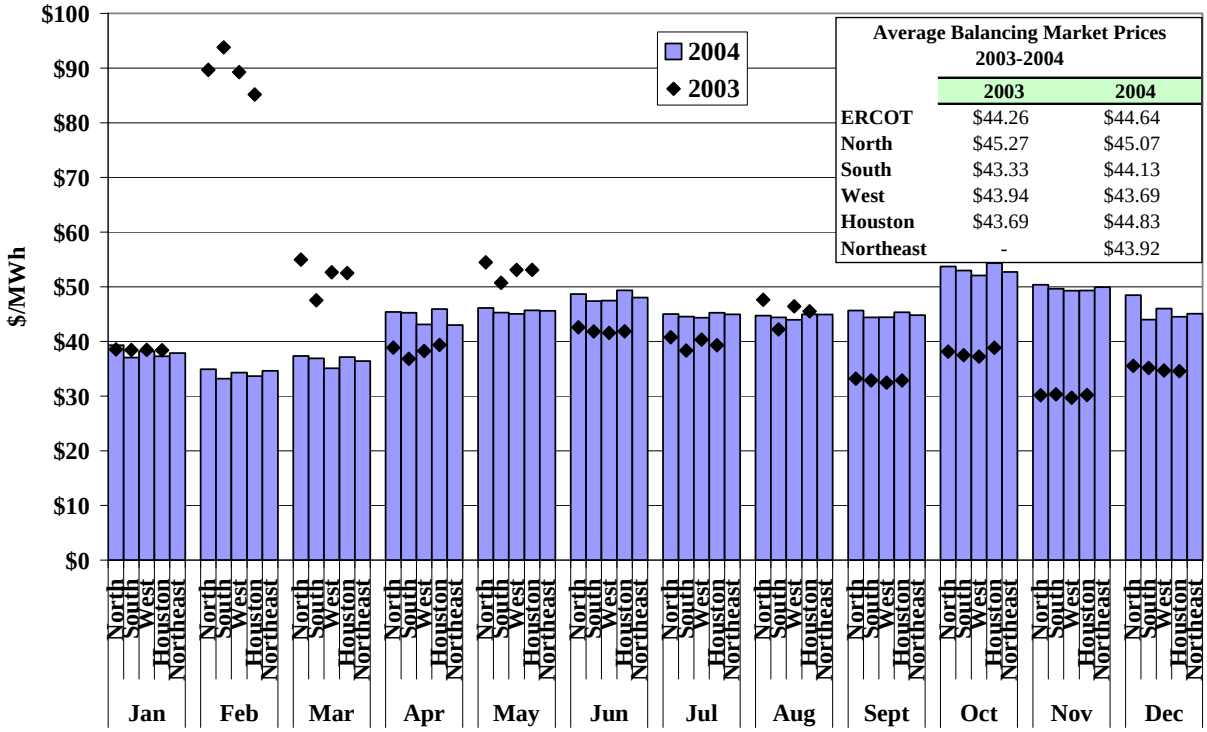
1. Balancing Energy Prices

The balancing energy market allows participants to make real-time purchases and sales of energy in addition to their forward schedules. While only a small portion of total 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").

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

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. The following figure shows the monthly average balancing energy prices in 2003 and 2004.

Balancing Energy Market Prices
2003 & 2004



Average Balancing Market Prices 2003-2004

	2003	2004
ERCOT	$44.26	$44.64
North	$45.27	$45.07
South	$43.33	$44.13
West	$43.94	$43.69
Houston	$43.69	$44.83
Northeast	-	$43.92

Balancing energy market prices in 2004 were similar to 2003 on an annual average basis, although the monthly average prices in the two years differed substantially. These differences were primarily due to fluctuations in natural gas prices. The prices in both years were more than 70 percent higher than in 2002. This was primarily due to considerable increases in natural gas prices. Average natural gas prices in 2003 and 2004 were both more than 65 percent higher than fuel prices in 2002. The effect of natural gas prices on electricity prices is consistent with expectations because the fuel costs constitute the majority of most generators' marginal costs (which should determine generators' balancing energy offers in a competitive market). Additionally, while all generation is not fired by natural gas, energy produced from natural gas-

fired resources are on the margin setting prices in the vast majority of the hours because most of the other resources are base-loaded.

However, the higher average prices during 2003 in February and March were primarily due to tight conditions in the natural gas market. These conditions were most severe on February 24-26, 2003 when balancing energy market prices exceeded $900 per MWh. These periods caused the prices in February 2003 to be 66 percent higher than they would have been without the three days of extreme prices. These three days increased the average prices for the year by 6.3 percent. The fact that such a small number of high-priced hours can have a significant effect on the average prices over the entire year illustrates the significant influence that price spikes can have on the economic signals provided by the market. It also reinforces the importance of ensuring that price spikes occur efficiently – i.e., that prices rise efficiently during periods of legitimate shortages and that price spikes do not result from withholding in the absence of a shortage.

The higher prices during the fall of 2004 can be partially attributed to higher natural gas prices during the fall 2004. However, offer patterns by a large supplier in the balancing energy market also contributed to these higher prices. We previously identified 95 intervals between October 27 and December 8 when these offer patterns contributed to prices that exceeded $200/MWh.[2] If these intervals were excluded, prices would have been 8.6 percent lower from October through December and 2.0 percent lower for all of 2004.

The figure also shows that the price differences between the zones tend to be relatively small, reflecting only moderate amounts of interzonal congestion. In both years, the North Zone exhibited the highest average prices, while the lowest prices occurred in the South Zone in 2003 and in the West Zone in 2004.

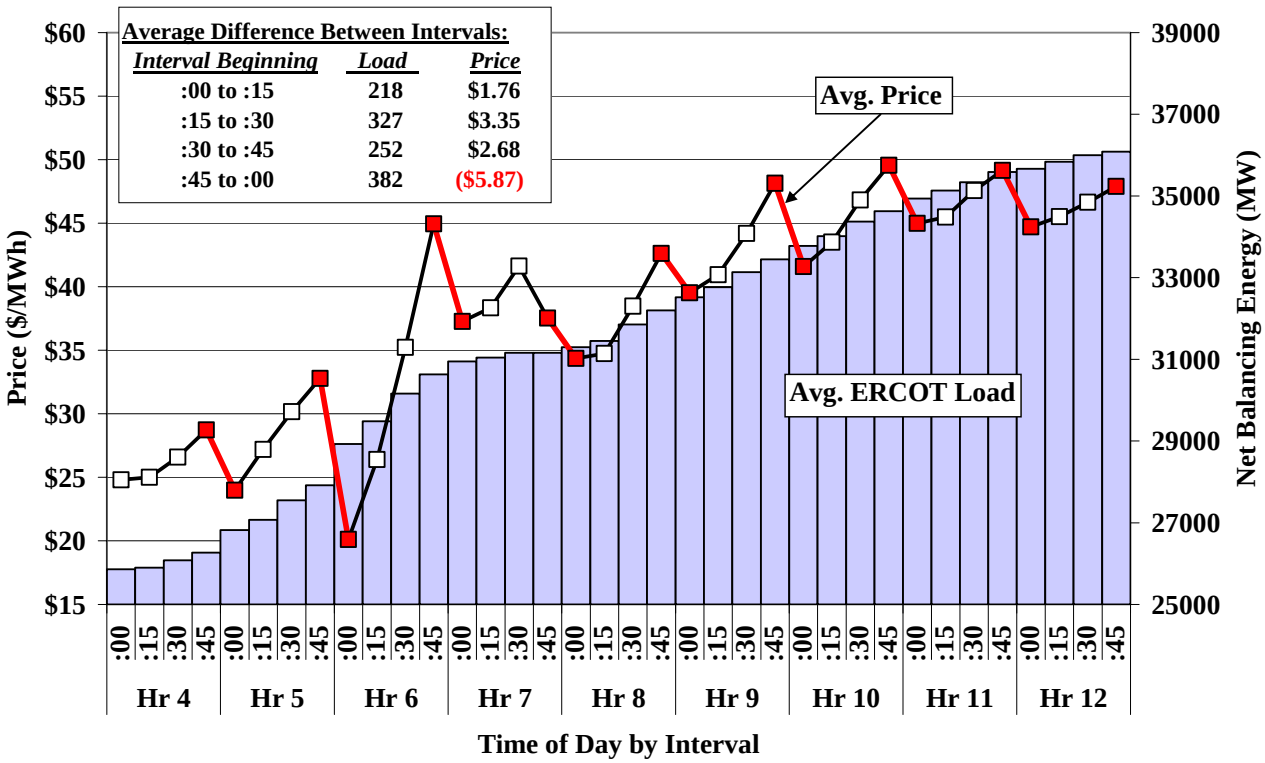
The report evaluates two other aspects of the balancing energy prices: 1) the primary determinants of the prices, and 2) the correlation of the prices with forward electricity prices in Texas. With regard to the determinants of balancing energy prices, one should expect that prices

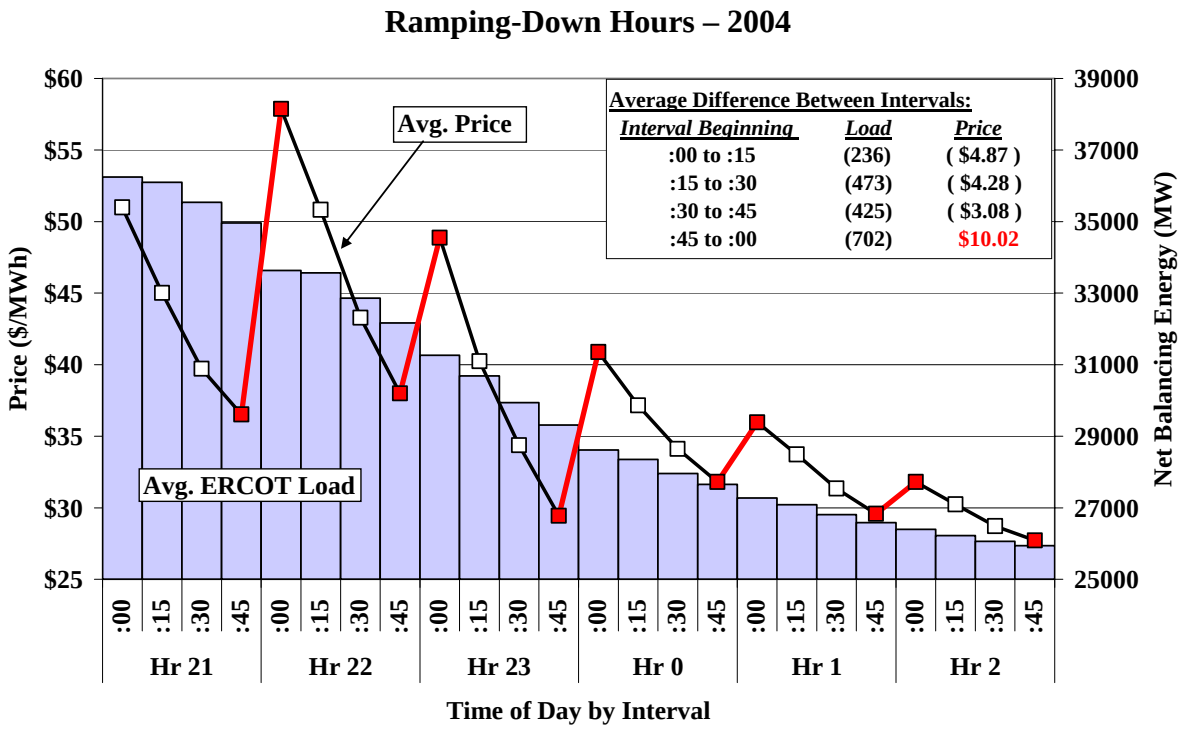
[2] **"**Investigation into the Causes for the Shortages of Energy in the ERCOT Balancing Energy Market and into the Wholesale Market Activities of TXU from October to December 2004", Potomac Economics, March 2005.

would be primarily determined by load levels and fuel prices in a well-functioning spot market. Although there is a strong relationship between fuel prices and balancing energy prices, we do not observe a strong relationship between prices and actual load levels in ERCOT. Instead, we observe a clear relationship between the net balancing energy deployments and the balancing energy prices, which is unexpected in a well-functioning market.

The report concludes that the observed relationship is primarily due to the hourly scheduling patterns of most of the market participants. We observe that 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. The following 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.

Average Balancing Energy Prices and Load by Time of Day
Ramping-Up Hours -- 2004

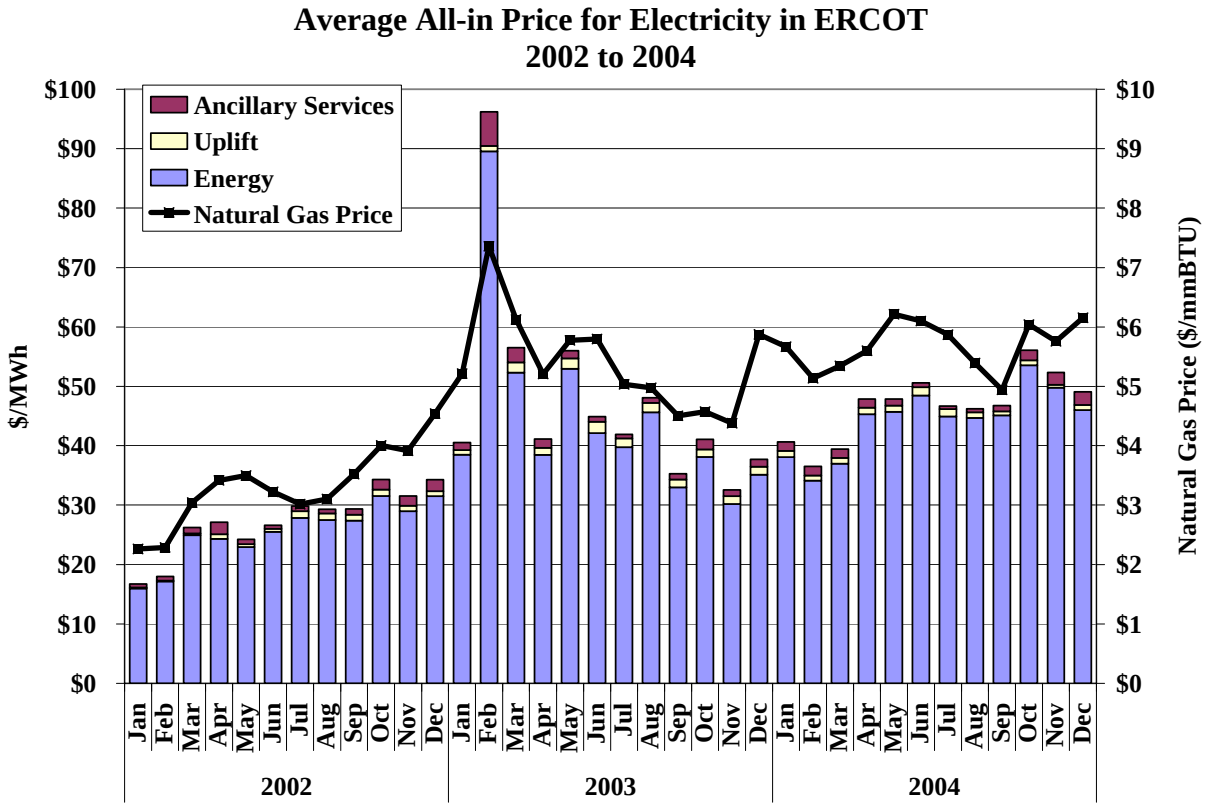




These pricing patterns and the fact that balancing energy prices are not as strongly correlated with actual load as expected raises significant efficiency concerns regarding the operation of the balancing energy market. These concerns and the recommendations we have made to address the concerns are discussed below.

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"), 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 three years are shown in the figure below along with a natural gas price trend.

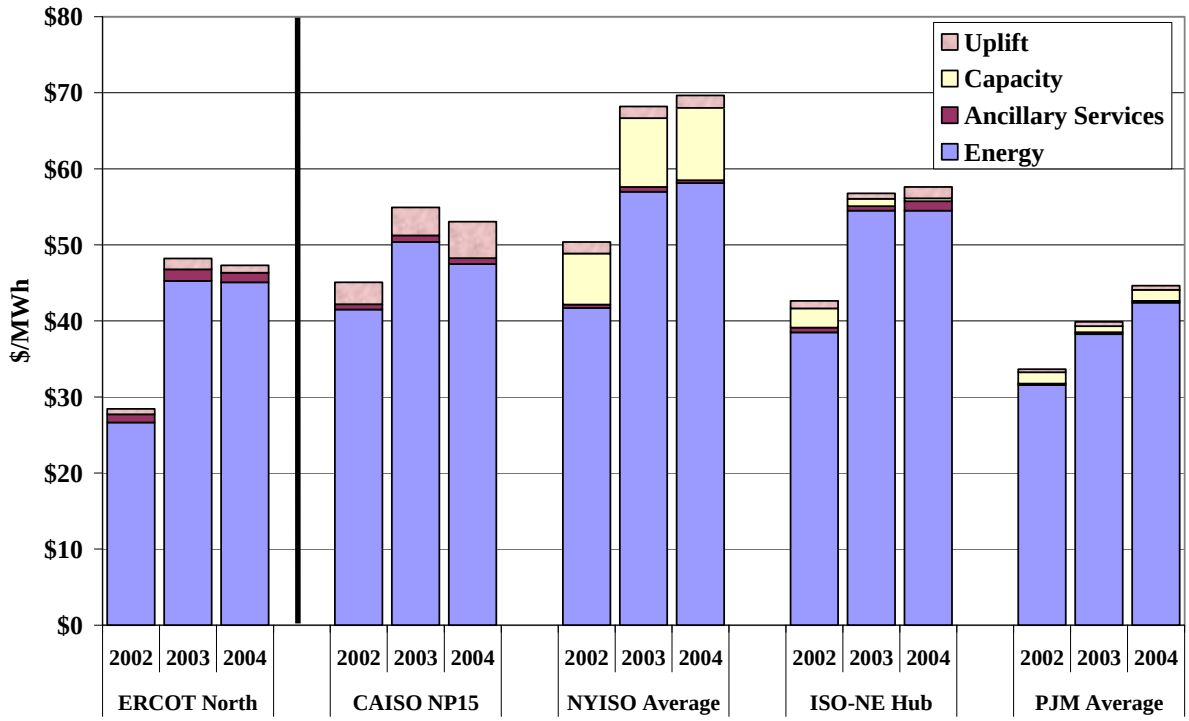


With the exception of the cold snap during February 2003, all-in prices are fairly stable from month-to-month. It is notable, however, that there is only a weak relationship between all-in prices and load levels. Energy prices have not risen significantly during the summer months, while ancillary services prices have actually gone down during the summer.

This figure indicates that natural gas prices were a primary driver of the trends in electricity prices from 2002 to 2004. 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 by more than 65 percent from 2002 to 2003 and by an additional 5 percent through 2004. Likewise, the all-in energy price rose by 72 percent from 2002 to 2003. However, the all-in price for electricity decreased by 1 percent because ancillary services costs decreased by 17 percent. The decrease in ancillary services costs was primarily due to a decrease in the average procurement of up and down regulation. There was also a 32 percent reduction in uplift costs for local congestion management in 2004, which is discussed below.

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. Figure 3 compares the all-in prices for the five major centralized wholesale markets in the U.S.: (a) ERCOT, (b) California ISO, (c) New York ISO, (d) ISO New England, and (e) PJM ISO. For each region, the figure reports the average cost (per MWh of load) for (a) energy, (b) ancillary services (reserves and regulation), (c) capacity markets (if applicable), and (d) uplift for economically out-of-merit resources.

Comparison of All-In Prices across Markets
2002 to 2004



Each market experienced a substantial increase in energy prices from 2002 to 2003 due to increased fuel costs, but prices were comparable between 2003 and 2004. Although the markets vary substantially in the portion of their generating capacity that is fueled by natural gas, these units are usually on the margin and set the wholesale spot prices in the majority of hours for all markets shown.

In 2002, ERCOT exhibited the lowest all-in price -- 18 percent lower than the next lowest-priced market. In 2003 and 2004, the all-in price in PJM, which experienced the lowest increase in

prices after 2002, was lower than in ERCOT. Natural gas-fired generation is on the margin less frequently in PJM than any of the other markets because PJM has access to large quantities of coal-fired generation within PJM itself and in the Midwest. The all-in prices in the ERCOT region are relatively low due in part to its substantial resource margin.

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.

Ancillary services prices were generally higher in 2003 and 2004 than in 2002. Much of this increase can be attributed to the increase 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 regulation can incur such opportunity costs if they reduce the output from economic units to make the capability available to provide these services.

The prices for reserves and regulation tend to be lower during the summer than at other times of the year. This is because the required quantities of reserves and regulation are relatively constant over the year while the supply of resources that can provide reserves and regulation (i.e., on-line capacity not scheduled for energy) tends to increase in proportion to load. The highest-priced periods for ancillary services shown in Figure 16 occurred at the end of 2004. This happened for two reasons. First, there was an increase in the frequency of price spikes in the balancing energy market during this period that raised the opportunity costs of providing ancillary services. Second, demand was relatively low so that less capacity was committed and therefore there was less capacity on line and available to provide ancillary services.

ERCOT continued to incur higher costs for reserves and regulation than other markets in 2002. This is due in part to the higher quantities of regulation and responsive reserves that are required in ERCOT due to its limited interconnections with adjacent areas. This report concludes that the ancillary services prices in ERCOT are generally higher than expected. For example, responsive

reserves prices averaged more than $11 per MWh, which is substantially higher than similar markets in other regions. We identify two explanations for this:

- A considerable portion of the available capability in ERCOT is not scheduled or offered in the ancillary services markets. Less than one-third of the regulation capability was scheduled or offered in the regulation market in 2004, while approximately 50 percent of the available responsive reserves capability and 25 percent of the non-spinning reserves capability were scheduled or offered.
- The sequential design of the ERCOT ancillary services and energy markets (ancillary services are procured in advance of the energy market rather than being jointly-optimized with the dispatch of energy) leads to higher costs because it results in an allocation of resources to provide ancillary services that is suboptimal. The only market with higher responsive reserves prices is PJM, which also does not jointly-optimize the procurement of reserves and energy.

We understand that co-optimization is being contemplated in the design of the Texas Nodal markets that are currently under consideration. If the Texas Nodal markets are adopted, we would encourage implementation of ancillary services markets that are jointly-optimized with the energy markets.

In the short-term, ERCOT plans to modify the procurement process for ancillary services under a new release of market software in September 2005, so that the markets for regulation, responsive reserves, and non-spinning reserves will clear simultaneously. This change is likely to result in increased prices in the responsive reserve market to reflect the higher marginal costs of providing non-spinning reserves. Since the costs of providing non-spinning reserves may be partly attributable to ERCOT's deployment procedures as discussed in the body of this report, it will be particularly important to consider potential improvements to these procedures.

4. Net Revenue Analysis

A final analysis of the outcomes in the ERCOT markets in 2004 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 and reflects 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 below long-run expected levels 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 prices 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 2002 to 2004 for two types of units, a natural gas combined-cycle generator and a natural gas single cycle turbine. The net revenue increased significantly from 2002 to 2004, largely due to higher natural gas prices and increased balancing energy purchases in 2004, both of which led to higher balancing energy prices. Despite this rise in net revenue, neither type of new generating unit would have earned sufficient net revenue to make the investment profitable. This is not surprising given the surplus of capacity that currently exists in ERCOT. However, net revenue should increase as retirements, mothballing, and load growth reduce the surplus capacity in the future.

Our analysis also shows that the net revenues in other markets are similarly insufficient to support new investment in generation due to capacity surpluses and mild weather conditions during 2004. There is one significant difference, however, between ERCOT and some of the other markets. ERCOT currently has no market mechanism that will ensure that its market sends economic signals that will allow it to maintain a sufficient base of generating resources once the surplus dissipates. There are two primary market mechanisms employed in other areas to ensure economic signals are sufficient to maintain adequate resources:

- 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.

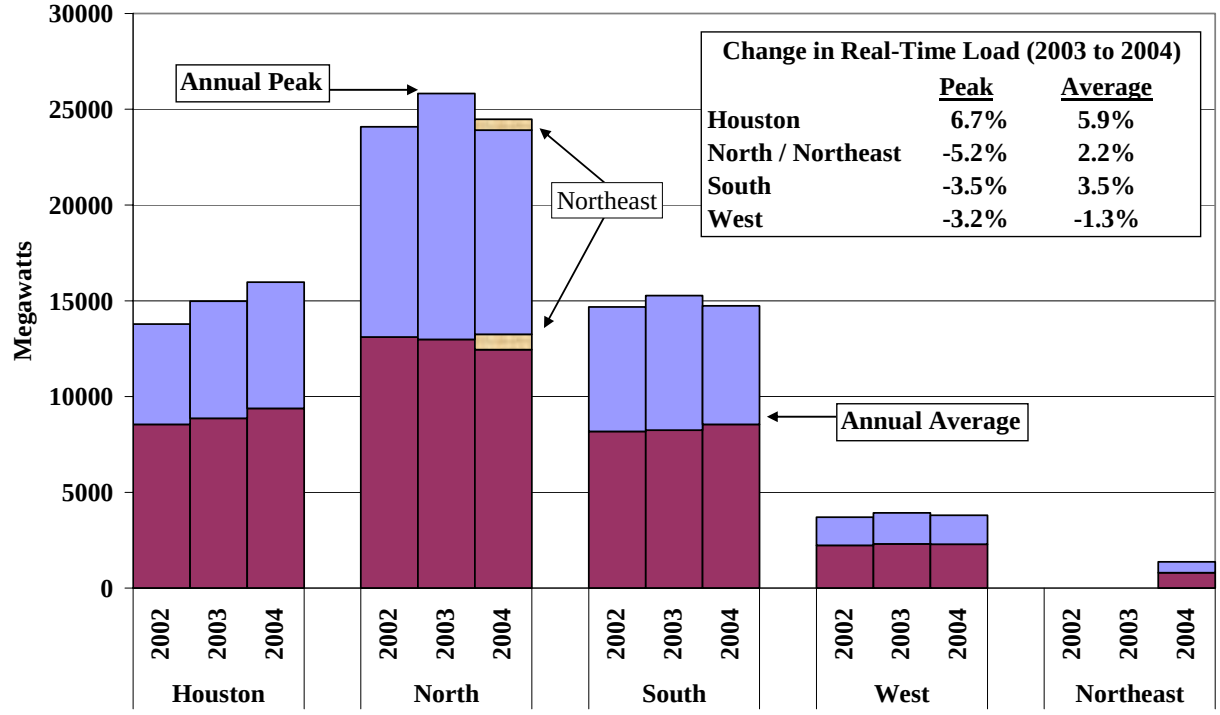
Absent one or both of these market mechanisms, ERCOT may ultimately have to rely on some form of mandated investment to maintain adequate resources once the current capacity surplus dissipates.

B. Demand and Resource Adequacy

1. Electrical Loads in 2004

Load levels remain one of the fundamental factors that determine the conditions in any electricity market. Because electricity cannot be stored, the electricity market must ensure that generation matches load on a continuous basis. The figure below shows that load increased on average by only 3 percent from 2003 to 2004. However, the peak demand decreased by 2.5 percent due to the cooler weather that occurred in 2004 than in 2003.

Annual ERCOT Load Statistics by Zone
2002 to 2004



Significant changes in these peak demand levels are very important because they determine the probability and frequency of shortage conditions, although no shortages occurred under peak demand conditions in 2004 due to ERCOT's relatively high resource margins. More broadly,

peak demand levels and capability of the transmission network are the primary factors that determine whether the existing generating resources are adequate to maintain reliability.

2. Generation Capacity In ERCOT

The report also provides an accounting of the current ERCOT generating capacity, which is dominated by natural gas-fired resources. These resources account for 73 percent of generation capacity in ERCOT as a whole, and 85 percent in the Houston Zone. This makes ERCOT particularly vulnerable to natural gas price spikes because the other resource types (coal and nuclear) are primarily base load units that are generally not the marginal source of supply.

Our analysis also shows that ERCOT has substantial excess capacity. Resource margins (the percentage by which total capacity exceeds peak demand) for ERCOT as a whole have remained relatively constant from 2003 to 2004. When import capability, resources that can be switched to the SPP, and Loads acting as Resources are excluded from the calculation, the resource margin in 2004 was 24 percent. When these classes of capacity are included, the resource margin is 33 percent. It is notable that both the peak load and the generating resources were approximately 1500 MW lower in 2004. Hence, if the 2003 peak load had been achieved in 2004, the actual resource margins would have decreased in 2004.

Although these resource margins are sizable, it is important to consider that electricity demand in Texas has been growing at a rapid pace and that a significant number of generating units in Texas are soon reaching or are already exceeding their expected lifetimes. These factors may cause the resource margins in ERCOT to diminish rapidly over the next three to five years.

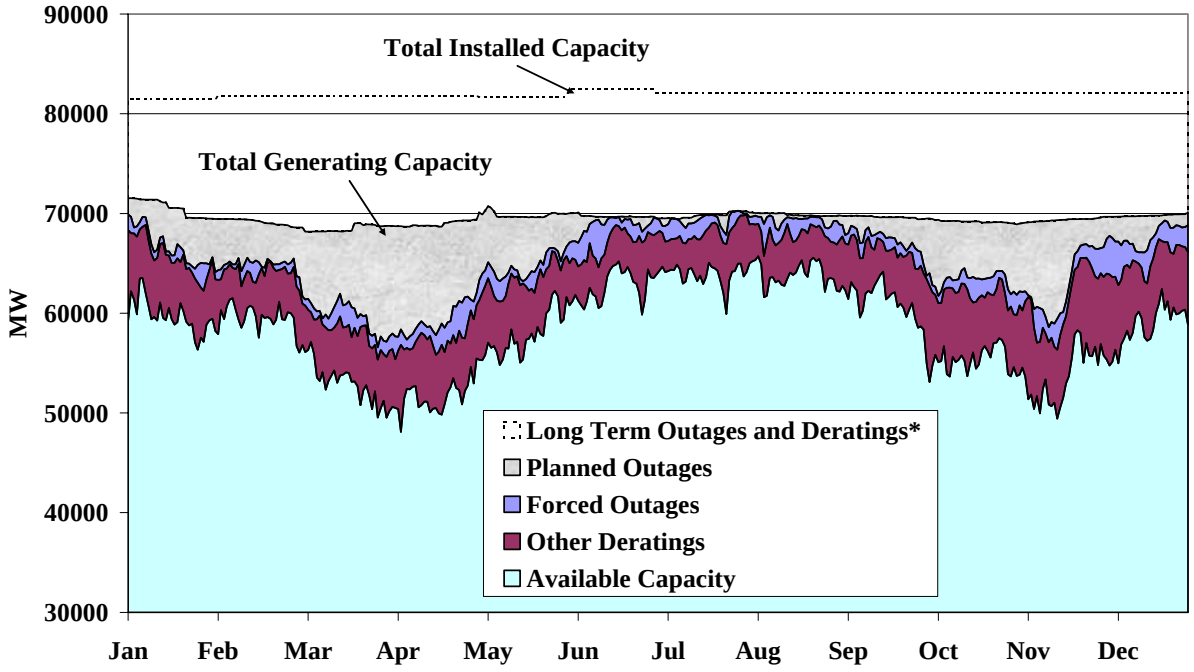
3. Generator Outages and Commitments

Despite the relatively high resource margins, 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 a generating resource's installed capability 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 following figure shows the daily available and derated capability of generation in ERCOT.

Short and Long-Term Deratings of Installed Capability
2004



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

This figure shows that long-term outages and deratings typically range from 10 GW to 14 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); 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 5 percent during the summer in 2003 to as high as 10 percent in other months. These deratings include outages not reported or correctly logged by ERCOT and natural deratings due to ambient conditions and other factors. These outages and deratings do not raise any significant issues.

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, operating reserves, and up regulation. Hence, it measures the total generation available for dispatch in excess of the electricity needs each day.

The report finds that excess capacity is significant in ERCOT, averaging almost 7,000 MW in 2004. These results show that the ERCOT system is generally over-committed, indicating significant inefficiencies in the outcomes of the current ERCOT markets. The tendency to over-commit capacity can be attributed in large part to the lack of a centralized day-ahead commitment process in ERCOT. Without a centralized commitment mechanism, each participant makes independent generator commitment decisions that, taken together, are not likely to be optimal. Hence, the introduction of day-ahead energy and operating reserves markets promises substantial efficiency improvements in the commitment of generating resources.

4. 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 2004, 63 resources totaling 1826 MW of capability were qualified as LaaRs. the amount of responsive reserves provided by LaaRs gradually increased from about 900 MW at the beginning of 2004 to 1,100 MW at the end of 2004. Currently, LaaRs are permitted to supply up to 1,150 MW of the responsive reserves requirement. This represents a relatively large share of the total 2,300 MW requirement for responsive reserves. Although the participants with LaaRs resources are qualified to participate in non-spinning reserves and balancing up energy markets, they have not participated in those markets up until now. 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, providing non-spinning reserves introduces a much higher probability of being curtailed. 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.

C. Transmission and Congestion

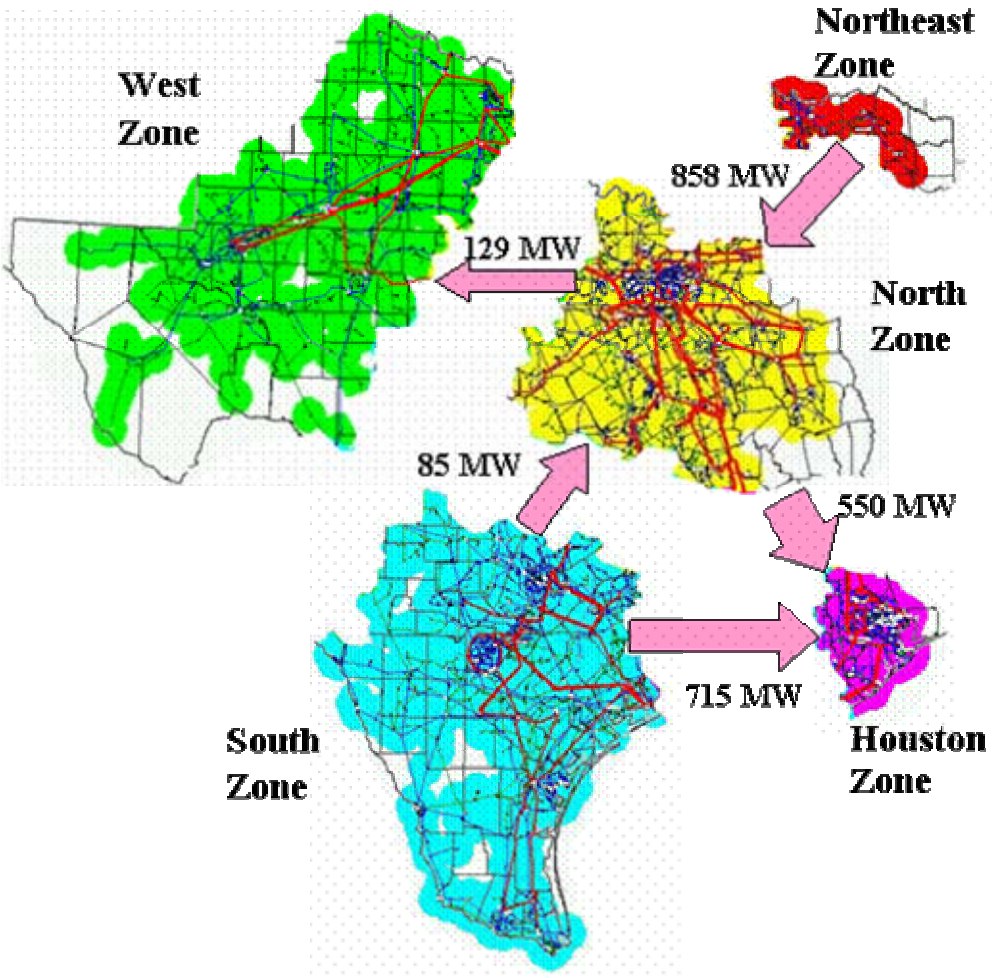
One of the most important functions of any electricity market is to manage the flows of power over the transmission network, limiting additional power flows over transmission facilities when they reach their operating limits. In ERCOT, constraints on the transmission network are managed in two ways.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

1. Electricity Flows between Zones and Interzonal Congestion

The balancing energy market uses the Scheduling, Pricing, and Dispatch ("SPD") software that dispatches energy in each zone 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 five transmission interfaces. The transmission interfaces are referred to as Commercially Significant Constraints ("CSCs"). The following figure shows the average flows modeled in SPD during 2004 over each of these CSCs.

Average Modeled Flows on Commercially Significant Constraints
2004



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 the three primary CSCs. When these flows cause congestion, it is beneficial to create a new CSC, such as the North to Houston CSC implemented by ERCOT in 2004, 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.
- SPD calculated net flows from the North Zone to the West Zone on average, while the West to North CSC was defined to only limit flows in the opposite direction. Not surprisingly, a new North to West CSC was defined for 2005 because ERCOT has found that congestion occurs in both directions.

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 remained modest in 2004, totaling approximately $41 million. This reflects an increase of $16 million from the interzonal congestion costs in 2003. Most of this increase can be attributed to the creation of the Northeast zone at the beginning of 2004. The congestion between the Northeast and North zones in 2004 had previously occurred as intrazonal or "local" congestion within the North zone in 2003.

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.

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. This shortfall on an annual basis decreased from approximately $10 million in 2003 to almost $8 million in 2004. This decrease

occurred even though the overall levels of interzonal congestion increased, indicating an increase in the consistency of the modeled limits and the TCR amounts for the CSCs.

The pricing of the congestion rights is also important because the revenue from the auction of the congestion rights is the primary means for the loads to receive the value of the transmission system that they pay for through regulated payments to transmission owners. 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 2002, the annual auction for the congestion rights resulted in prices that substantially over-valued the congestion rights on the South to North and South to Houston interfaces. In 2003, the congestion rights auction prices for all of the interfaces decreased considerably, resulting in a much closer convergence with the actual value of the congestion rights. In 2004, the convergence improved further. Convergence was good even on the new CSCs created in 2004, for which the participants had no historic information. This indicates that market participants' ability to forecast interzonal congestion and the overall liquidity of the TCR market have improved, resulting in better valuation and pricing of the transmission rights.

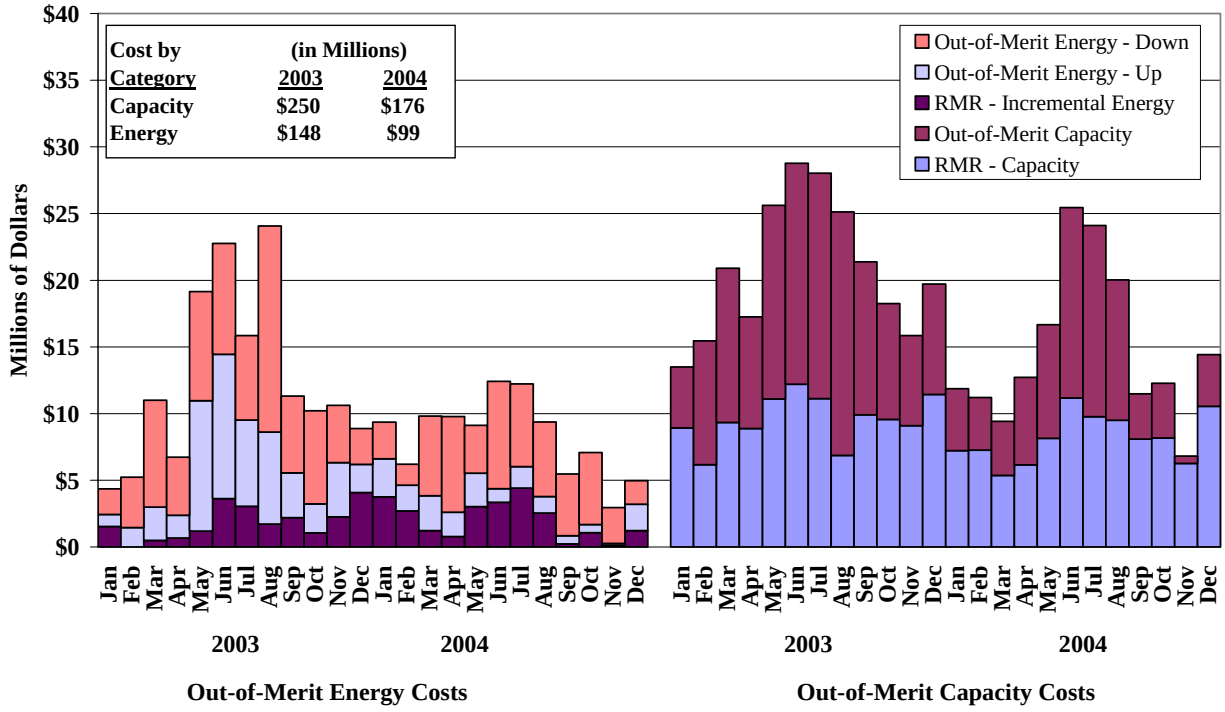
2. Local Congestion and Local Capacity Requirements

ERCOT manages local (intrazonal) congestion using out-of-merit dispatch ("OOME up" and "OOME down"), which causes units to depart from their scheduled quantities. 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. RMR agreements were signed with certain generators needed for local reliability. When these units are called out-of-merit order, they receive revenues specified in the agreements rather than standard OOME or OOMC payments. Understanding the causes and patterns of local congestion is important. The following figure shows the out-of-merit energy and capacity costs, including RMR costs, for each month in 2003 and 2004.

The figure shows that OOME costs and incremental energy costs from RMR units declined from $148 million to $99 million from 2003 to 2004, a decrease of 33 percent. Likewise, the costs of OOMC and the capacity costs from RMR units declined 30 percent in 2004. The most

substantial percentage decrease in these costs between 2003 and 2004 was associated with payments for OOME-Up, which declined 64 percent. The figure also shows that all classes of out-of-merit costs tend to increase during the summer when higher loads increase the need for ERCOT operators to take out-of-merit actions to manage local congestion and reliability needs.

Expenses for Out-of-Merit Capacity and Energy
2003-2004



The report finds three primary factors that contributed to the reduction of these out-of-merit costs. First, the addition of the Northeast Zone at the beginning of 2004 allowed a significant amount of congestion that had been local to become interzonal congestion between the Northeast and revised North zones in 2004. This change reduced the OOME Down dispatch within the Northeast and the OOME Up in DFW and other areas within the North zone. Second, the definition of an additional CSC from the North zone to Houston reduced local congestion by allowing the zonal energy market to manage the congestion on the transmission facilities connecting the two areas. Third, the formula for OOMC payments was revised, which reduced the incentive for suppliers to wait for ERCOT to commit their units through the OOMC process on days when the units would otherwise be economic.

3. Conclusions regarding Transmission Congestion in ERCOT

The results in this area of the report confirm prior findings in the 2003 SOM Report and the Market Operations report that:

- the vast majority of congestion in ERCOT is intrazonal, which is difficult for loads to hedge and is not transparent;
- the current zonal market can result in large inconsistencies between the interzonal flows calculated by SPD and the actual flows over the CSC interfaces; and
- these inconsistencies can result in under-utilized transmission capability and difficulties in defining transmission rights whose obligations can be fully satisfied.

The most complete long-run remedy for both the interzonal and intrazonal issues identified in this report would be to implement nodal markets, an option that is currently being evaluated in ERCOT. These markets would provide transparent prices for both generators and loads that would fully reflect all transmission constraints on the ERCOT network.

Absent implementation of nodal markets, we continue to recommend the following changes from the Market Operations Report to improve the management of interzonal and local congestion.[3]

- Improve the process for designating zones to minimize the effects of the simplifying zonal assumptions.
- Improve the process for evaluating and revising CSC definitions.
- Modify the calculation methodology of the zonal average shift factor to exclude generation whose output is generally fixed (e.g., nuclear units).
- Provide ERCOT the operational flexibility to temporarily modify the definition of a CSC associated with topology changes.
- Modify the multi-step balancing energy market optimization to recognize the interactions between its local congestion management and zonal balancing energy deployments to minimize the costs of both classes of deployments.

Protocol Revision Requests ("PRR") have been approved by the Technical Advisory Committee to address the first four recommendations.[4] However, a decision on the last recommendation has

---

[3] The Commission has opened Project No. 30634, Activities Related to Implementation of Recommendations from the Potomac Economics 2004 Report on the Operation of the ERCOT Wholesale Electricity Markets, to address these recommendations.

[4] See PRRs 587, 589, and 592.

been deferred pending a decision on whether ERCOT will move to a nodal market design. This last recommendation is particularly important because local congestion management can have large indirect effects on portfolio energy deployments and the balancing energy prices. In the Market Operations Report, we concluded that current multi-step process does not efficiently consider the interaction between actions taken to resolve local congestion versus those taken to resolve interzonal congestion, resulting in inefficient market results and artificial price spikes in the balancing energy market. The last recommendation addresses this concern.

In addition, we continue to recommend that ERCOT consider the feasibility and benefits of creating a new zone for Dallas-Fort Worth, which would likely cause much of the remaining OOME costs to be reflected in the balancing energy prices and would reduce uplift costs. However, we recognize that there are a number of important issues that would need to be considered in making this change in the short-run.

D. Balancing Energy Offers and Schedules

QSEs play an important role in the current ERCOT markets. QSEs must submit balanced schedules with scheduled resources that match their scheduled load. With the introduction of "relaxed balanced scheduling" in November 2002, there is no longer a requirement that the balanced schedules closely follow the QSE's actual load. The energy schedules are a primary input to determine the net supply and demand for balancing energy. In general, 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. QSEs also submit balancing energy offers to increase or decrease their energy output from the scheduled level. The balancing up offers correspond to the unscheduled output from the QSE's 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 QSE's schedules, offers, and resource plans.

1. Scheduling Patterns

We evaluate forward scheduling patterns by comparing load schedules to actual real-time load. In the aggregate, load schedules tend to be under-scheduled by an average of almost 1 percent and by higher amounts under peak demand conditions. In some hours, the load is under-scheduled by 10 to 20 percent, which creates a sizable demand for balancing energy. This under-scheduling together with the balancing energy offer patterns described below sometimes result in large balancing energy price increases.

The North and Houston zones are under-scheduled most significantly with the under-scheduling amounts ranging from approximately 4 percent on average to 9 percent in high-load periods. Persistent load imbalances are not necessarily a problem. It can reflect the fact that more energy from economic resources is typically available in the balancing energy market. On the other hand, over-scheduling of load in other zones such as the West zone reflects that under Relaxed Balanced Scheduling, the load schedules do not have to reflect the load that is actually expected. Rather than selling power to the balancing energy market through energy imbalances or deployments in the balancing energy market, QSEs that over-schedule load sell into the balancing market through load imbalances. This poses no operational concerns and is a mechanism by which some suppliers may more fully utilize their portfolio.

However, QSEs with generators in *locally*-constrained areas can benefit from systematic over- and under-scheduling. The local congestion management process provides incentives for QSEs to over-schedule in export-constrained areas and under-schedule in import-constrained areas. Our analysis in this report shows that this has occurred for resources that are frequently committed or dispatched out of merit.

2. 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 400 MW each 15-minute interval.

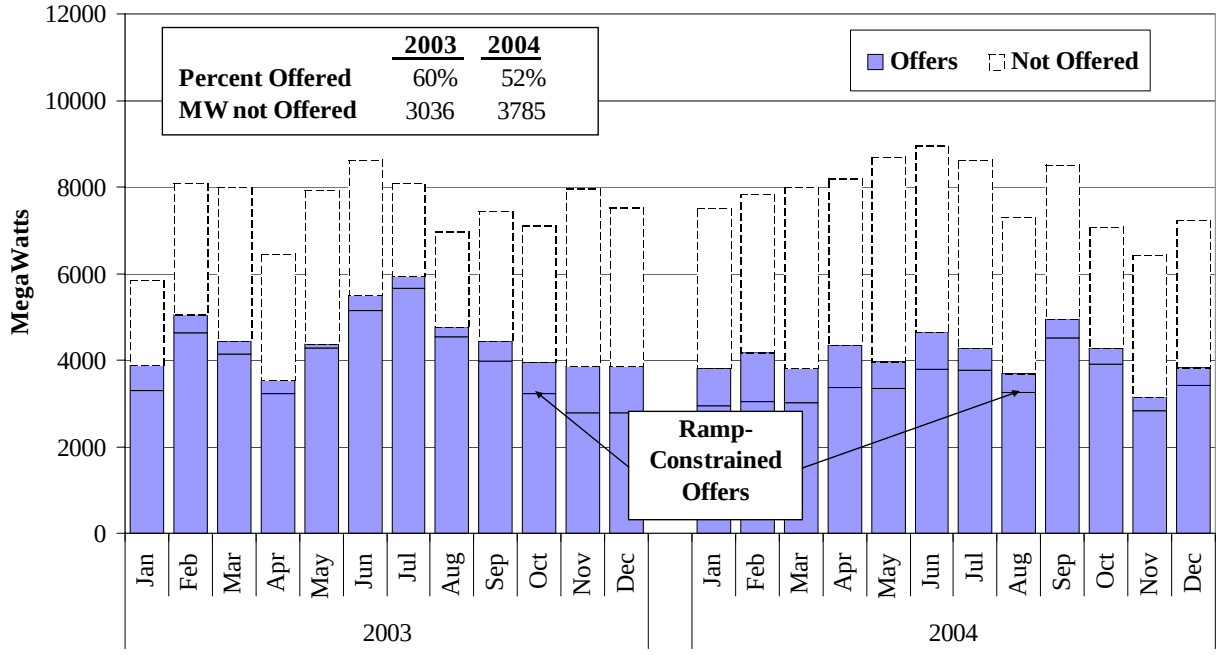
- Although QSE's can modify their schedules each interval, most only change their schedules hourly, resulting in schedule changes averaging 1000 to 3000 MW in these hours (and sometimes significantly larger).
- The inconsistency between the changes in schedules and actual load in these hours places an enormous burden on the balancing energy market, resulting in the erratic pricing patterns shown above.
- The largest two QSEs schedule much more flexibly than the other QSEs and generally help to mitigate these problems.

To address this issue and improve the performance of the balancing energy market, the report recommends changes that may increase the willingness of QSEs to submit flexible schedules (i.e., schedules that change every 15 minutes). These recommendations were considered by ERCOT and its participants, but were not proposed in any pending PRR.

3. 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 figure below shows the total available energy versus the amount offered in the balancing energy market on average in each month during 2003 and 2004.

Available Balancing Energy vs. Balancing Energy Offers
Daily Peak Load Hours – 2003 and 2004



This figure and the other analysis of the portfolio offers indicate that:

- In general, approximately half of the available energy is offered in the balancing energy market.
- The largest QSEs offer a much higher share of their available energy than smaller suppliers.
- Participants generally offer little more than the amount that can be deployed in a single interval (the additional amount is labeled “ramp-constrained offers” in the figure).

It is a significant concern that not all available capacity is offered into the market. Part of this problem can be attributed to the fact that gas turbine capacity is difficult to effectively offer in the balancing energy market, of which ERCOT currently has more than 3000 MW.

The report also identifies a number of concerns regarding the difficulties of offering all available ramping capability from online or quick-starting resources. In particular, the current market rules and portfolio bidding framework results in ramp limitations that are much lower than the true physical ramp limitations of the individual generating units. This reduces the ability of the market to fully utilize the generating resources and can result in inefficient transitory fluctuations in balancing energy prices.

The report includes a number of recommendations to address the portfolio ramp limitations and allow gas turbine capacity to be included in the portfolio offers. These recommendations include:

- Considering the feasibility of allowing QSEs to offer multiple ramp rates that vary by output level;
- Modifying the treatment of ramp limitations in the balancing energy market to recognize ramping capability that is used/made available associated with QSEs' schedule changes.
- Encouraging QSEs to submit multiple "sub-QSE" portfolio offers to reduce the ramp limitation effects of having all of a QSE's supply subject to a single ramp constraint.

E. 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.

1. Summary of Resource Plan Revisions

This subsection of the report summarizes changes in the resource plans between the day ahead and real time, and evaluates the different reasons underlying the resource plan changes. QSEs make changes to their resource plans that reflect changes in information between the day-ahead and the operating period. The following factors explain most changes made to the resource plans.

- *Changes in the Load Forecast* – Weather forecasts and load expectations are constantly changing up until real-time. When expected load increases, QSEs respond by committing additional generation and increasing planned generation. Conversely, when the load forecast decreases, QSEs respond by de-committing resources and decreasing planned generation.
- *Out-Of-Merit Commitments by ERCOT* – When ERCOT commits generation for reliability, it leads to more on-line capacity overall. Frequently, QSEs respond to an OOMC instruction by de-committing other resources to maintain the same overall level of capacity. On occasion, these de-commitments have led to additional reliability issues.

- *Plant Technology and Portfolio Composition* – This evaluation finds that there is some variation in resource plan revision trends due to variations in plant characteristics and portfolio characteristics.

2. Resource Plans and Out-of-Merit Actions

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 provides evidence that market participants have engaged in strategies that increase:

- *OOMC Commitment* – Our analysis suggests that QSEs with resources that frequently receive OOMC instructions regularly delay the decision to commit those units until after ERCOT determines which resources to select for OOMC. This behavior forces ERCOT to make more OOMC commitments, resulting in higher local congestion uplift costs.
- *OOME Up Dispatch* – QSEs with resources that frequently receive OOME Up instructions typically under-schedule these resources in the final real-time resource plan. This behavior leads ERCOT to deploy these resources upward for local congestion management, resulting in higher local congestion uplift costs.
- *OOME Down Dispatch* – There is some evidence that QSEs with resources that frequently receive OOME Down instructions may over-schedule these resources in the final real-time resource plan. This kind of behavior leads ERCOT to deploy these resources downward for local congestion management. This results in higher local congestion uplift costs and higher balancing energy prices due to reduced supplies in the balancing market.

These analyses indicate that the current procedures for OOME and OOMC provide incentives for participants to submit resource plans that do not reflect anticipated real-time operations. This stems 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 act strategically to garner additional uplift payments.

F. 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 when load obligations are considered, the suppliers in ERCOT are rarely pivotal. However, because a large portion of the available energy from online resources is routinely not offered in the balancing energy market, we found that a supplier was pivotal in 3 percent of all hours and in 10 percent of hours when real-time load exceeded 40 GW. Although the balancing market may not reflect traditional market power, the factors described above that prevent full utilization of the available energy in ERCOT make the balancing market more vulnerable to manipulation. The report shows that the energy offered in the balancing energy market decreased in 2004 from 2003, which generally increases the frequency with which one or more suppliers can significantly increase the balancing energy prices. Part of this decrease can be attributed to the reduction of excess online generation caused by a reduction in OOMC commitments by ERCOT during the fall of 2004.

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 analyze physical and economic withholding in order to further evaluate competitive performance of the ERCOT market. Based on the analyses conducted in this area, the report finds little evidence of systematic physical or economic withholding of generating resources during 2004.

However, based on the results of an investigation published earlier this year focused on the period from October 27 to December 8, 2004,[5] a period during which a large number of price spikes occurred, we found:

- TXU's balancing energy offers associated with its gas turbines were not consistent with competition and contributed to a significant increase in balancing energy prices during

[5] *Investigation Into the Causes for the Shortages Of Energy in the ERCOT Balancing Energy Market and into the Wholesale Market Activities of TXU From October 27 To December 8, 2004*, Potomac Economics, April 2005.

the study period.

- Prices during the high-priced intervals would generally have cleared at roughly 50 percent lower had TXU offered its gas turbines at competitive price levels.

Consistent with the patterns we discuss above, we identified a relatively large quantity of available energy that could have been produced from on-line and quick-start resources by rival suppliers that was not offered in the balancing energy market during the period we investigated. If all of this energy had been offered, the price spikes would not have occurred. This confirms the conclusion that when significant quantities of available energy are not offered, it compromises the competitiveness of the balancing energy market. It also reinforces the need to implement changes that will increase the incentive for suppliers to offer their available energy in the balancing energy market, including the recommendations regarding portfolio scheduling and ramp limitations.

I. REVIEW OF MARKET OUTCOMES

A. Balancing Energy Market

1. Balancing Energy Prices During 2004

The balancing energy market is the spot market for electricity in ERCOT. As is typical, 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.

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).[6] 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 2004.

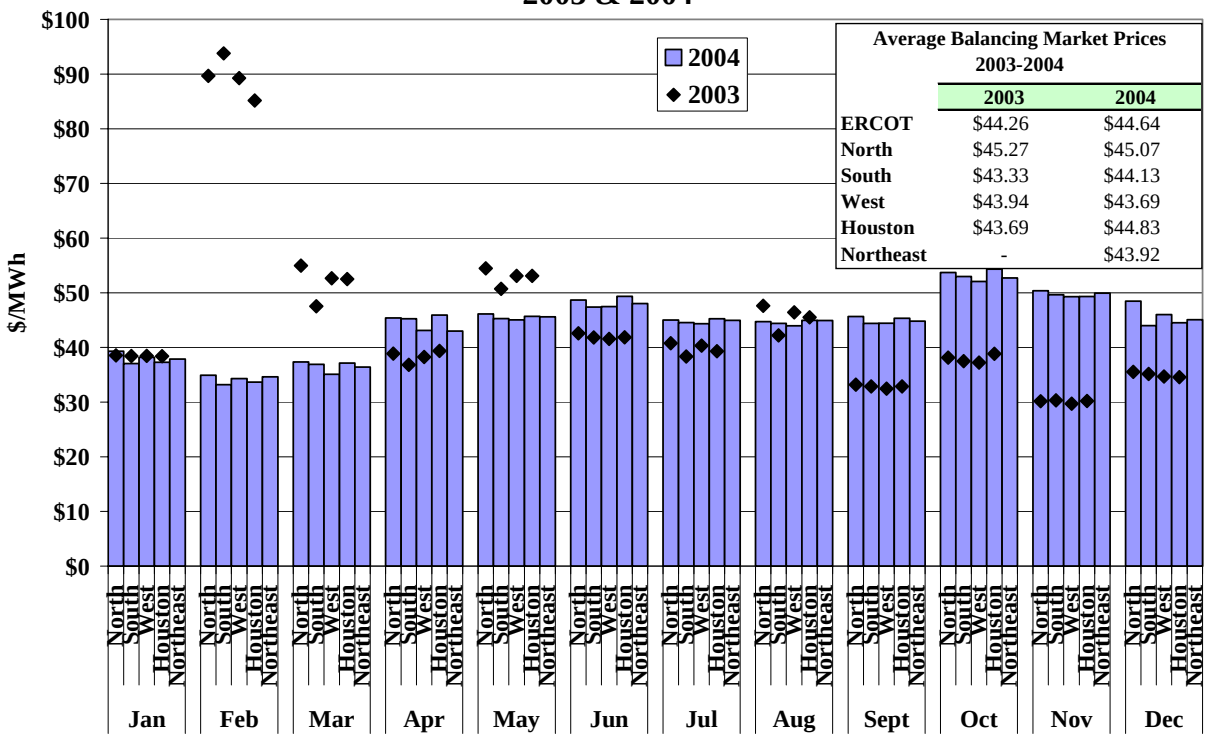
Balancing energy market prices in 2004 were similar to 2003 on an annual average basis, although the monthly average prices in the two years differed substantially. These differences were primarily due to fluctuations in natural gas prices. To summarize the price levels during the past two years, Figure 1 shows the load-weighted average balancing energy market prices in each of the ERCOT zones in 2003 and 2004.[7]

---

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

7 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 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.

Figure 1: Average Balancing Energy Market Prices
2003 & 2004



Average Balancing Market Prices 2003-2004

	2003	2004
ERCOT	$44.26	$44.64
North	$45.27	$45.07
South	$43.33	$44.13
West	$43.94	$43.69
Houston	$43.69	$44.83
Northeast	-	$43.92

Figure 1 shows that the prices in 2004 were significantly lower in February and March and considerably higher in September to December than in 2003. Most of these differences can be explained by changes in natural gas prices. The higher average prices during 2003 in February and March were primarily due to tight conditions in the natural gas market. These conditions were most severe on February 24-26, 2003 when balancing energy market prices exceeded $900 per MWh. These periods caused the prices in February 2003 to be 66 percent higher than they would have been without the three days of extreme prices. These three days increased the average prices for the year by 6.3 percent.

The fact that such a small number of high-priced hours can have a significant effect on the average prices over the entire year illustrates the significant influence that price spikes can have on the economic signals provided by the market. It also reinforces the importance of ensuring that price spikes occur efficiently – i.e., that prices rise efficiently during periods of legitimate shortages and that price spikes do not result from withholding in the absence of a shortage.

The higher prices during the fall of 2004 can be partially attributed to higher natural gas prices during the fall 2004. However, offer patterns by a large supplier in the balancing energy market also contributed to these higher prices. We previously identified 95 intervals between October 27 and December 8 when these offer patterns contributed to prices that exceeded $200/MWh.[8] If these intervals were excluded, prices would have been 8.6 percent lower from October through December and 2.0 percent lower for all of 2004.

Figure 1 also shows that the price differences between the zones tend to be relatively small, reflecting only moderate amounts of interzonal congestion. In both years, the North Zone exhibited the highest average prices, while the lowest prices occurred in the South Zone in 2003 and in the West Zone in 2004. The average difference in prices between the highest and lowest priced zones was approximately 4.5 percent in 2003 and 3.2 percent in 2004. The Northeast Zone was created at the beginning of 2004 from within the North Zone because it is an area that had exhibited frequent export constraints. The difference in prices between the Northeast and North Zones was 2.6 percent in 2004.

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". (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.) 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 2004.

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

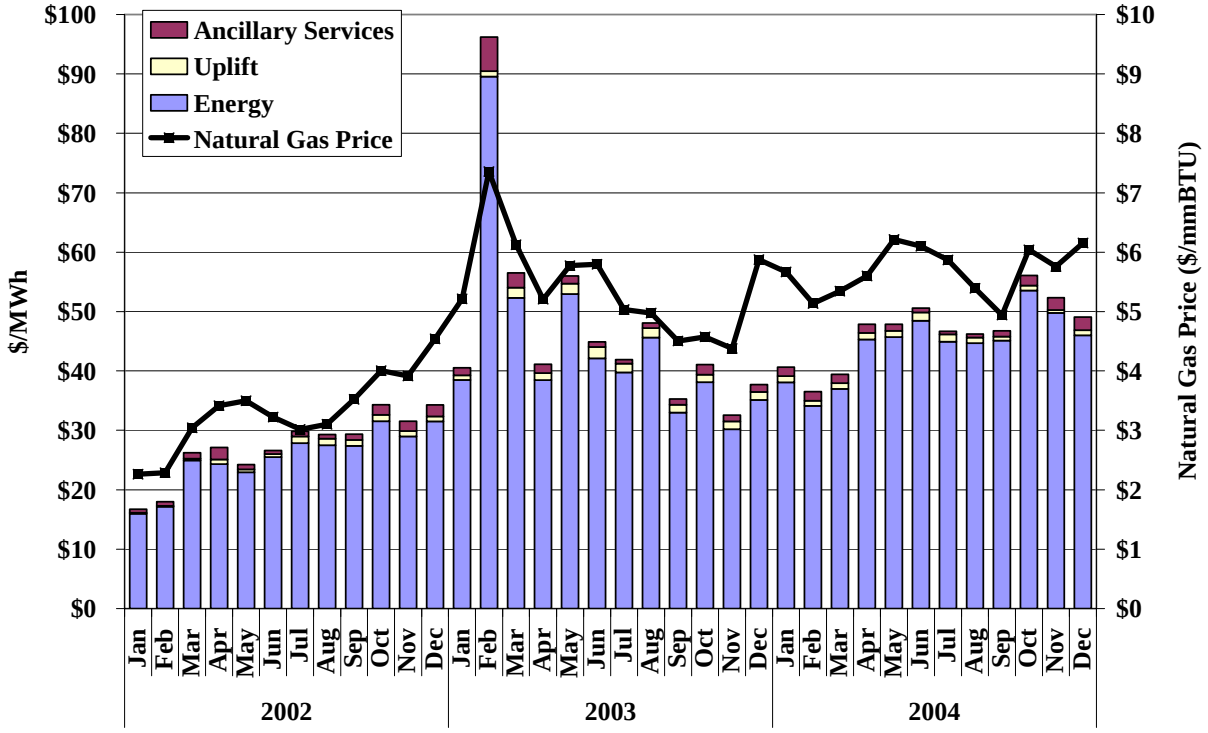
- Energy costs: Balancing energy market prices are used to estimate energy costs, under the assumption that the price of bilateral energy purchases converges with balancing energy market prices over the long-term, as discussed above.

[8] **"**Investigation into the Causes for the Shortages of Energy in the ERCOT Balancing Energy Market and into the Wholesale Market Activities of TXU from October to December 2004", Potomac Economics, March 2005.

- 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 2004



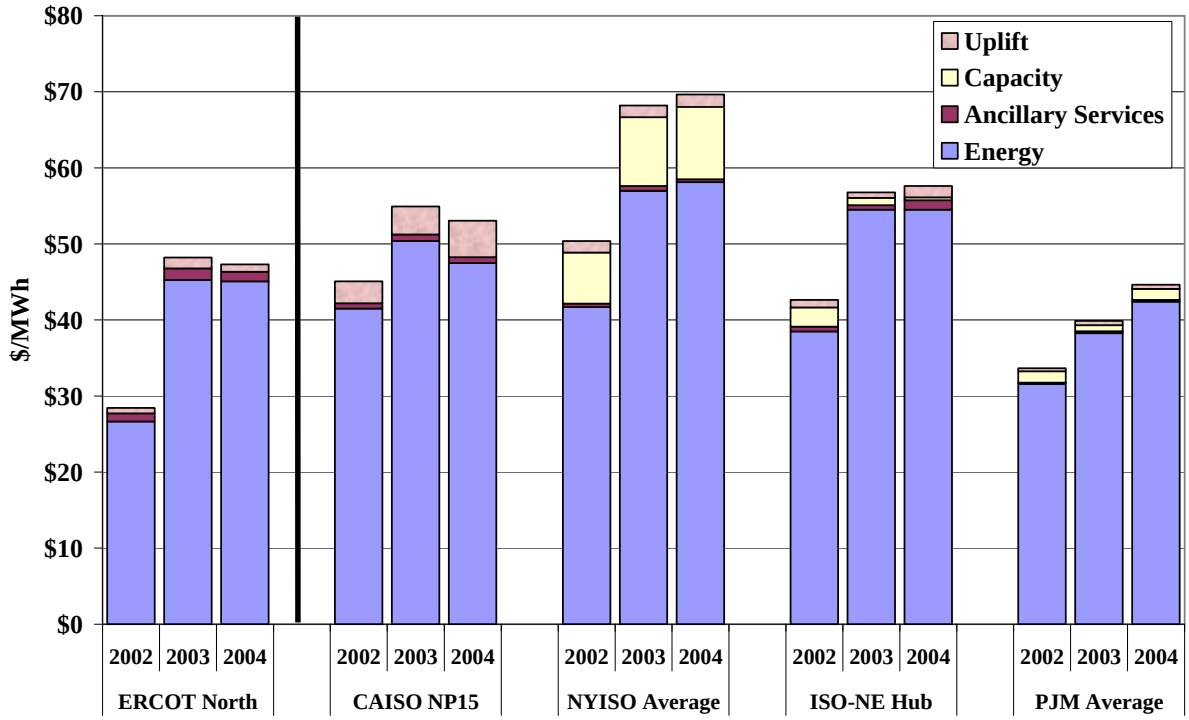
With the exception of February 2003, all-in prices are fairly stable from month-to-month. It is notable, however, that there is only a weak relationship between all-in prices and load levels. Energy prices have not risen significantly during the summer months, while ancillary services prices have actually gone down during the summer. The anomalous all-in prices in February were the result of electricity price spikes over three days (February 24-26) when prices rose as high as $990 per MWh in the balancing energy market. These price spikes occurred in response to a spike in natural gas prices and unusually high loads associated with a period of extremely cold weather.

Figure 2 indicates that natural gas prices were a primary driver of the trends in electricity prices from 2002 to 2004. 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 2003 by more than 65 percent from 2002 levels on average while the all-in price for electricity increased by 72 percent. Natural gas prices increased by an additional 5 percent in 2004 compared to 2003, leading to slightly higher energy prices in 2004. However, the all-in price for electricity decreased by 1 percent because ancillary services costs decreased by 17 percent. The decrease in ancillary services costs was primarily due to a decrease in the average procurement of up and down regulation. There was also a 32 percent reduction in uplift costs for resolving local congestion in 2004, which is discussed below.

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. Figure 3 compares the all-in prices for the five major centralized wholesale markets in the U.S.: (a) ERCOT, (b) California ISO, (c) New York ISO, (d) ISO New England, and (e) PJM. For each region, the figure reports the average cost (per MWh of load) for (a) energy, (b) ancillary services (regulation and reserves), (c) capacity markets, and (d) uplift for economically out-of-merit resources.

Figure 3: Comparison of All-In Prices across Markets 2002 to 2004



Each market experienced a substantial increase in energy prices from 2002 to 2003 due to increased fuel costs, but prices were comparable between 2003 and 2004. Although the markets vary substantially in the portion of their generating capacity that is fueled by natural gas, these units are usually on the margin and set the wholesale spot prices in the majority of hours for all markets shown.

In 2002, ERCOT exhibited the lowest all-in price -- 18 percent lower than the next lowest-priced market. In 2003 and 2004, the all-in price in PJM, which experienced the lowest increase in prices after 2002, was lower than in ERCOT. Natural gas-fired generation is on the margin less frequently in PJM than any of the other markets because PJM has access to large quantities of coal-fired generation within PJM itself and in the Midwest. The all-in prices in the ERCOT region are relatively low due in part to its substantial resource margin.

Our next analysis of all-in prices (shown in Figure 4) indicates how the market costs vary by ERCOT zone.

Figure 4: Average All-In Price of Electricity by Zone 2002 to 2004

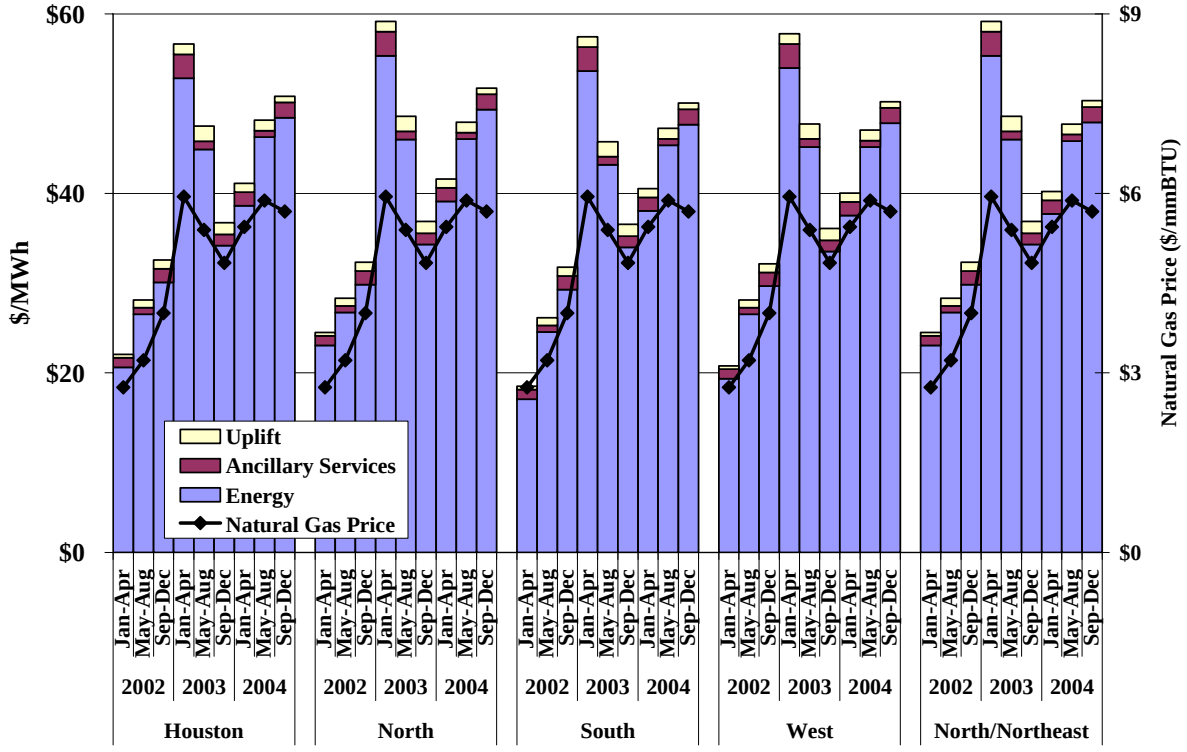
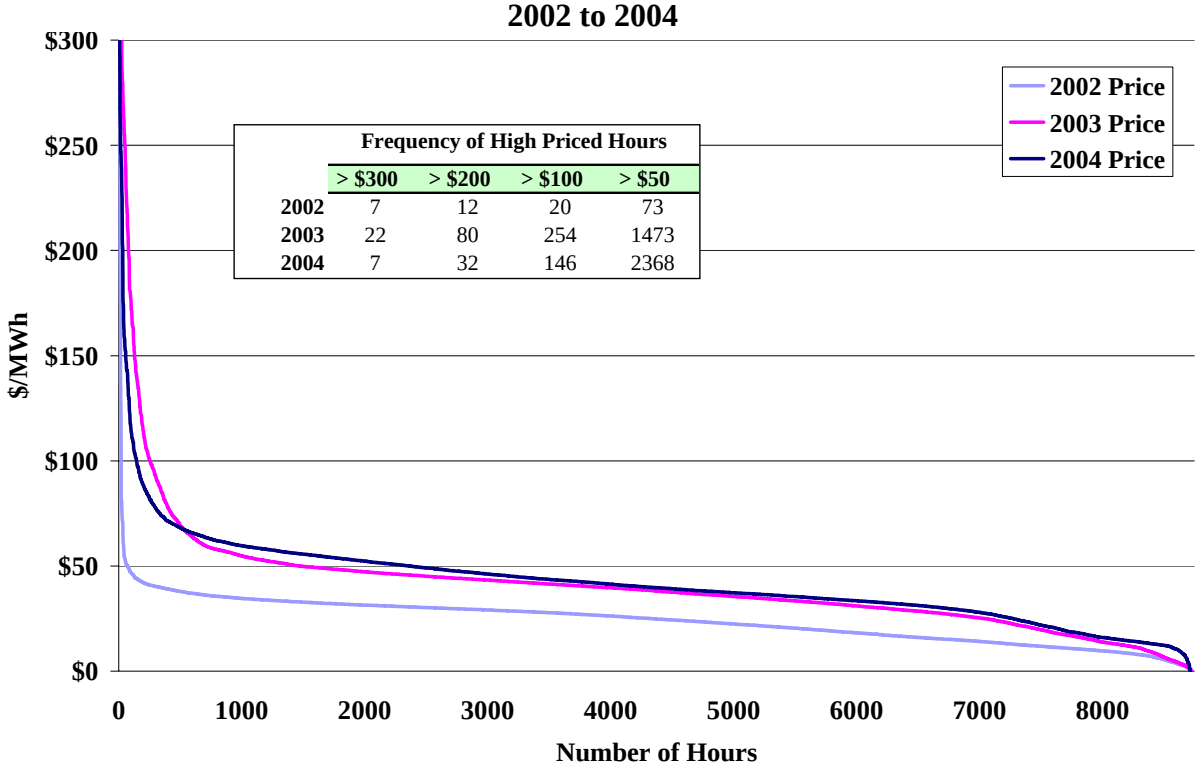


Figure 4 shows that there is relatively little difference in prices between zones. The largest interzonal price differences occurred at the beginning of 2002 before ERCOT began to directly assign the costs of interzonal congestion. The figure also shows that the uplift costs were comparable in size to reserves and regulation costs during all three years. Uplift costs were typically higher than reserves and regulation costs during the summer months and lower during other periods.

Figure 5 presents price duration curves for the balancing energy market in 2002, 2003, and 2004. A price duration curve indicates the number of hours that the price is at or above a certain level. The prices in this figure are hourly load-weighted average prices for the ERCOT balancing energy market.

Figure 5: ERCOT Price Duration Curve
2002 to 2004



Frequency of High Priced Hours				
	> $300	> $200	> $100	> $50
2002	7	12	20	73
2003	22	80	254	1473
2004	7	32	146	2368

This figure shows that prices were relatively low in 2002, exceeding $50 in only 73 hours. In contrast, almost 1,500 hours in 2003 and 2,400 hours in 2004 exhibited prices higher than $50. This clearly illustrates the effect of higher fuel prices, which increased electricity prices in 2003 and 2004 over a broad range of hours. This occurs because higher natural gas prices raise the marginal production costs of the generating units that set the prices in the balancing energy

market in most intervals. However, Figure 5 shows significant differences between 2003 and 2004 balancing energy market prices that are not explained by average fuel price levels. In 2004, there were nearly 60 percent more hours when prices were above $50 than in 2003. However, 2003 showed significantly more price spikes when prices exceeded $100.

To better observe the highest-priced hours, Figure 6 shows a narrower set of data that focuses on the highest-priced five percent of hours. The prices in these hours play a significant role in providing economic signals to invest in new and retain existing generation.

Figure 6: Price Duration Curve
Top Five Percent of Hours – 2002 to 2004

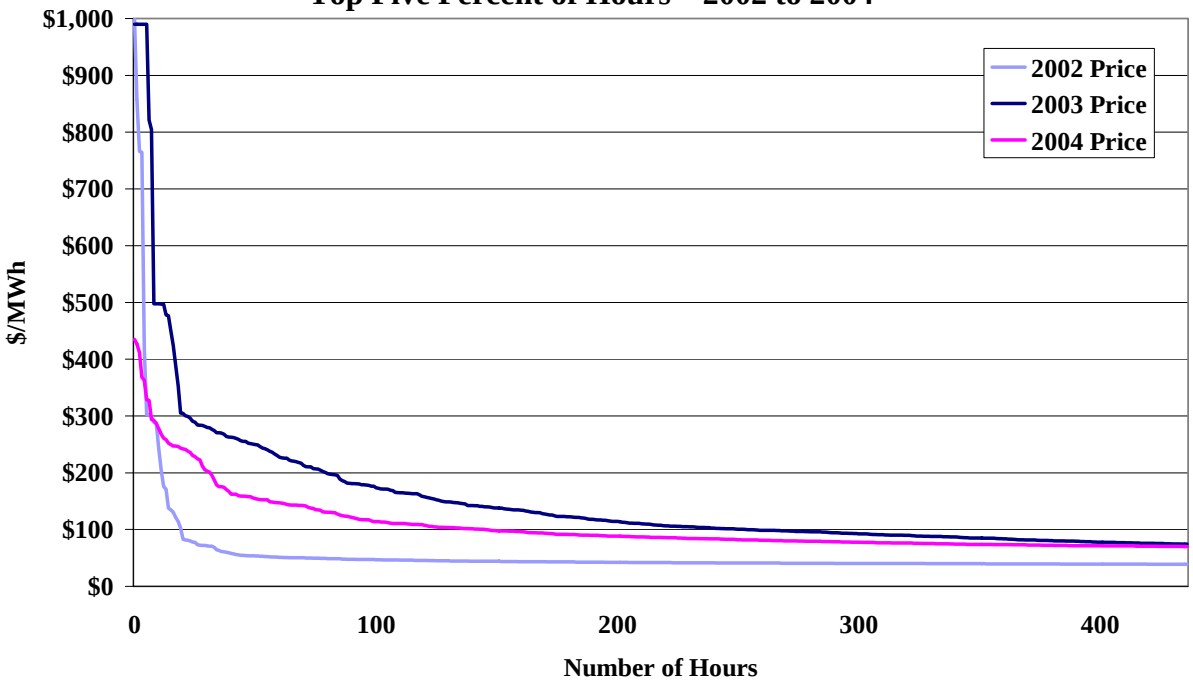
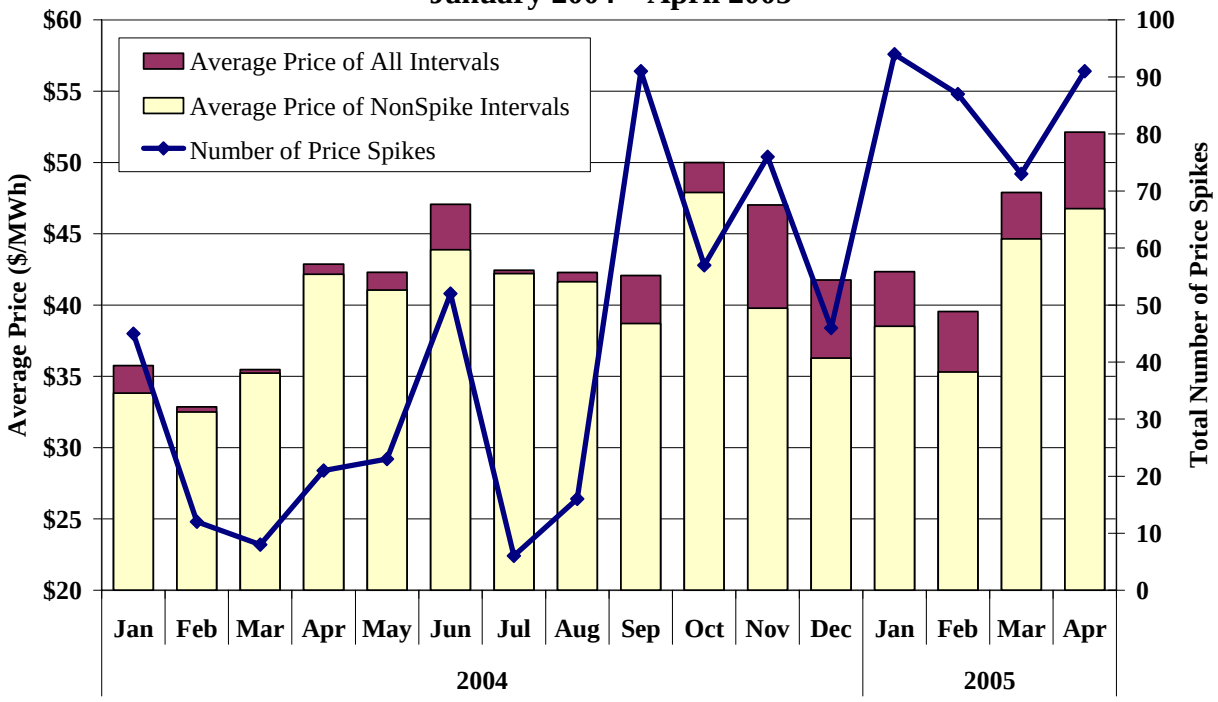


Figure 6 shows a more significant difference between 2003 and 2004 in the highest-priced hours than in all other hours. In 2004, there were only 146 hours with prices over $100 per MWh and only 32 hours with prices over $200 per MWh. In contrast, prices in 2003 exceeded $100 per MWh in 254 hours and exceeded $200 per MWh in 80 hours. Normally, one would expect the highest-priced hours to occur during the summer peak-demand conditions. However, in 2004 only 26 percent of the hours when price exceeded $100 occurred during the summer months (June through September). Furthermore, 46 percent of the hours above $100 occurred during the period between October 27 and December 8, 2004. In general, most of the price spikes during 2004 occurred under moderate load conditions and could not be explained by natural gas prices.

The fact that these prices occurred despite a substantial excess of generating capacity in ERCOT raises issues regarding the efficiency and competitiveness of the balancing energy market that are examined below.

In our final analysis of balancing energy prices, we show average prices and the number of price spikes in each month. 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). This analysis is shown in Figure 7, and includes four months in 2005.

Figure 7: Average Balancing Energy Prices and Number of Price Spikes
January 2004 – April 2005



As the figure shows, the number of price spikes increased sharply after August, 2004. There were 183 price spike intervals during the first eight months of 2004. The number of price spike intervals more than tripled to 615 during the subsequent eight months. 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 $1 per MWh during the first eight months and more than $4 per MWh during the latter period. Even though price spikes account for a small portion of the total intervals, they have a significant impact on overall price levels.

Price spikes in the markets for ancillary services have also risen significantly over this period. During the first eight months of 2004, there were 45 price spike hours for regulation up, 20 for regulation down, and 59 for responsive reserves. However, from September 2004 through April 2005, the number of price spike hours rose dramatically to 303 for regulation up, 412 for regulation down, and 217 for responsive reserves.[9] 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 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 lower cost resources are not efficiently utilized, it raises costs to consumers and sends inefficient economic signals.

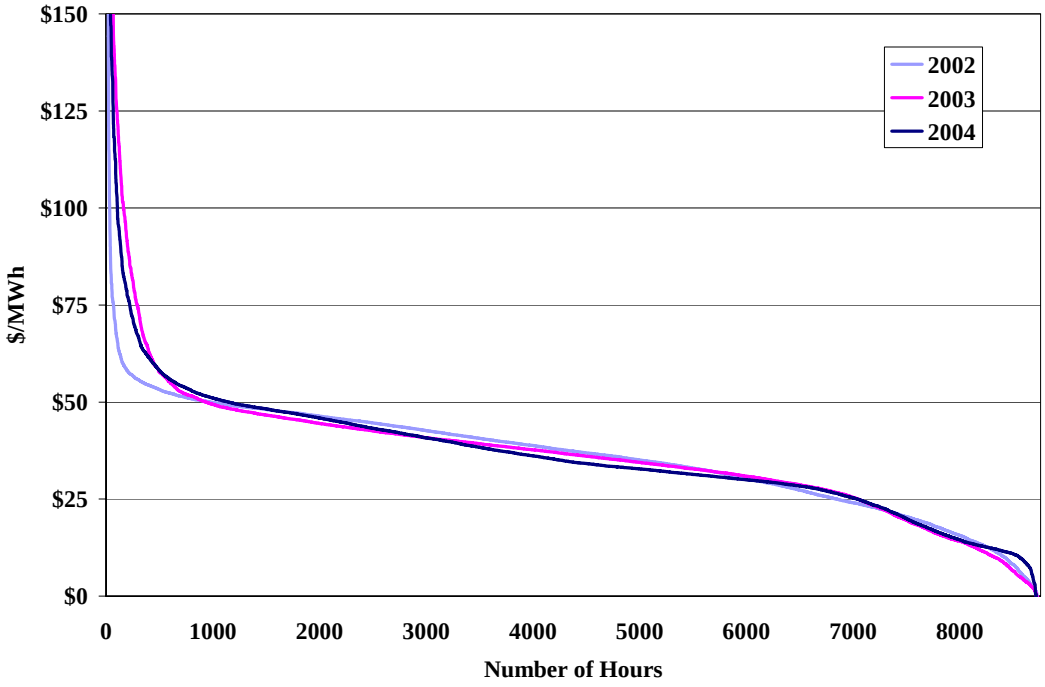
2. Fuel Price-Adjusted Balancing Energy Prices

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 isolate the effects of factors other than fuel prices, we produce two of the figures shown above in this sub-section adjusted for changes in natural gas prices. To do this, we divide the electricity price by the natural gas price (i.e., an “implied heat rate”) and multiply to $5 per MMBTU. In other words, the results in the figures below show electricity prices

[9] Price spikes are defined as hours where the price exceeds a threshold of $40 per MW for regulation up and regulation down, and $30 per MW for responsive reserves.

adjusted to reflect a fixed $5 per MMBTU gas price.[10] The first figure shows a revised version of the price duration curves shown above for 2002 to 2004.

Figure 8: Fuel Price-Adjusted Price Duration Curve
2002 to 2004

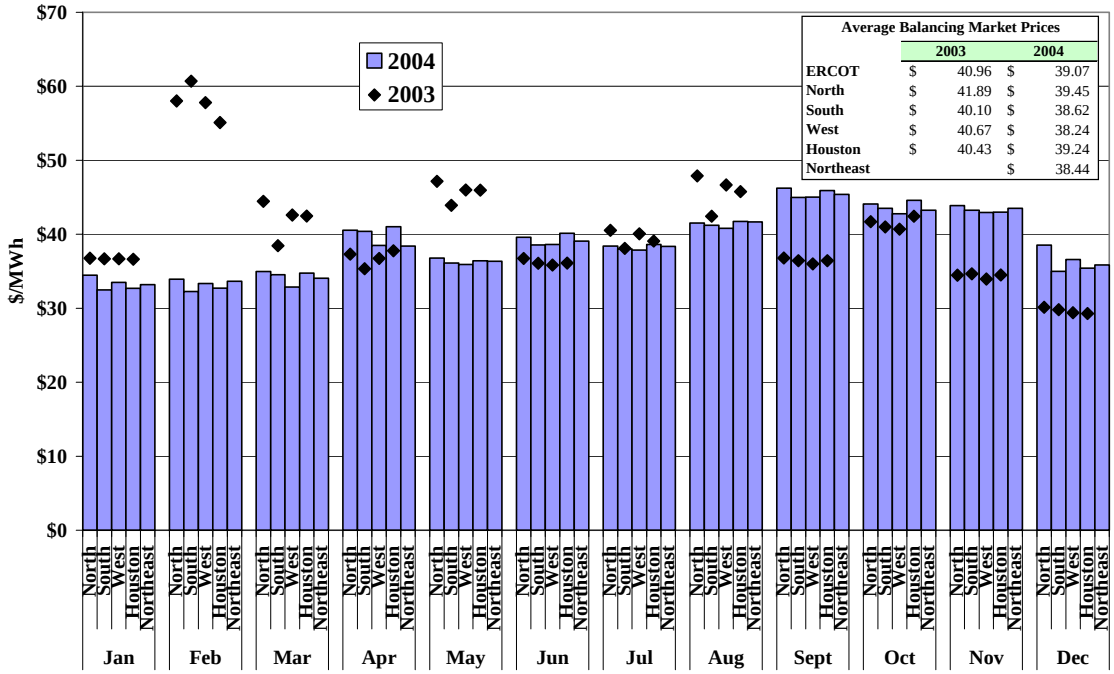


In contrast to Figure 5 above, Figure 8 shows that the fuel-adjusted prices in the three years are comparable. The unadjusted results showed that 2002 with significantly lower-priced, which was primarily due to the lower natural gas prices that prevailed in 2002. The differences that had existed in the highest-priced hours continue in the fuel price-adjusted results because these differences are not due to changes in natural gas prices.

The next figure shows the fuel price-adjusted prices on a monthly basis in each of the ERCOT zones from 2003 to 2004. This figure is the fuel price-adjusted version of Figure 1 in the prior sub-section.

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

Figure 9: Average Balancing Energy Market Prices
2003 & 2004



Average Balancing Market Prices

	2003	2004
ERCOT	$ 40.96	$ 39.07
North	$ 41.89	$ 39.45
South	$ 40.10	$ 38.62
West	$ 40.67	$ 38.24
Houston	$ 40.43	$ 39.24
Northeast		$ 38.44

The changes resulting from the fuel-price adjustment are not obvious in this figure (compared to Figure 1). The two most notable changes are in September and February. The relatively low natural gas prices in September 2004 cause the adjustment to increase the price in that month relative to the other months in 2004, causing September to be the highest-priced month. This occurs because price spikes occur most frequently in September, as shown in Figure 7 above.

The second noticeable change in this figure is the result for February 2003. The prices in this month decrease from an unadjusted average of approximately $90 per MWh to an adjusted value of approximately $60 per MWh. This adjustment likely overestimates the effect of the fuel prices on the electricity prices because the price spikes that occurred during the cold snap in February 2003 were only partially related to the high natural gas prices during that period. Nevertheless, the figure continues to show that the cold snap caused average prices in February 2003 to exceed the average prices in all other months in 2003 and 2004.

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. One important step taken to address the first condition (i.e., to reduce barriers between the markets), was the implementation of relaxed balanced schedules in November 2002. By allowing QSEs to increase and decrease their purchases in the balancing energy market, they should be better able 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.[11]

It should be noted, however, that the current balancing energy market does not reveal the full value of energy in the ERCOT market. Intrazonal constraints associated with "local congestion" are not reflected in balancing energy prices, which tends to undervalue energy in locally-constrained areas. Instead, these congestion costs are borne in market-wide uplift charges that cannot be hedged through forward energy contracts. Hence, neither the balancing energy prices nor the forward energy prices will include the costs of managing local congestion.

There are several ways to measure the degree of price convergence between forward and real-time markets. In our analysis, we measure two aspects of convergence. The first method investigates whether there are systematic differences in prices between forward markets and the

[11] The volatility in balancing energy prices, which became more prevalent in 2003 and 2004, creates risk that may cause some participants to be willing to pay a premium to purchase energy in the bilateral markets and should, therefore, result in a premium in the bilateral market prices above the balancing energy prices over time.

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 calculate the difference between the average forward price and the average balancing energy price in 2002, 2003, and 2004. This measures 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 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 Table 1 for 2002 to 2004.

Table 1: Convergence Between Forward and Real-Time Energy Prices 2002 to 2004

		Average price of power on weekdays from 6 AM to 10 PM		Convergence (as a Percent)	
		Day-Ahead Price	Balancing Energy Price	Avg. Day-Ahead minus Avg. Balancing	Average Difference
2002	**All Days**	$29.06	$26.57	9%	17%
2003	**All Days**	$46.56	$48.96	-5%	27%
	Excluding Feb. 24 & 25	*$46.21*	*$45.06*	*2%*	*20%*
2004	**All Days**	$48.95	$51.07	-4%	18%
	January to August	*$49.04*	*$48.58*	*1%*	*13%*
	September to December	*$48.78*	*$54.29*	*-11%*	*26%*

Note: Day-Ahead Price based on Megawatt Daily peak day-ahead prices when five or more trades were reported.

The table shows the statistics for all hours on weekdays between 6 AM and 10 PM. These are the peak hours that are commonly traded in the forward market. The much lower volumes in the off-peak hours make the forward prices for these hours much less reliable. For 2002, the Table indicates that there was a 9 percent premium in the day-ahead prices relative to the balancing energy prices. In 2003 and 2004, prices were generally higher and there was a closer correspondence between day-ahead and real-time prices in percentage terms. This is likely due, in part, to the introduction of relaxed balanced schedules near the end of 2002 that increased participants' flexibility to arbitrage prices between the day-ahead forward market and the balancing energy market, as discussed above.

Although forward market prices generally converge with spot market prices, unexpected spot market events can result in large systematic differences between forward and spot prices. In 2003, forward prices were 5 percent lower than balancing energy market prices, but if February 24th and 25th are excluded, forward prices would have been significantly closer and actually 2 percent higher. In 2004, there was a 1 percent price premium in the forward market from January to August. However, between September and December, the frequent unanticipated price spikes contributed to the balancing energy prices exceeding the forward prices by 11 percent.

Several factors explain the unexpected rise in prices at the end of 2004 and are discussed in subsequent sections. These include, primarily:

- A reduction in the level of on-line and quick-start capacity relative to energy and ancillary services demand;
- High balancing energy offers by TXU;
- Failure of participants to fully offer their balancing energy capability in the balancing energy market; and
- The multi-step congestion management process for jointly managing local and inter-zonal congestion in ERCOT.

The last two factors point to significant issues related to the design and operation of the ERCOT markets. These issues are analyzed and discussed in detail in a study we performed last fall.[12] This study finds that most of these issues would most fully be resolved through the implementation of the nodal energy markets being considered in ERCOT. Apart from implementing nodal markets, the study also provides fourteen recommended changes to the existing markets to improve their operation and resolve some of these issues. Notwithstanding these potential changes, market participants should improve over time in their ability to recognize these factors that can lead to sharp price increases in their expectations, so that forward prices are not persistently higher or lower than balancing energy market prices.

Table 1 also shows that the average absolute price difference increased from 17 percent in 2002 to 27 percent in 2003, before decreasing in 2004 to 18 percent. As noted above, the average absolute difference measures volatility. It can capture wide price movements that are missed in a simple difference in the average prices. The general rise in the frequency of price spikes after 2002 has made balancing energy market prices more volatile, and thus inherently more difficult to predict. Taking into account the rise in volatility, convergence was actually better in 2004 than in 2002.

The results in this section indicate that the effectiveness of the ERCOT market in achieving convergence between the day-ahead bilateral prices and the balancing energy prices has generally improved over time, although the volatility of the balancing energy market and the unexpected high prices at the end of 2004 have undermined price convergence.

[12] "2004 Assessment of the Operation of the ERCOT Wholesale Electricity Markets", Potomac Economics, November 2004, (hereinafter "Market Operations Report").

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 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 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 10 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 10: Average Quantities Cleared in the Balancing Energy Market 2002 to 2004

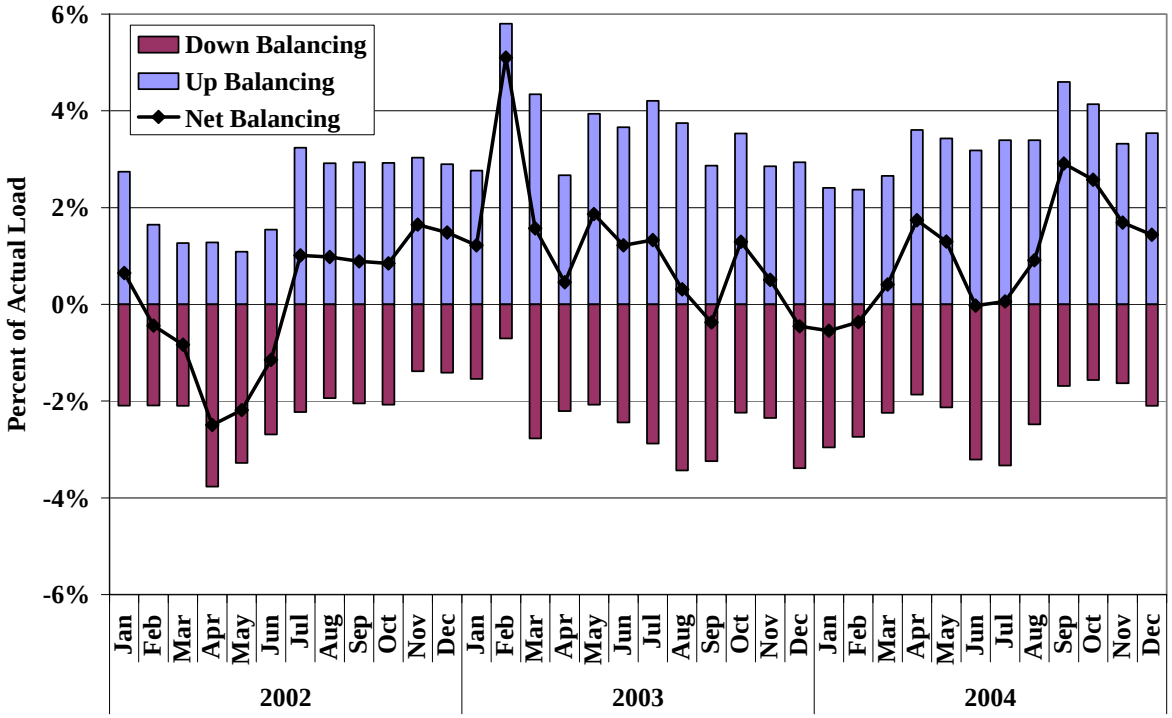


Figure 10 shows that the total volume of balancing up and balancing down energy as a share of actual load increased from an average of 4.6 percent in 2002 to 6.1 percent in 2003 and 5.7 percent in 2004. Thus, there was a general increase in trading through the balancing energy market after 2002. In addition, participants have generally been net purchasers of balancing energy rather than net sellers. Hence, they generally schedule less than their full load and rely on the balancing energy market to satisfy the remaining unscheduled load. One factor that influenced these patterns is the implementation of relaxed balanced schedules in November 2002.

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 increasingly operate like a centralized energy spot market and has contributed to improved price convergence, although the

increase in the volume of energy traded through the balancing energy market is not as large as some expected.

Figure 10 also shows that the largest quantity of net up balancing energy sales for any single month occurred in February 2003. From February 24th to March 6th, load serving entities purchased an average of 2,465 MW of net balancing up energy, approximately 7.8 percent of load. This high level of purchases was due in part to natural gas curtailments and generation outages that compelled some of the load-serving entities to turn to the balancing energy market to purchase additional energy to serve load. These factors were identified by Wholesale Market Oversight (formerly, the Market Oversight Division or "MOD") in a report that focused on the most extreme portion of this period, from February 24 to 26.[13]

Aside from the introduction of relaxed balanced schedules in 2002, another reason the balancing energy quantities have increased is 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.

The second aspect of the increase in trading volume that is important is the increase in *net* balancing energy quantities. 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. One reason this can occur is to arbitrage the forward energy and balancing energy markets. Figure 10 shows that the average monthly net balancing energy volume has fluctuated significantly over the period, although it has been positive in most months in 2004.

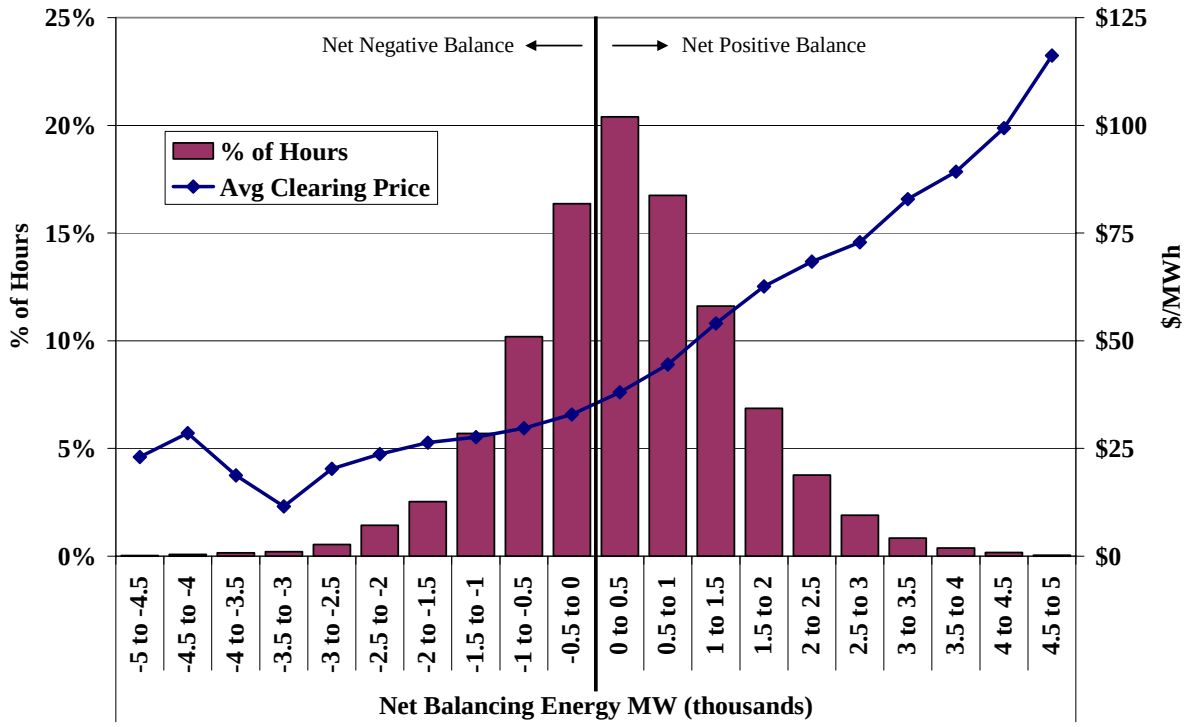
If large hourly under-scheduling or over-scheduling occurs suddenly, the balancing energy market can lack the ramping capability and sometimes the volume of energy offers necessary to achieve an efficient outcome. In these cases, large net balancing energy purchases can lead to

[13] Public Utility Commission of Texas, Market Oversight Division, "Market and Reliability Issues Related to the Extreme Weather Event on February 24-26, 2003," report filed in Project Number 25937 (May 19, 2003.

transient price spikes when excess capacity exists but is not available in the 15-minute time frame 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 11 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.

Figure 11: Magnitude of Net Balancing Energy and Corresponding Price 2004



Each bar in Figure 11 shows the portion of the hours during 2004 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) in over 20 percent of the hours in 2004.

Figure 11 shows a relatively symmetrical distribution of net balancing energy purchases centered between 0 and 0.5 gigawatts. This is consistent with Figure 10 which showed that there were substantial net balancing up quantities on a monthly average basis in 8 of the 12 months in 2004. Figure 11 also shows that approximately 64 percent of the hourly observations show net purchases or sales between -1.0 gigawatts and 1.0 gigawatts.[14] Hence, there were many hours when the net balancing energy traded was relatively low, indicating that in many hours the total scheduled energy is close to the actual load.

The line plotted in Figure 11 shows the average balancing energy prices corresponding to each level of balancing energy volumes. In an efficiently functioning spot market, there should be little relationship between the balancing energy prices and the net purchases or sales. Instead, one should expect that prices would be primarily determined by more fundamental factors, such as actual load levels and fuel prices. However, this figure indicates a relatively clear relationship, showing the balancing energy prices increasing as net balancing energy volume increases. This provides an indication that the balancing energy market is thinly traded, which can undermine its efficiency. We analyze the potential reasons for this apparent relationship in the next sub-section.

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 12 shows the average balancing energy price and the actual load in the peak hour of each weekday during 2004. The figure shows that the highest prices (e.g., greater than $100/MWh) do not reliably correspond to the highest load levels. Indeed, the clearing price was approximately $61 per MWh at the system peak, which is lower than on many other days. Likewise, prices throughout the summer were generally not positively correlated to peaks in load.

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

Figure 12: Daily Peak Loads and Prices

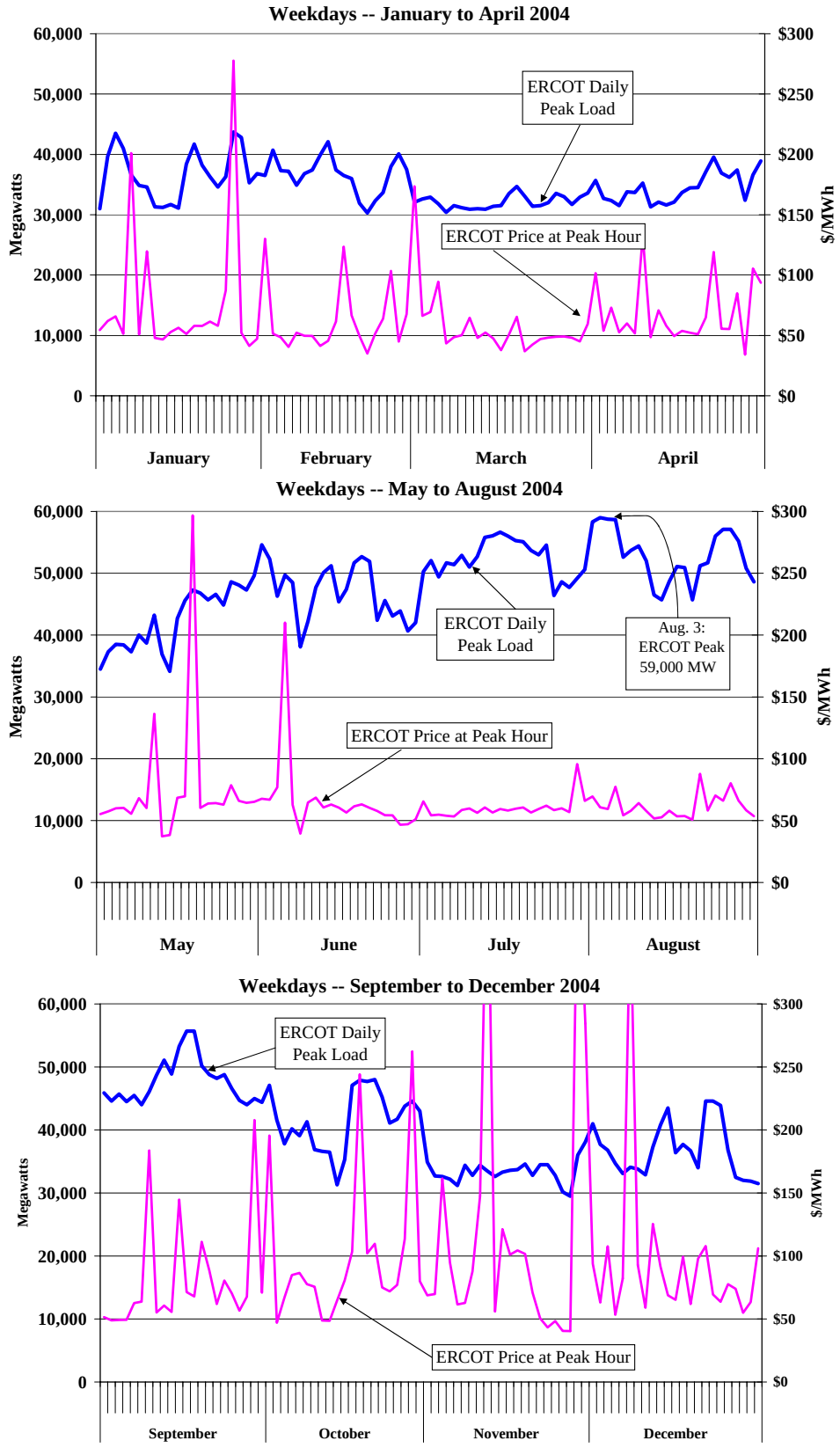


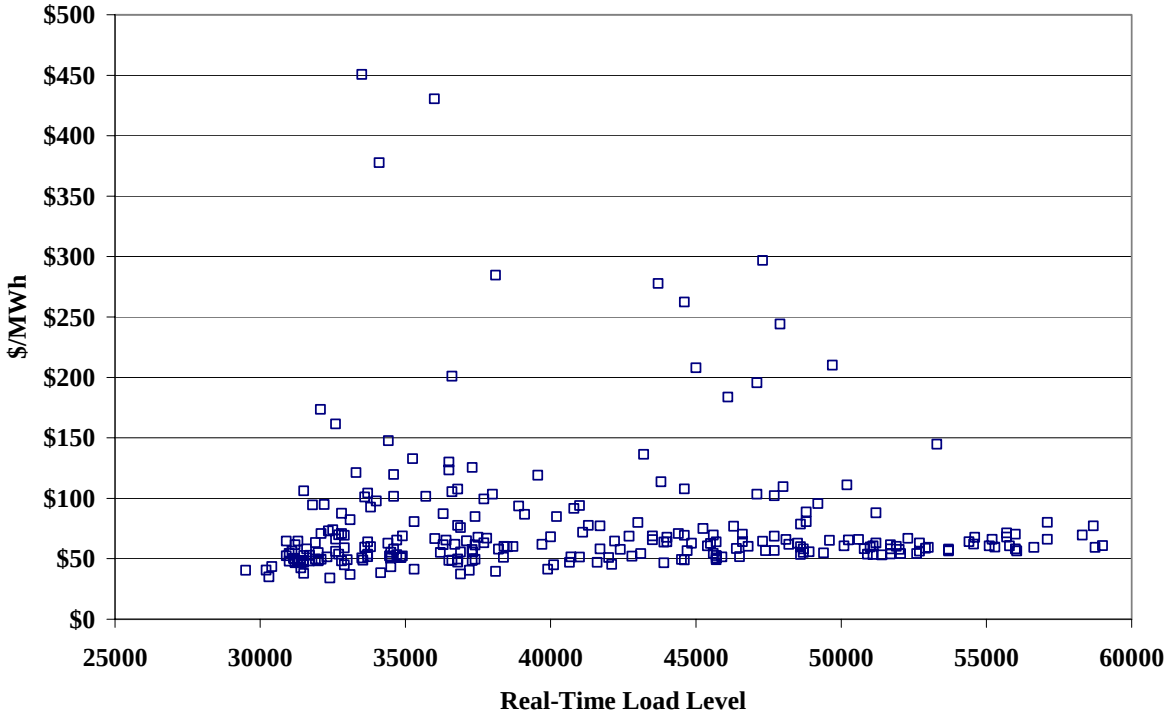
Figure 12 shows that on most days the average balancing energy price is between $50 and $60 per MWh during the peak load hour. Although there were a number of days when the price in the peak load hour rose above $100 per MWh, there was only one such day during high-load months of June to August. Indeed, the figure indicates little relationship between load levels and balancing energy prices.

In some cases, the relatively high prices occurred on the highest load days of the month. This was the case in April, for example, when high prices tend to be due to large portions of ERCOT's resources being out-of-service for maintenance. At other times, such as between September and December, price spikes exhibited little or no relationship to fluctuations in demand. As noted above, the price spikes that occurred during the late fall were the result of certain factors that we investigated in a separate report.[15] In that report, we found that the balancing energy offers of TXU, together with the lack of offers from other suppliers, caused the balancing energy price spikes.

To further examine the relationship between actual load in ERCOT and the balancing energy prices, Figure 13 shows the same data as Figure 12, but plots the average balancing energy prices versus the daily peak 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.

[15] "Investigation into the Causes of the Shortage of Energy in the ERCOT Balancing Energy Market, etc.," *Op Cit*.

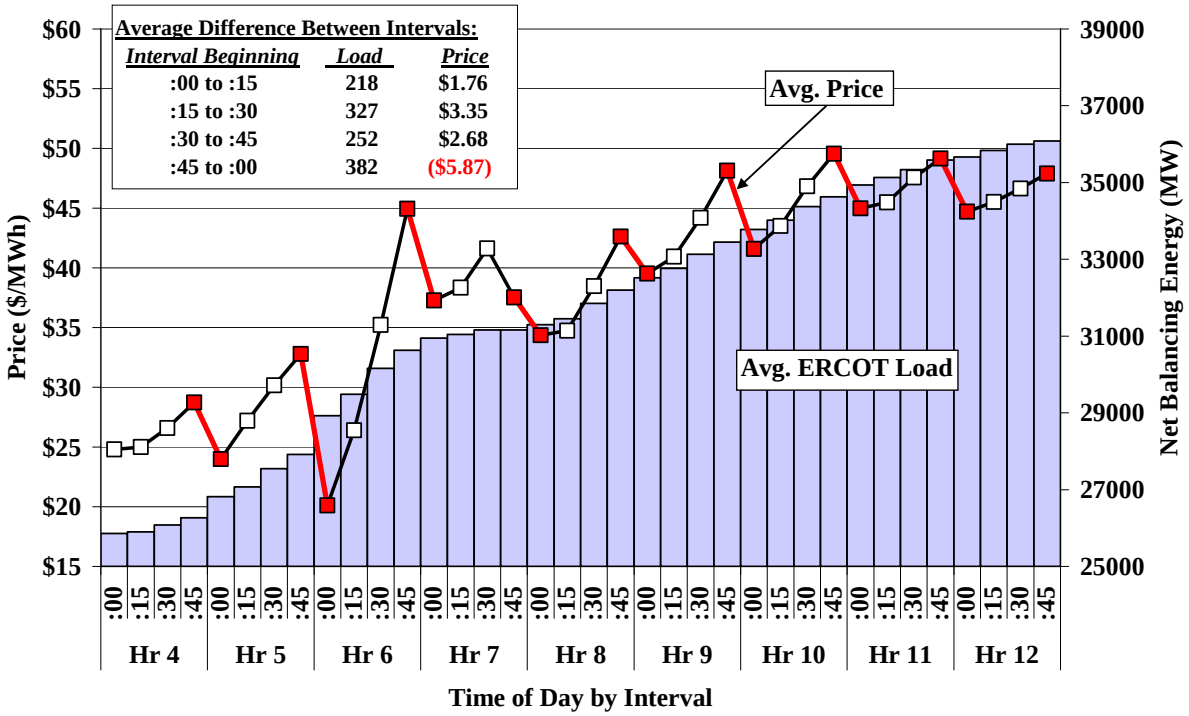
Figure 13: ERCOT Balancing Energy Price vs. Real-Time Load
Weekdays -- Peak Load Hour -- 2004



These prices are generally tightly clustered with very slight upward trend as load increases. However, if one examines the relatively high prices, i.e., those greater than $100 per MWh, there is little discernable relationship between these occurrences and the actual load in ERCOT. In fact, the majority of price spikes occur when load is less than 40 GW. Alternatively, the analysis shown above in Figure 11 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 26 GW to 36 GW. This usually occurs over a nine-hour period. Thus, the change in load averages 1,100 MW per hour (275 MW per 15-minute interval) during the morning and early afternoon. Figure 14 shows the average load and balancing energy price in each interval from 4 AM through 1 PM in 2004. The price is plotted as a line in the figure while the average load is shown with vertical bars.

Figure 14: Average Clearing Price and Load by Time of Day
Ramping-Up Hours – 2004



The figure shows that the load steadily increases in every interval and prices generally move upward from about $25 per MWh at 4:00 AM to $47 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 14 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 -$5.87 per MWh. This occurs because participants tend to change their schedules once per hour, bringing on additional supply 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 1,600 MW per hour (400 MW per 15-minute interval) during the late evening. Figure 15 shows this decrease in load by interval, together with the average balancing energy prices for the intervals from 9 PM to 3 AM.

Figure 15: Average Clearing Price and Load by Time of Day
Ramping-Down Hours – 2004

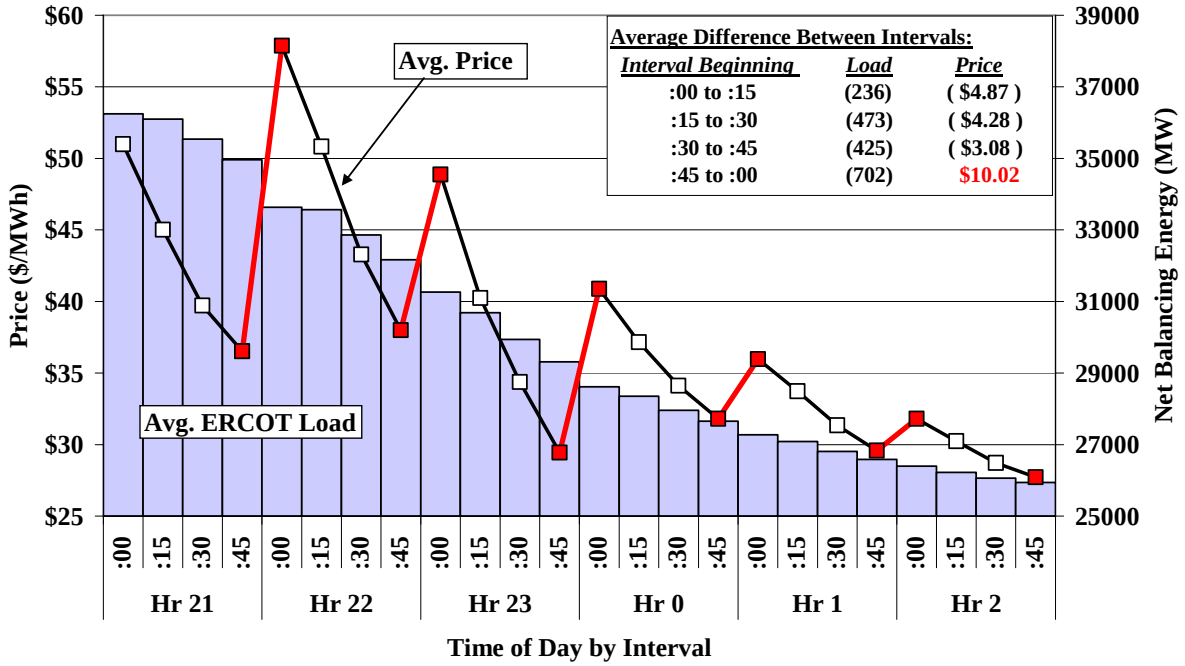


Figure 15 shows that while balancing energy prices decrease over these intervals, they follow a similar pattern as exhibited in the ramping-up hours. The balancing energy price decreases in each interval of the hour before rising substantially in the first interval of the following hour. The balancing energy price increases by an average of $10.02 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.

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

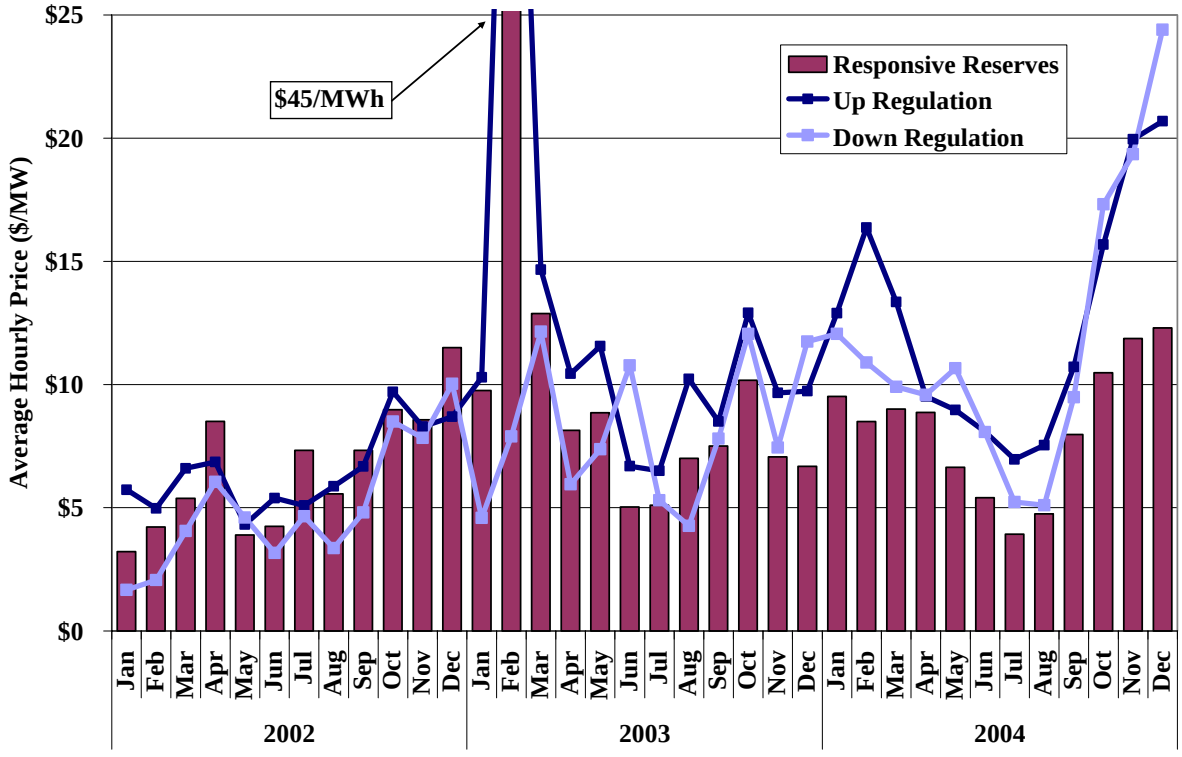
B. Ancillary Services Market Results

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 reserves and regulation markets in 2004.

1. Reserves and Regulation Prices

Our first analysis in this section provides a summary of the ancillary services prices over the past three years. Figure 16 shows the monthly average ancillary services prices between 2002 and 2004. Average prices for each ancillary service are weighted by the quantities required in each hour.

Figure 16: Monthly Average Ancillary Service Prices 2002 to 2004



This figure shows that ancillary services prices were generally higher in 2003 and 2004 than in 2002. Much of this increase can be attributed to the increase 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, although these expected costs are likely to be lower than the costs of providing up regulation. This is consistent with the prices in all three years, although the average premium on up regulation was reduced in 2004.

The figure also shows that the prices for up regulation exceeded prices for responsive reserves in all months of 2003 and 2004. This is consistent with expectations because a supplier must incur the same 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. This pricing relationship between responsive reserves and up regulation was not consistent in 2002 -- five of the twelve months in 2002 exhibited responsive reserves prices that were higher than the up regulation prices. This improvement in price consistency since then has occurred despite the fact that regulation requirements were reduced in 2003 and again in 2004, which would tend to reduce regulation prices.

Figure 16 also shows that reserves and regulation prices tend to be lower during the summer than at other times of the year. This is because the required quantities of reserves and regulation are relatively constant over the year while the supply of resources that can provide reserves and regulation (i.e., on-line capacity not scheduled for energy) tends to increase in proportion to load. The additional supply puts downward pressure on reserves and regulation prices. In some other markets, this effect is outstripped by the increase in prices during the summer, which causes ancillary services prices to rise during the summer. However, the ERCOT market has not exhibited significantly higher balancing energy prices during the summer peak load conditions.

Other than February 2003, the highest-priced periods for ancillary services shown in Figure 16 occurred at the end of 2004. This happened for two reasons. First, there was an increase in the frequency of price spikes in the balancing energy market during this period that raised the opportunity costs of providing ancillary services. Second, demand was relatively low so that less

online capacity was committed and unscheduled for energy, making it available to provide ancillary services.

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 2 shows such an analysis, comparing the average prices for responsive reserves and non-spinning reserves over the past three years in those hours when ERCOT procured non-spinning reserves. It also shows average prices for 2002 without April 29 and 30 when prices ranged above $990 for 13 hours. These two days are excluded from the table because they tend to obscure the overall price relationship between the two services.

Non-spinning reserves were purchased in approximately 18 percent of the hours during 2002, 25 percent of hours during 2003, and 24 percent of hours during 2004. Like the relationship between regulation and responsive reserves prices, responsive reserves prices should exceed non-spinning reserves prices because responsive reserves are a higher quality of reserves. Resources capable of providing responsive reserves can also be used to provide non-spinning reserves, but the reverse is not true. Hence, the price for non-spinning reserves should never exceed the price of responsive reserves under an efficient market design.

Table 2: Responsive Reserves and Non-Spinning Reserves Prices 2002 to 2004

	2002*	2003	2004
Non-Spinning Reserve Price	$6.30	$9.82	$7.51
Responsive Reserve Price	$8.37	$10.73	$9.03

* Excludes April 29-30, 2002. Including these days, the average prices were $14.43 for Non-spin and $9.19 for Responsive.

Table 2 shows the expected relationship between average prices for responsive and non-spinning reserves. However, non-spinning reserves prices were still higher than responsive reserves prices in 17 percent of hours during 2002, 35 percent of hours during 2003, and 9 percent of hours during 2004. Although non-spinning reserves are a lower quality product than responsive reserves, these price reversals may not be as counterintuitive as they appear because providing non-spinning reserves may actually be more costly for some resources than responsive reserves.

When a resource providing reserves is actually deployed to produce energy, the deploying QSE is paid for the output at the balancing energy price. There is no guarantee that the balancing energy price will be higher than the cost of dispatching the resource. When the balancing energy price is lower, no additional payment is made to the QSE. In fact, it is likely most units providing reserves have production costs higher than the balancing energy price because these units are the lowest cost providers of reserves (these units incur no lost profits by not producing energy). Hence, these units will be running at a loss if they are deployed, and the risk of losses associated with reserve deployments should be included in the operating reserves offer prices by suppliers. The two determinants of the expected value of these losses are: (a) the average difference between the resource's production cost and the balancing energy price, and (b) the probability of being deployed for energy. It is the second factor that can cause the marginal cost of supplying non-spinning reserves (and hence the clearing prices for non-spinning reserves) to be higher than for responsive reserves.

In 2004, less than 0.1 percent of the responsive reserves were actually deployed, while 3.2 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, which could contribute to counterintuitive results in some hours. In general, the purpose of operating reserves is to protect the system against unforeseen contingencies (e.g., transmission line or generator outages), rather than for meeting load. The balancing energy market deployments in the 15-minute timeframe and regulation deployments in the 4-second timeframe are the primary means for meeting the load requirements.

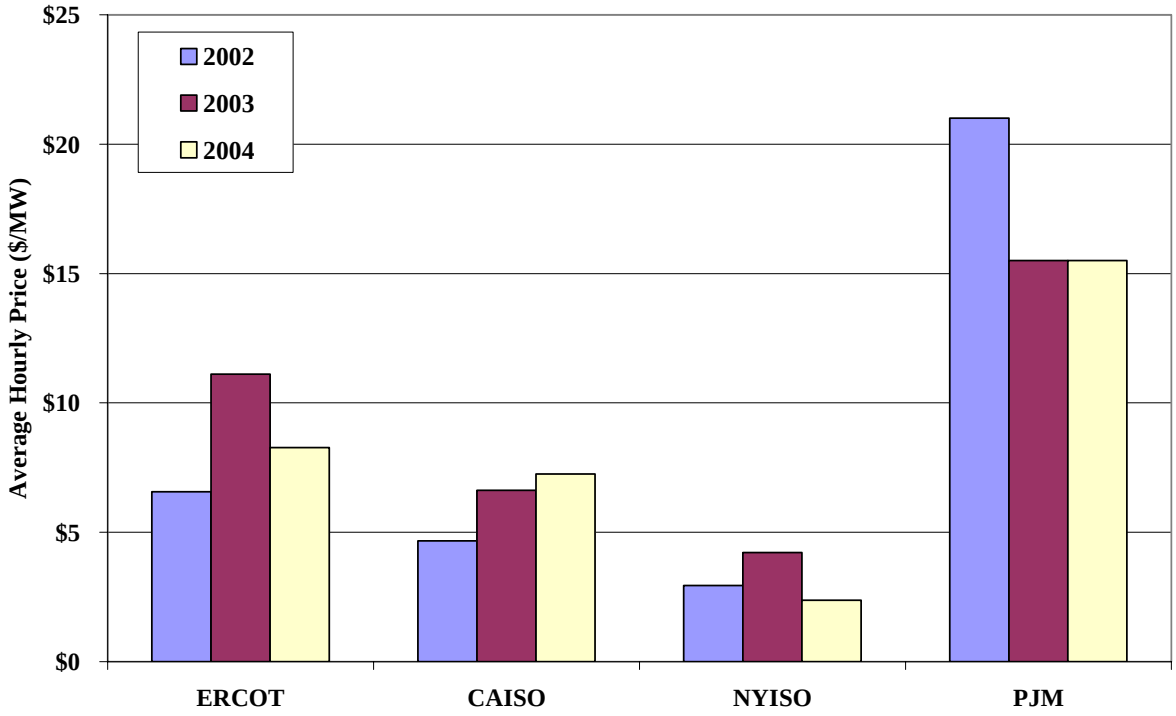
However, in cases when the resources in the balancing energy and regulation markets may not be sufficient to satisfy the energy demand while meeting the responsive reserve requirement, we understand that ERCOT will frequently procure and deploy non-spinning reserves. This process is a means for ERCOT to implement supplemental generator commitments to increase the supply of energy in the balancing energy market. While supplemental generator commitments can be necessary for a variety of reasons, this is not a typical or desirable use of an operating reserve market.

Ultimately, the objective in the long-run should be to jointly-optimize each of the ancillary services markets with the energy market. In a market where ancillary services are jointly optimized with energy, the marginal cost of providing non-spinning reserves can never be higher than the marginal cost of providing responsive reserves. As in ERCOT, a jointly optimized market will deploy non-spinning reserves more frequently than responsive reserves because responsive reserves are more critical for reliability and are therefore more valuable. However, when non-spinning reserves are deployed in the context of a jointly optimized market, there is no risk that the clearing price will be insufficient for these units to recover their production costs since they will contribute to setting the energy prices. A jointly-optimized market recognizes the energy offer prices for all resources that are dispatched.

ERCOT plans to modify the procurement process for ancillary services in July 2005, so that the markets for regulation, responsive reserves, and non-spinning reserves will clear simultaneously. This change is likely to result in increased prices in the responsive reserve market to reflect the higher marginal costs of providing non-spinning reserves. Since the costs of providing non-spinning reserves may be partly attributable to the deployment procedures discussed above which do not co-optimize ancillary services with the balancing energy market, it will be particularly important to consider potential improvements to these procedures.

Responsive reserves prices declined in 2004 to about $8/MW from the relatively high levels in 2003. Figure 17 shows how the annual average prices in ERCOT from 2002-2003 compare to the responsive reserve prices in the California, PJM, and New York wholesale markets. This figure shows that the responsive reserve prices in ERCOT were somewhat higher than comparable prices in California and New York, but lower than prices in PJM. Only 2003 and 2004 prices are shown for PJM, which instituted a market for spinning reserves in December 2002.

Figure 17: Responsive Reserves Prices in Other RTO Markets 2002 to 2004



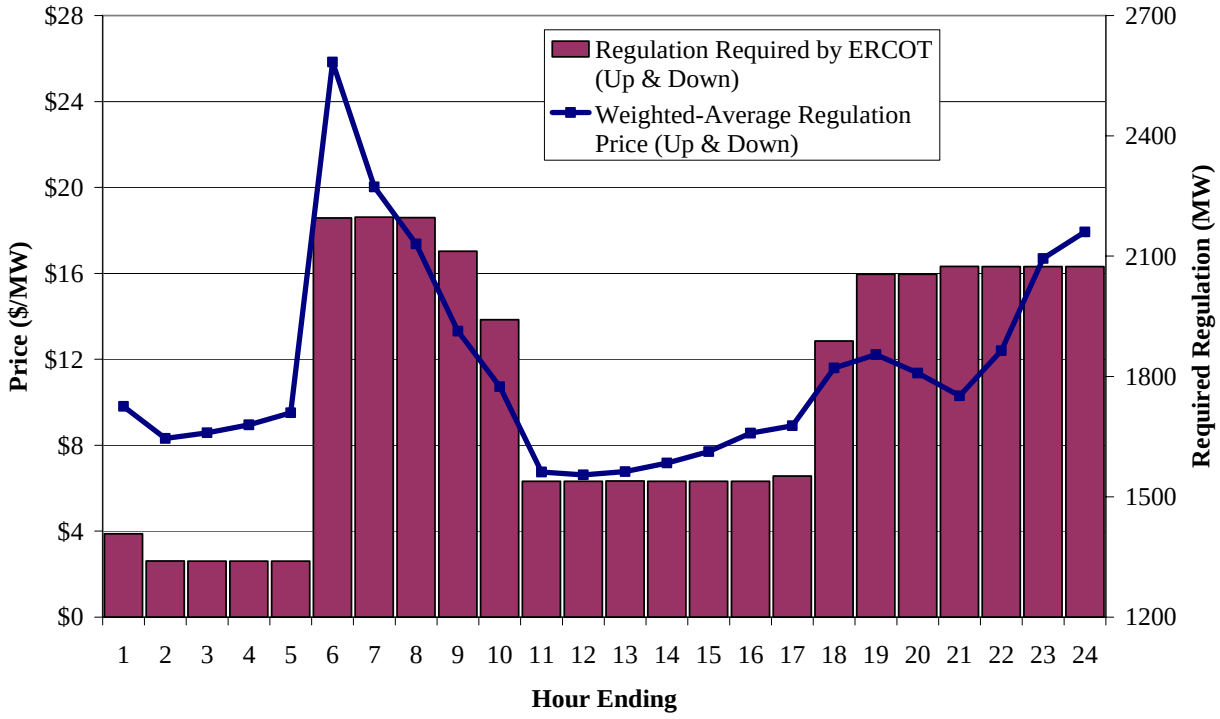
There are a number of reasons why the responsive reserve prices in ERCOT are higher than prices in some of the other regions. First, ERCOT procures substantially more responsive reserves relative to its load than New York, which satisfies a large share of its operating reserve requirements with non-spinning 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 does not jointly-optimize ancillary services and energy markets, like in California and PJM. 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, more regulation is 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 generally needs approximately 50 percent more regulating capacity during ramping hours. When demand rises, higher-cost resources must be employed and prices should increase.

Figure 18 shows the relationship between the quantities of regulation demanded by ERCOT and regulation price levels. This figure compares regulation prices to the average regulation quantity (both up and down regulation together) 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.

Figure 18: Regulation Prices and Requirements by Hour of Day 2004



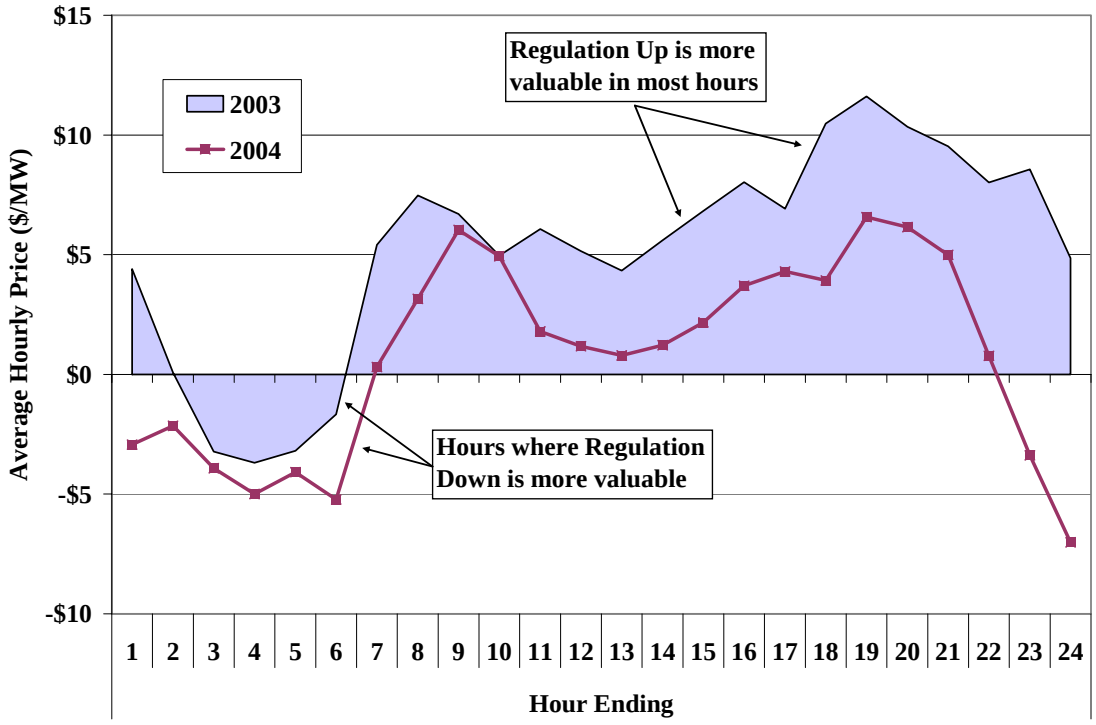
This figure shows that ERCOT requires approximately 1,300 MW of capability prior to the initial ramping period (beginning at 6 AM). The requirement then jumps up to 2,200 MW during the steepest ramping hours from 6 AM to 9 AM. The requirement declines to 1,500 MW during the late morning and afternoon 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 2,100 MW. On average, the quantities of regulation required by ERCOT in 2004 were 200 to 300 MW lower than in 2003. This has helped reduce the total costs of ancillary services.

Figure 18 indicates that average regulation prices are closely correlated with the regulation quantities purchased. During non-ramping hours, such as overnight, late morning, and in the afternoon, regulation prices average from $7 to $9 per MW. During the ramping hours in early morning and evening, average regulation prices range from $10 to $26 per MW. The higher prices during ramping hours also occur because a larger portion of regulation capability is actually deployed during ramping hours. Additionally, the price range exhibited in the ramping hours was wider in 2004 than in 2003. This is largely due to the regulation prices that occurred in the last quarter of 2004, when lower levels of excess online generating capacity and increased balancing energy market volatility resulted in higher and more volatile regulation prices.

Regulation prices are particularly high during hours ending 6, 7, 23, and 24. Less supply is available during these hours, because many regulation-capable units in ERCOT start after 7 am and shutdown before 10 pm. This reduces the amount of capacity available to supply regulation, which leads to higher prices.

While up and down regulation are relatively close substitutes and are generally supplied from the same resources, ERCOT runs separate regulation markets reflecting the fact that the marginal costs of providing up and down regulation can differ substantially. Like the comparative analysis of responsive reserve and non-spinning reserve prices presented earlier in this sections, our next analysis examines the differences between up and down regulation prices to determine whether they exhibit a rational pattern that is consistent with market fundamentals. Figure 19 shows the average up regulation price minus the average down regulation price in each hour of the day for 2003 and 2004 separately.

Figure 19: Comparison of Up Regulation and Down Regulation Prices
Up Regulation Price Minus Down Regulation Price



The figure reveals a distinct intertemporal variation in the regulation price differences. The opportunity costs associated with providing regulation helps explain the inter-temporal pattern of regulation prices. Down regulation prices tend to rise during the off-peak hours—when energy prices are low and there is greater risk that cost will exceed price when a generator is operating above its minimum output level. This is because suppliers of down regulation must operate sufficiently above minimum output levels so they have the ability to reduce output when called on to regulate down in real time. In addition, the overall supply of down regulation is lowest in the early morning hours because fewer units are online and they are operating at relatively low operating levels. Alternatively, up regulation is most expensive during the peak hours when the potential opportunity costs of not producing energy are the highest.

Figure 19 also shows a significant downward shift in the price difference from 2003 to 2004, which means that up regulation became less expensive relative to down regulation. The price difference becomes largest during periods of acute capacity shortage. For example, the difference between up regulation and down regulation prices was $37 per MW during February 2003. This explains a large portion of the downward shift in Figure 19 from 2003 to 2004. In

addition, the general reduction in regulation procured during 2004 would reduce the prices for both up and down regulation, as well as the difference between the prices for the two services.

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 20. This figure summarizes the quantities of ancillary services offered and ancillary services self-arranged relative to the total capability and the typical demand for each service.

The bottom segment of each bar in Figure 20 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.

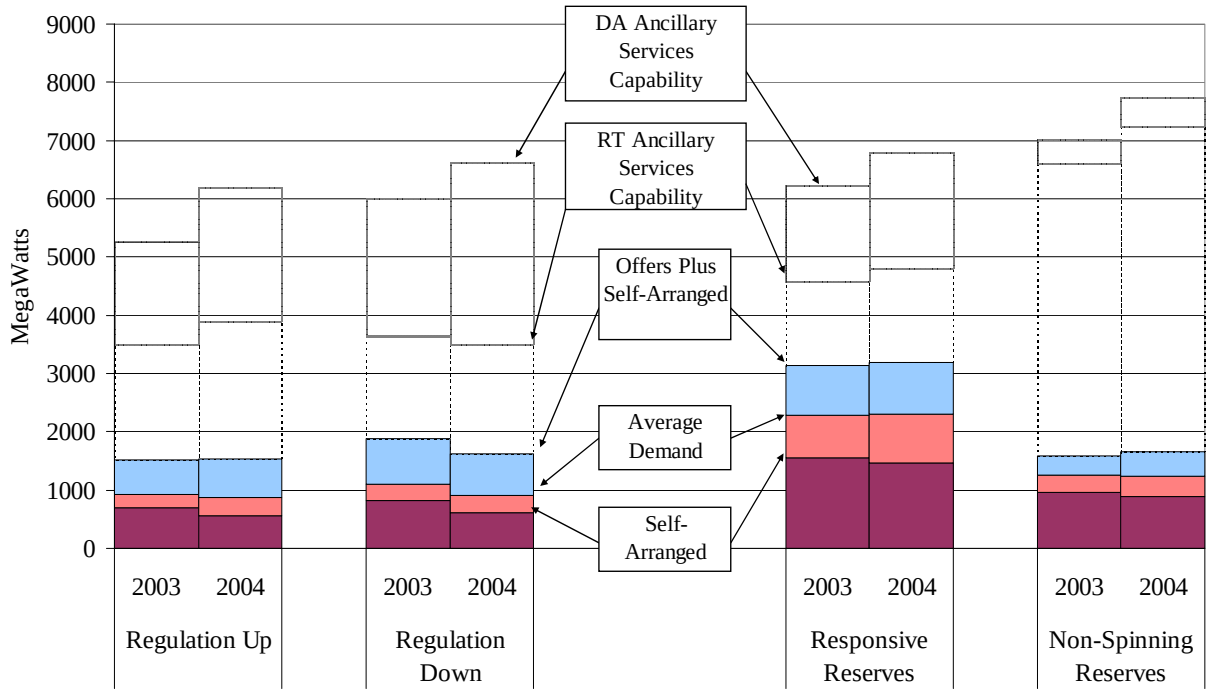
The third segment of each bar is the quantity offered into the auction market that is not cleared. Therefore, the sum of the second and third segments is the total quantities offered in each ancillary services auction on average, including the quantities cleared and not-cleared. The empty segments correspond 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).

The capability shown in Figure 20 incorporates ERCOT's requirements and restrictions for each service type. 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 20 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 1,000 MW of the demand for responsive reserves was satisfied by Loads acting as Resources ("LaaRs"). However, 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 20 includes the capability of units that QSEs indicate are able to ramp-up in thirty minutes and able to start-up on short notice. However, it should be noted that any on-line resource with available capacity can provide non-spinning reserves, so the actual capability is larger than shown in the figure. 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 20: Reserves and Regulation Capacity, Offers, and Schedules 2003 & 2004



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

Figure 20 shows that less than one-half of each type of ancillary services capability was offered during 2003 and 2004. 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.

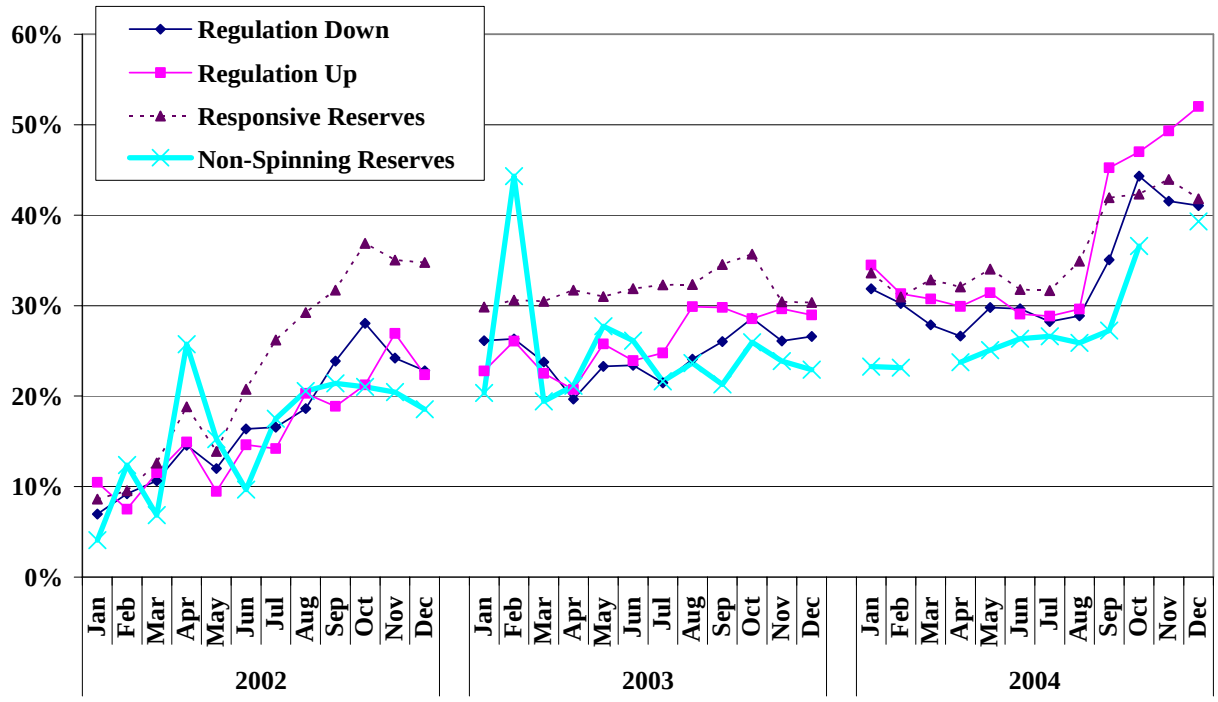
In addition, participants may not offer the capability of resources they do not expect to commit for the following day. This explanation is less likely because suppliers could submit offer prices high enough to ensure that their costs of committing additional resources to support the ancillary services offers are covered. Nonetheless, there is a substantial quantity of reserves that remain available in real time, but are 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. The analysis in this section is not sufficient to make that determination given that there are multiple factors that may be contributing to these offer patterns.

Figure 20 shows modest increases in the amount of day-ahead ancillary services capability between 2003 and 2004. The installation of several gigawatts of new capacity during 2003 and 2004 has contributed to this overall increase, although continued mothballing and retirement of certain units in 2004 mitigated the increase. The figure also indicates a small growth between 2003 and 2004 in regulation up and responsive reserves capability in real-time, but a slight decline for regulation down. This indicates that resources have more headroom on average in 2004 and, therefore, more capability to move up rapidly. The rise in responsive reserves capability is also attributed to a steady increase in the amount of demand response capability in the form of LaaRs.

Finally, Figure 20 shows that a relatively high share of these services is 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 21 shows the share of each type of ancillary service that is purchased through the ERCOT market.

Figure 21: Portion of Reserves and Regulation Procured Through ERCOT 2002 to 2004



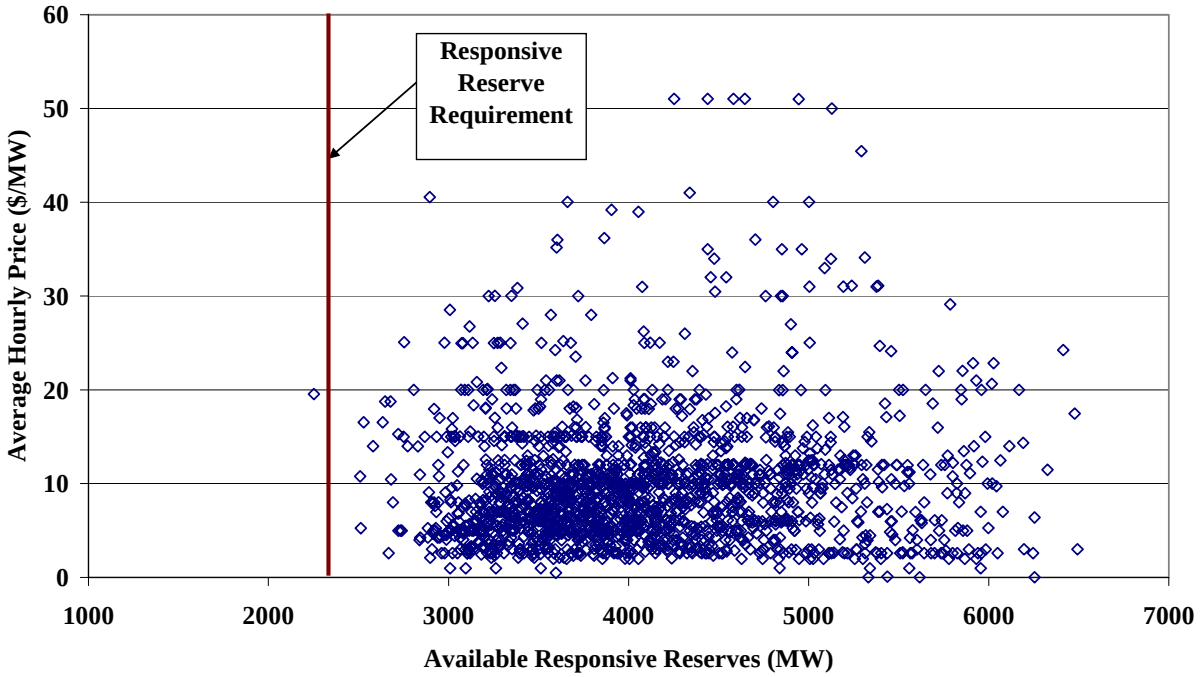
This figure shows that purchases of all ancillary services from the ERCOT markets have increased consistently over the past three years, rising sharply after the summer of 2004. As noted earlier, there was a significant rise in balancing energy prices during the fall of 2004, and since ancillary services providers must forego energy sales, this has likely increased the opportunity cost of providing ancillary services. When buyers of ancillary services face higher bilateral contract prices, it can push more of their purchases into the ERCOT market until prices between the two markets converge.

We expect that as market participants gain more experience with the ERCOT markets, larger portions of the available responsive reserves and regulation capability will be offered into the market, thereby increasing the market's liquidity and reducing ancillary services prices. Based on the results shown in Figure 21, this appears to be the case. Jointly-optimizing the reserves and energy markets would serve to increase the liquidity of these markets further by reducing the economic costs of selling ancillary services under the current sequential market design.

The final analysis in this section evaluates the prices prevailing in the responsive reserves market during 2004. Prices in this market are significantly higher than in other markets that co-optimize the dispatch of energy and responsive reserves. Lower prices occur in co-optimized markets because 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.

Hence, Figure 22 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. The figure also shows the responsive reserves requirement of 2,300 MW to show the amount of the surplus in each hour.

Figure 22: Hourly Responsive Reserves Capability vs. Market Clearing Price Afternoon Peak Hours – 2004

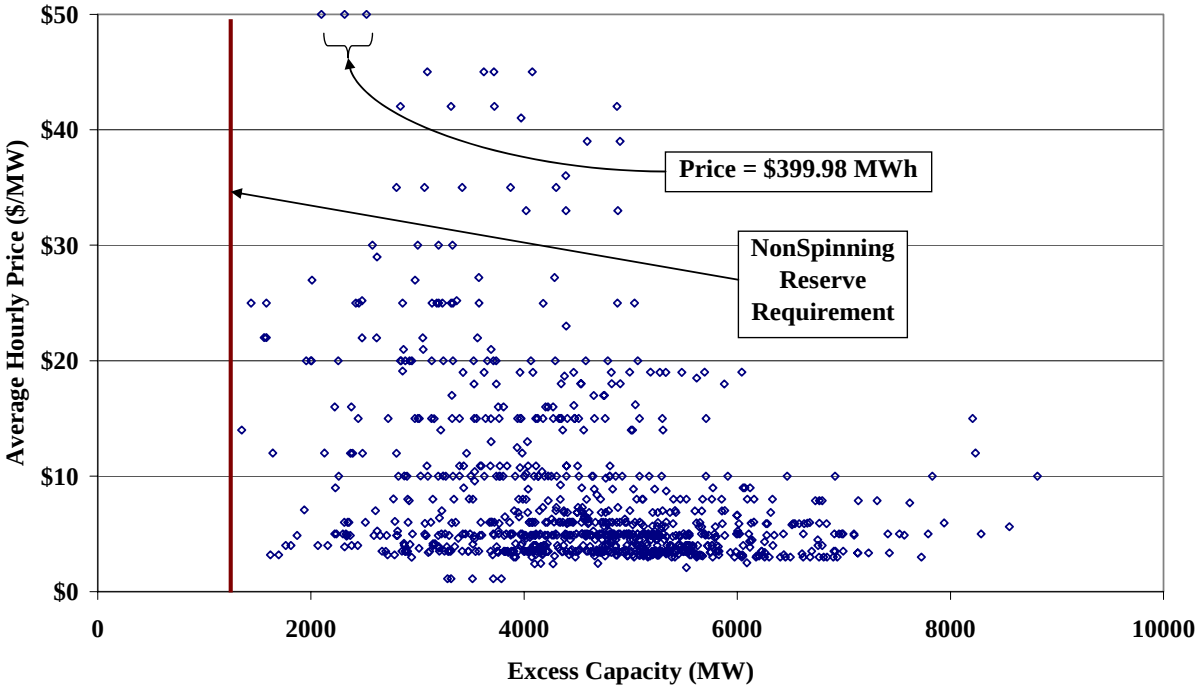


This figure indicates a somewhat random pattern of responsive reserves prices in relation to the hourly available responsive reserves capability in real time. In a well functioning-market for responsive reserves, we would expect excess capacity to be negatively correlated with the clearing prices, but this was not the case in 2004. Although a slight negative relationship existed in 2003, the dispersion in prices in both years raise significant issues regarding the performance of this market. Particularly surprising is the frequency with which the price exceeds $10 per MW when the available responsive reserves capability is more than 2,000 MW higher than the requirement. In these hours, the marginal costs of supplying responsive reserves should be zero. These results reinforce the potential benefits promised by jointly optimizing the operating reserves and energy markets, which we would recommend in the context of the alternative markets designs currently under consideration.

Non-spinning reserves are purchased on an as-needed basis whenever ERCOT predicts a balancing energy shortage at least one hour in advance. Non-spinning reserves are resources that can be brought on-line 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 23 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 2004.

Figure 23 shows that there were more than 2,000 MW of excess capacity capable of providing non-spinning reserves in virtually every hour when it was purchased. Although not obvious from the scatter plot, prices are negatively correlated to the non-spinning reserves capability, which should be expected. However, the dispersion of prices is wide. Again, the lack of co-optimized markets for energy, regulation, and reserves may be a primary contributing factor to the high prices for non-spinning reserves when there are large quantities of excess capacity available.

Figure 23: Hourly Non-Spinning Reserves Capability vs. Market Clearing Price Afternoon Peak Hours – 2004



C. Net Revenue Analysis

Net revenue is defined as the total revenue that can be earned by a generating unit less its variable production costs. Hence, it is the revenue in excess of short-run operating costs and is available to recover a unit's fixed and capital costs. Net revenues from the energy, operating reserves, and regulation markets together provide the economic signals that 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:

a. New capacity is not needed because there is sufficient generation already available;

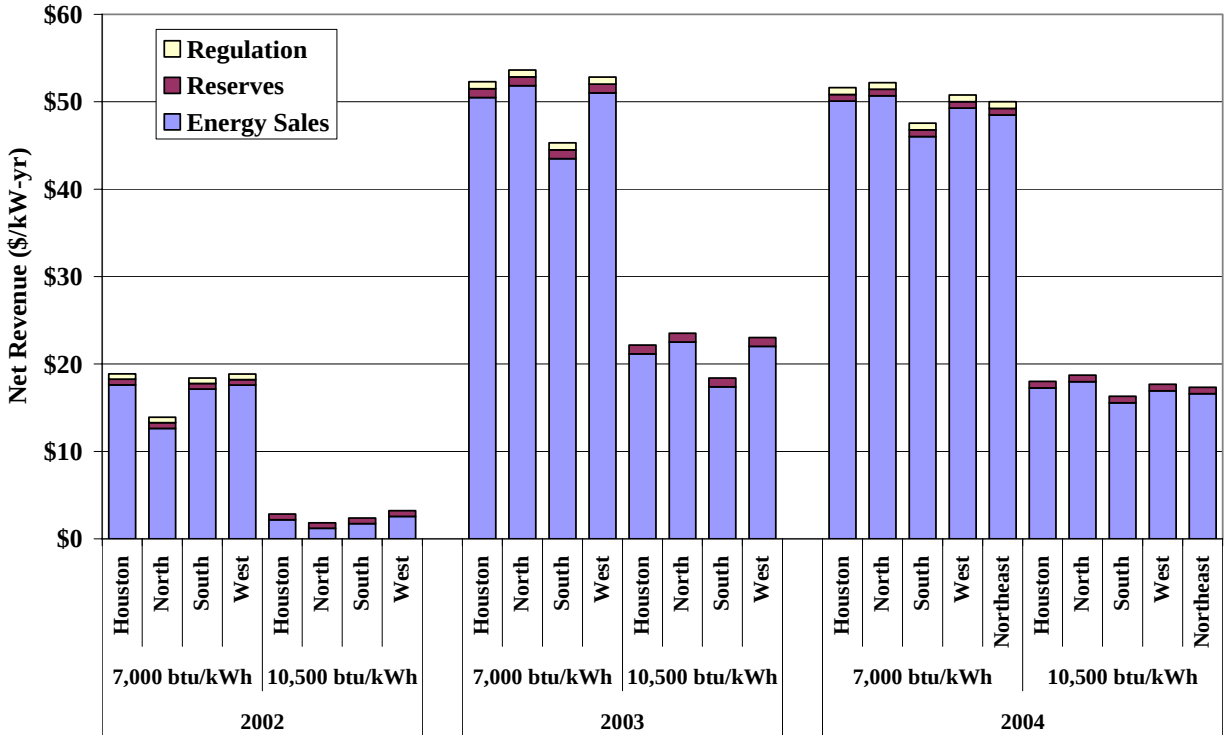
b. Load levels, and thus energy prices, are temporarily below long-run expected levels (this could be due to mild weather or other factors); or

c. Market rules are causing revenues to be reduced inefficiently.

Likewise, the opposite would be true if prices 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 2004 by various types of generators in each zone.

Figure 24 shows the results of the net revenue analysis for two types of units. The first type is a gas combined-cycle (with an assumed heat rate of 7,000 BTU/kWh). The second type is a gas turbine (with an assumed heat rate of 10,500 BTU/kWh). 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. 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 24: Estimated Net Revenue 2002 to 2004



The revenues in Figure 24 are reduced to account for the assumed outage rate for each unit. For purposes of this analysis, we assume the heat rates cited above for each unit, $4 per MWh

variable operating and maintenance costs, and a total outage rate (planned and forced) of 10 percent. Some units, generally those in unique locations that are used to resolve local transmission constraints, also receive a substantial amount of revenue through uplift payments (i.e., Out-of-Merit Energy, Out-of-Merit Capacity, and Reliability Must Run payments). This source of revenue is not considered in this analysis.

Figure 24 shows that the estimated net revenue was significantly higher in 2003 and 2004 than in 2002. This is largely because the rise in natural gas prices led to substantially higher energy prices. While the higher natural gas prices also lead to higher costs for these new units, these units are more efficient than the resources that set prices in some hours. Therefore, an increase in natural gas prices can increase the margin for the new units and result in higher net revenue.

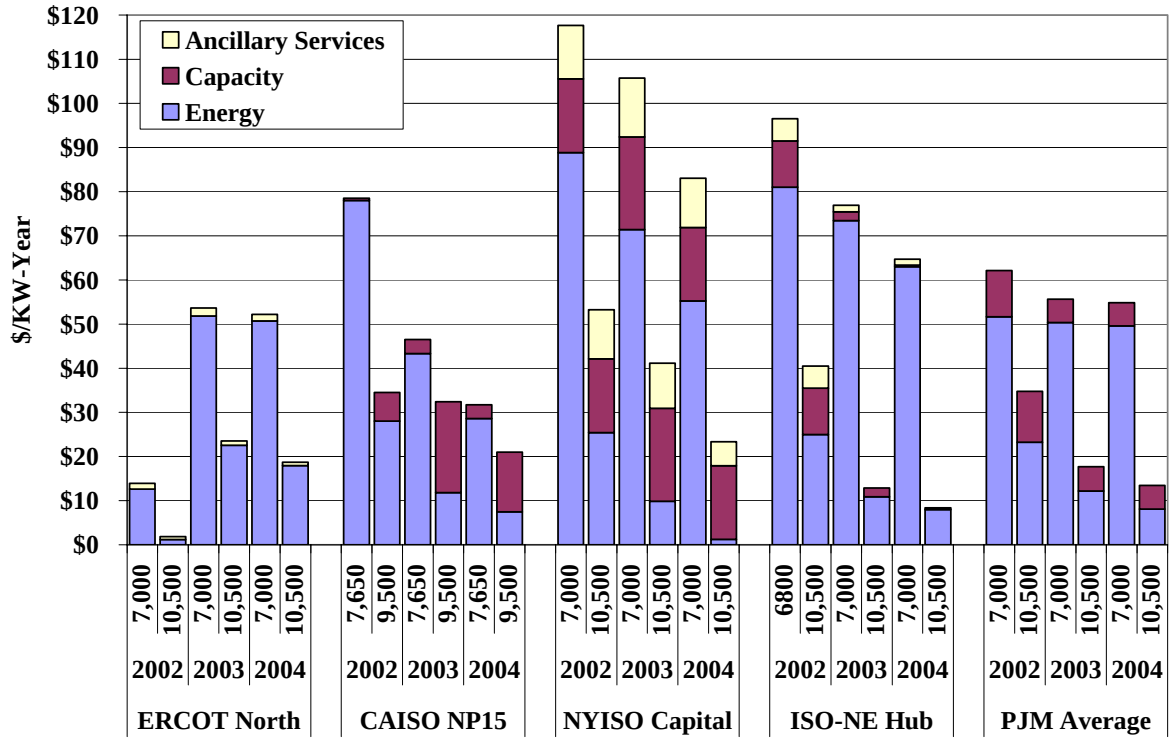
Importantly, there were also a much larger number of relatively high-priced hours after 2002, which contributed to the higher net revenues. These higher prices frequently occurred under moderate load conditions when one would not expect high prices. We have determined that these prices are due to scheduling and ramping issues under the current market design and to the fact that a large share of the available balancing energy capability is not offered in the balancing energy market. These issues are analyzed in Section II below. In addition, balancing energy offers of TXU during fall of 2004 contributed to the high prices.

Although net revenues were substantially higher in 2003 and 2004, neither type of new generating unit would have earned sufficient net revenue to make the investment profitable. Based on our estimates of investment data for new units, the net revenue required to satisfy the annual fixed costs (including capital carrying costs) of a new gas turbine unit is $70 to $80 per kW-year. For a new combined cycle unit, net revenue requirements are more that $100 per kW-year. Although the net revenue increased considerably from 2002 to 2004, it remained at less than half of the amount necessary to support new investment in 2004. Hence, the net revenue of both types of hypothetical new units would need to increase by $50 to $60 per kW-year to be profitable. This is not surprising given the surplus of capacity that currently exists in ERCOT. However, net revenue should increase as retirements, mothballing, and load growth reduce the surplus capacity in the future.

Figure 24 shows that the estimated net revenue is relatively stable from zone-to-zone. In 2004, the highest prices were in the North while the lowest prices were in the South. However, the net revenue for a combined cycle unit in the North is just 10 percent higher than for the same type of unit in the South. This illustrates that the current zonal pricing does not provide sufficient incentive to invest in areas where capacity is needed most.

We also compared the net revenue in the ERCOT market with net revenue in other centralized wholesale markets. Figure 25 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, and (e) the PJM ISO. The figure includes estimates of net revenue from (a) energy, (b) reserves and regulation, and (c) capacity. ERCOT does not have a capacity market, and thus, does not have any net revenue from capacity sales.[16]

Figure 25: Comparison of Net Revenue between Markets 2002 to 2004



16 The California ISO does not report capacity and ancillary services net revenue separately, so it is shown as a combined block in Figure 25. 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.

Based on Figure 25, net revenues fell moderately throughout the country from 2003 to 2004, since most areas experienced a very mild summer in 2004. Net revenues decreased slightly or remained flat from 2002 to 2003 for every market except ERCOT, where estimated net revenue increased by a factor of three for a theoretical combined-cycle unit and by more than a factor of ten for a theoretical gas turbine. This difference can be explained by a number of factors. First, ERCOT is much more dependent on natural gas than the other markets. The sharp increase in natural gas prices in the other regions does not translate as directly into higher 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.

Third, a substantial amount of new capacity has been installed over the last three years in the Northeast and load levels were lower in 2004 than in 2002 and 2003 due to milder weather in the summer. These factors also contribute to lower net revenue in the Northeast. Finally, the increased frequency of relatively high electricity prices in ERCOT, as discussed above, also contributed to the increase in net revenue. The sources of these increases are evaluated in Section II.

Despite increases in the estimates of net revenue for ERCOT since 2002, they are still lower for a combined cycle unit than in the other markets and roughly comparable for a new gas turbine. None of these markets produces net revenue close to the amounts needed to support investment in such resources. This may not be troubling because all of the markets currently exhibit a capacity surplus outside of certain constrained areas.

There is one significant difference, however, between ERCOT and some of the other markets. ERCOT currently has no market mechanism that will ensure that its market sends economic signals that will allow it to maintain a sufficient base of generating resources. There are two primary market mechanisms employed in other areas to ensure economic signals are sufficient to maintain adequate resources:

- A capacity market; and/or
- Shortage pricing in the energy and ancillary services markets.

A capacity market is a market that causes generators' revenues to rise significantly if the generating margins decreased to levels that are inadequate to maintain reliability. Most of the current RTO markets have some form of capacity markets, which places less reliance on high prices in the energy and ancillary services markets alone to generate the revenues necessary to maintain adequate resources.

Two of the existing markets (New York and New England) employ shortage pricing provisions that complement their capacity markets, although shortage pricing can replace the capacity markets altogether if the prices are high enough during shortage conditions. The shortage pricing provisions in these markets ensure the prices in these markets reflect the true costs of the actions that are taken during true shortages (e.g., curtailment of load, sacrifice of reserves, etc.). Hence, the price rises sharply during periods of true shortages and provides transparent signals to suppliers that new generating resources are needed in the area. Absent one or both of these market mechanisms, ERCOT may ultimately have to rely on some form of mandated investment to maintain adequate resources once the current capacity surplus dissipates.

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, among other things, units that are expected to be online and producing energy. Under ERCOT's relaxed balanced schedules policy, the load schedule is not required to approximate the QSE's projected load. When a QSE's forward 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 forward 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 forward scheduling by QSEs. The analysis focuses on the degree to which forward 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.

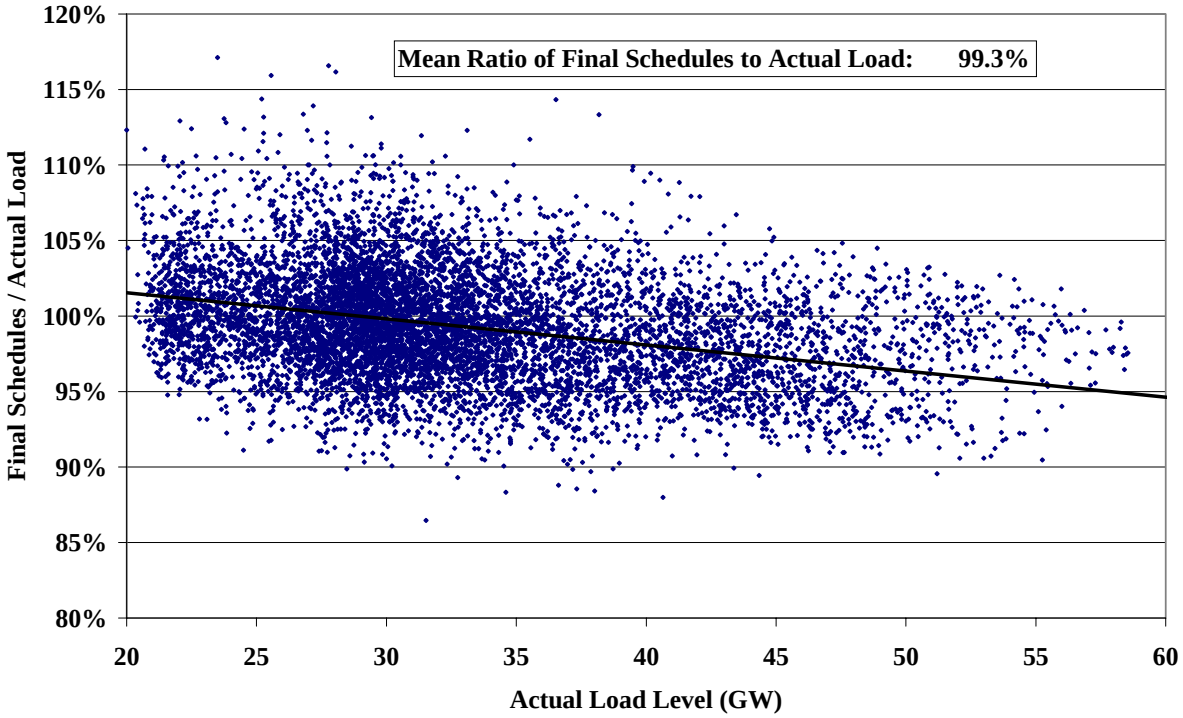
A. Forward Load Scheduling

In this subsection, we evaluate forward load scheduling patterns by comparing forward load schedules to actual real-time load. We focus on the forward load schedules at two points in time.

First, we will refer to the final load schedule, which is the last schedule submitted by the QSE prior to the operating hour. Second, we will refer to the day-ahead schedule, submitted by the QSE in the day ahead.

To provide an overview of the scheduling patterns, Figure 26 shows a scatter diagram that plots the ratio of the final load schedules to the actual load level during 2004.

Figure 26: Ratio of Final Load Schedules to Actual Load
All ERCOT – 2004



The ratio shown in Figure 26 will be greater than 100 percent when the final load schedule is greater than the actual load. Therefore, in general, the figure shows that final load schedules come very close to actual load in the aggregate, as indicated by an average ratio of the final load schedules to actual load of 99.3 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 28 GW and declines to 96 percent for load above 50 GW. The overall pattern shown in the figure above is similar to 2003, which exhibited the same downward trend in final load schedules relative to actual load, although the average ratio was only 98.4 percent.

This result runs counter to what would typically be expected. Normally, one would expect balancing energy prices to be more volatile at high load levels. Therefore, if market participants were generally risk averse, they would be expected to schedule forward to cover their load during high load periods rather than reducing their forward scheduling levels during those periods. There may be several explanations for this scheduling pattern. 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 the attendant price risk. Financial contracts or derivatives may be in place to protect market participants from the price risk in the balancing energy market, such as a contract for differences. Second, they can cover themselves by bidding enough generation to cover their load needs in the balancing energy market. Last, the fact that balancing energy prices have not risen predictably with actual load levels (as shown above) may provide an incentive for some market participants to purchase peak energy from the balancing energy market that they need to satisfy their load.

Figure 27 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 27: Average Ratio of Final Load Schedules to Actual Load by Load Level All Zones – 2004

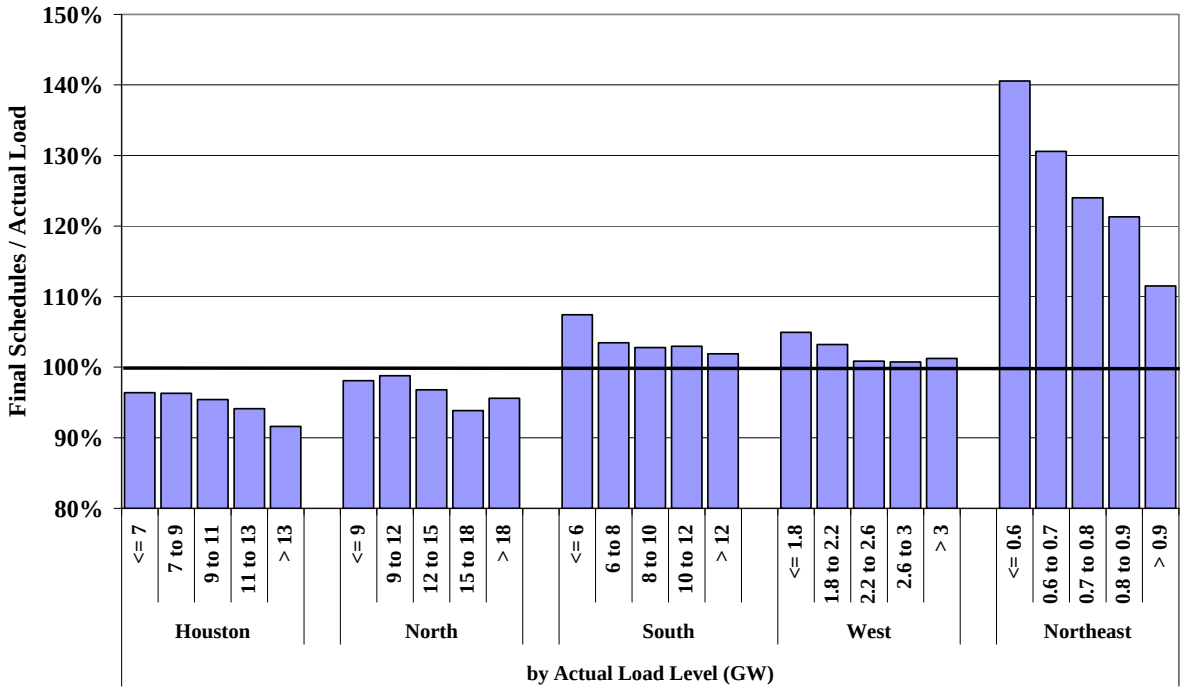


Figure 27 shows that:

- The final schedule quantity decreases in each of the five zones as actual load increases, with the exception of the North and West zones where a slight increase in the ratio occurs at the highest load level.

- The West Zone and South Zone are generally over-scheduled slightly, although the ratios still decline 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 is usually less than 300 MW.

- The North Zone and Houston are under-scheduled at each load level, ranging from 1 percent at lower load levels up to 9 percent at high load levels.

The pattern in Figure 27 differs somewhat from load scheduling in the previous year. In 2003, the West Zone consistently over-scheduled load by a substantial amount, while the North Zone was under-scheduled by 5 to 10 percent. Given that the Northeast Zone was not separated from the North until 2004, this indicates a substantial upward shift in load scheduling for the North Zone from 2003 to 2004.

The result of these scheduling patterns is that the QSEs in the North Zone and 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.

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, even if they have not sold the power in a bilateral contract. Rather than selling power to the balancing energy market through energy imbalances or 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.

However, QSEs with generators in locally-constrained areas can benefit from systematic over- and under-scheduling. The local congestion management process rewards QSEs that over-schedule in export-constrained areas and under-schedule in import-constrained areas. When QSEs over-schedule generation in a local area, there is a tendency for them to over-schedule load in the corresponding zone. Later in this section, we perform a resource-level analysis to identify at least one factor that may contribute to this scheduling pattern.

We next evaluate the day-ahead load schedules relative to actual load in Figure 28. The figure is analogous to Figure 27. It shows the ratio of day-ahead load schedules to actual load by load level for each of the four zones in ERCOT. The load levels are divided into five roughly equal groups.

Figure 28: Average Ratio of Day-Ahead Load Schedules to Actual Load by Load Level All Zones – 2004

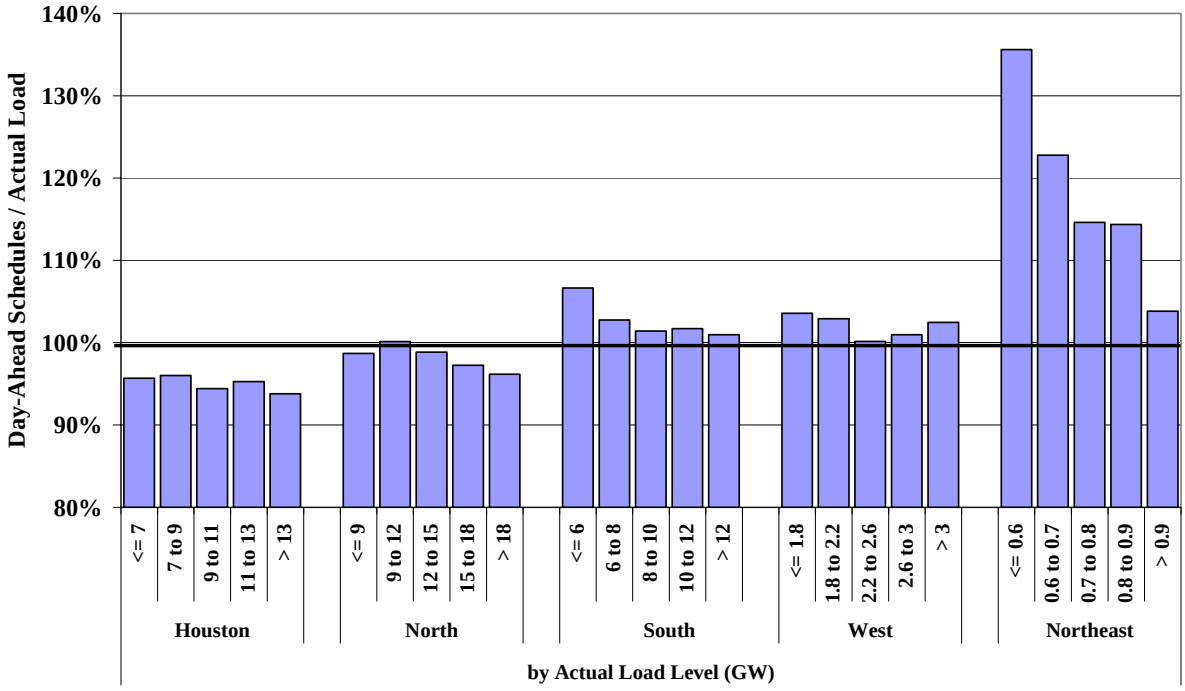
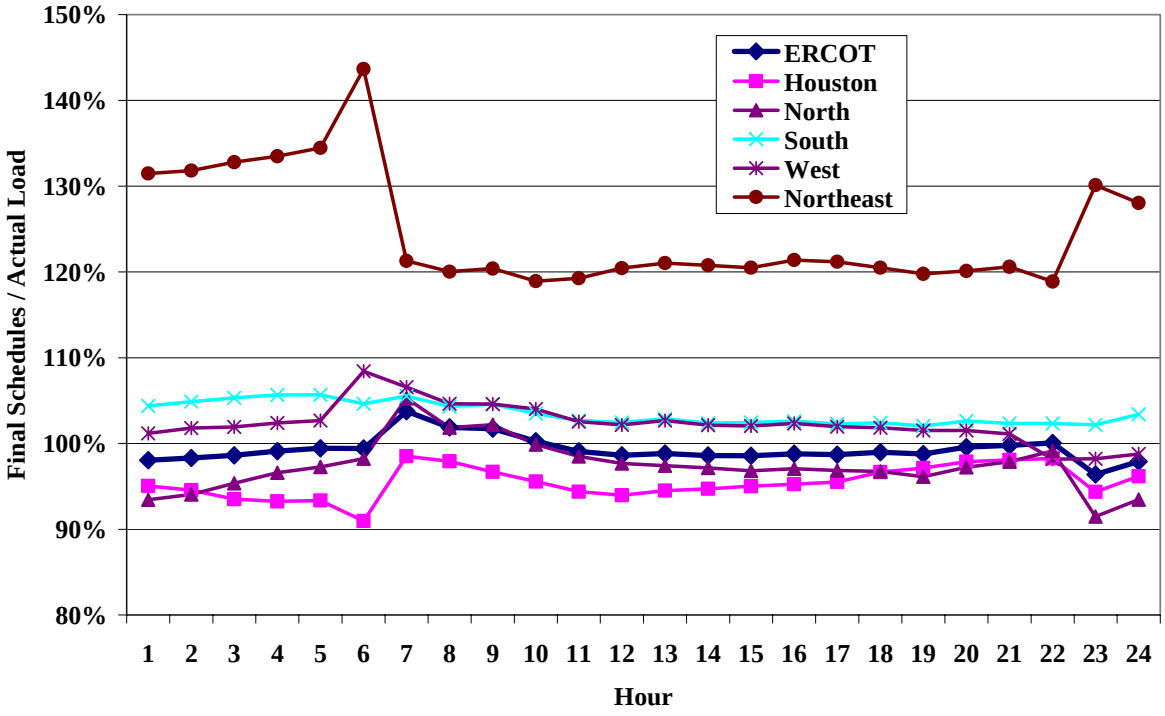


Figure 28 shows that day-ahead scheduling results are generally comparable, both in magnitude and pattern, to the scheduling levels shown in Figure 27 for final load schedules. Day-ahead load schedules in the Houston, South, and West Zones are negatively correlated with actual load levels, but to a lesser degree than the final load schedules. There is also less over- and under-scheduling in the North and Northeast Zones in the day-ahead than in real time.

Although there is no obvious explanation for the differences in scheduling patterns between the day-ahead and real-time in the North Zone and Northeast Zone, a more detailed analysis of out-of-merit commitment and dispatch described below provides some insight. In addition, Section III of this report analyzes changes in QSEs' resource plans from the day-ahead to the real-time timeframe that are consistent with changes in the QSEs' schedules.

To further analyze forward scheduling, Figure 29 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 29: Average Ratio of Final Load Schedules to Actual Load All Zones by Hour of Day - 2004



This figure shows that on an ERCOT-wide basis, final schedules are close to actual load (between 98 percent and 100 percent) during hours ending 1 to 6. At hour ending 7, the ratio rises to 104 percent, the highest of any hour. During the remainder of the day, the ratio declines to between 96 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 29 is that participants tend to submit schedules consistent with their bilateral

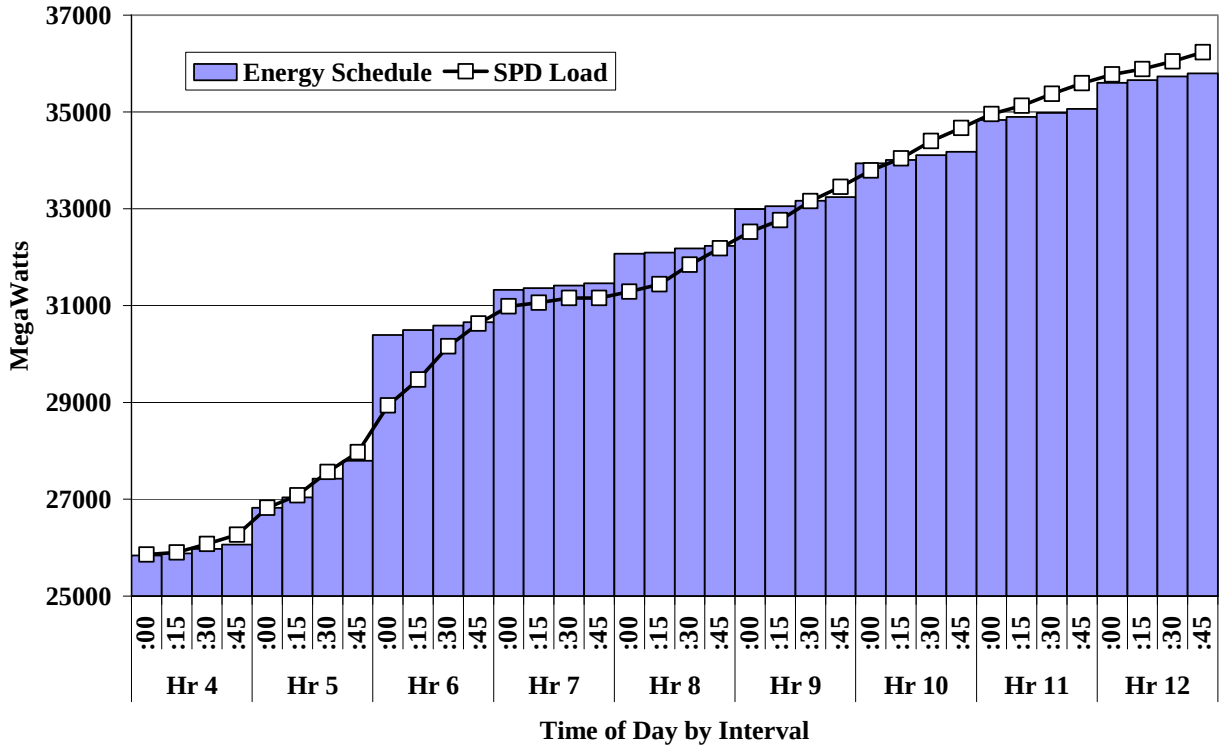
transaction positions. This is not irrational if the market participants also submit balancing energy offers to optimize the energy that is actually deployed. In addition, this pattern of forward scheduling is consistent with the notion that market participants bear additional price risk in ramping hours (as shown in the prior section), accounting for their propensity to schedule a larger portion of their needs during these periods. However, the latter explanation does not explain the sharp change in scheduling in hours ending 7 and 22.

B. Balancing Energy Market Scheduling

In the previous section, we analyzed balancing energy prices and load and found that actual load is not the primary determinant of balancing energy prices. 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 30 shows average energy schedules and actual load for each interval from 4 AM to 1 PM during 2004.

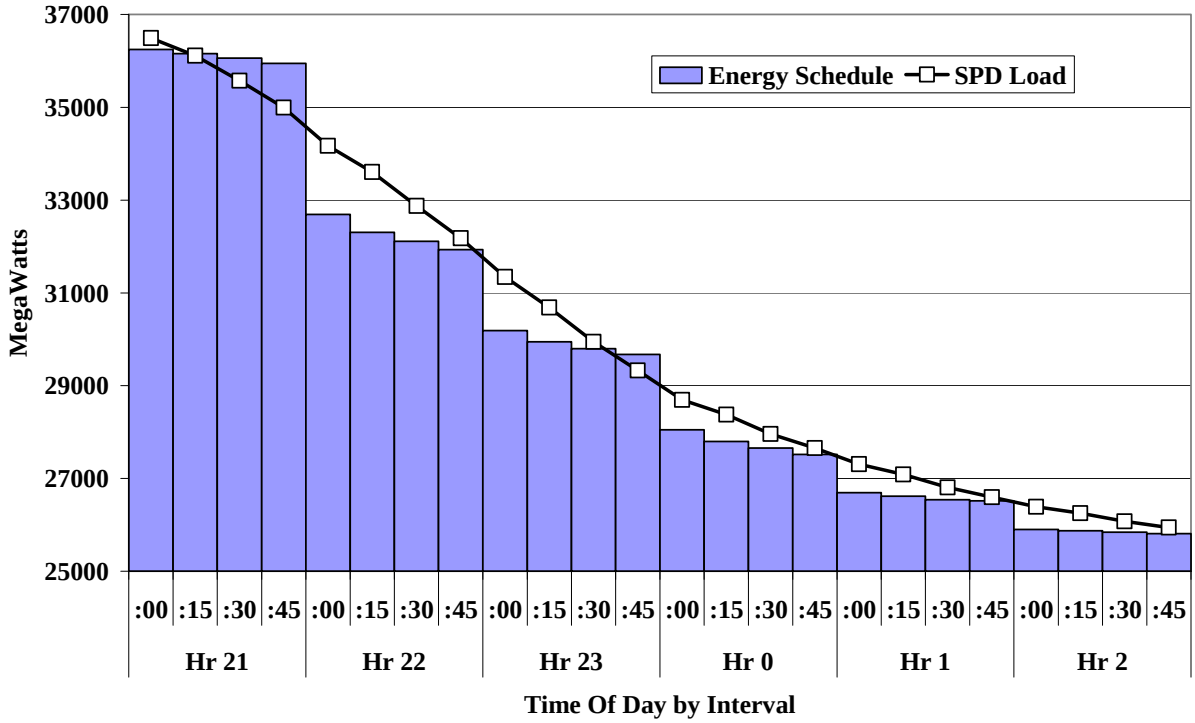
Figure 30: Final Energy Schedules during Ramping-Up Hours 2004



In general, 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 26 GW to more than 36 GW in the nine hours shown in Figure 30. The average increase per 15-minute interval is approximately 290 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 to the fact that the daily peak occurs in the morning during certain times of year.

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 more than 2 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 31 shows average energy schedules and load for each interval from 9 PM to 3 AM during 2004.

Figure 31: Final Energy Schedules during Ramping-Down Hours 2004



On average, load drops from approximately 36 GW to less than 26 GW in the six hours shown in Figure 31. The average decrease per 15-minute interval is approximately 420 MW, although the rate of decrease is greatest from 9:45 PM to midnight. The progression of load during ramping-down hours is steady relative to the progression of energy schedules. As during the ramping-up hours, energy schedules decrease in relatively large steps at the top of each hour. For instance, the average energy schedule drops approximately 3 GW from the last interval before 10 PM to the interval beginning at 10 PM.

The sudden changes in energy schedules that occur at the beginning of each hour during ramping-up hours and at the end of each hour during ramping-down hours arise from the fact that approximately one-half of the generation in ERCOT is scheduled by QSEs that submit energy schedules that change hourly, while the other half is scheduled by QSEs that submit energy schedules that change every 15 minutes. Deviations between the energy schedules and actual loads will result in purchases or sales in the balancing energy market. Specifically, net balancing up energy equals real-time load minus scheduled energy. Hence, Figure 30 indicates that during ramping-up hours, QSEs tend to purchase balancing energy on net at the end of each hour and sell balancing energy at the beginning of each hour. On the other hand, Figure 31 indicates that during ramping-down hours, QSEs tend to sell balancing energy on net at the beginning of each hour and purchase balancing energy at the end of each hour.

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 30 and Figure 31). This analysis is similar to that shown in Figure 14 and Figure 15, except instead of showing balancing energy prices relative to load, we show balancing energy prices relative to balancing energy deployments. Figure 32 shows the analysis for the ramping-up hours.

Figure 32: Balancing Energy Prices and Volumes
Ramping-Up Hours – 2004

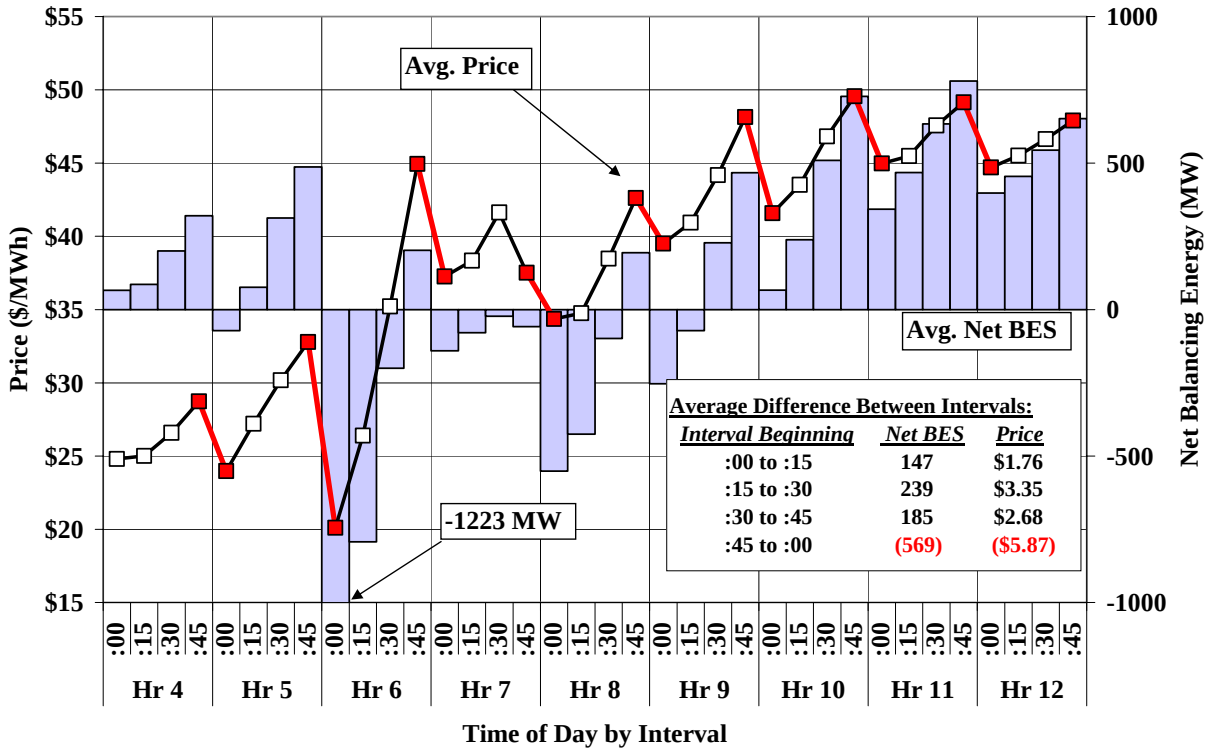
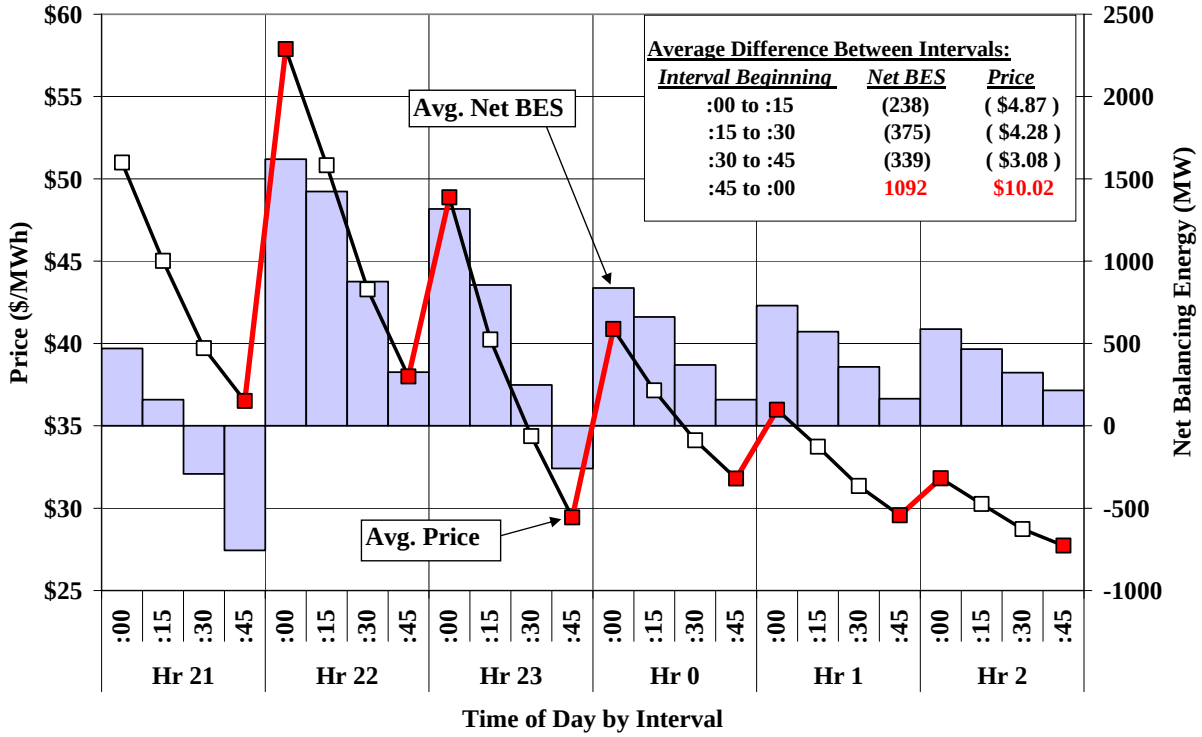


Figure 32 reveals two key aspects of the balancing energy market. First, as discussed above, balancing energy prices are highly correlated with balancing energy deployments. Second, there is a distinct pattern of increasing purchases during the hour. At the beginning of the hour, purchases tend to be 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 33 shows the same analysis for the ramping-down hours.

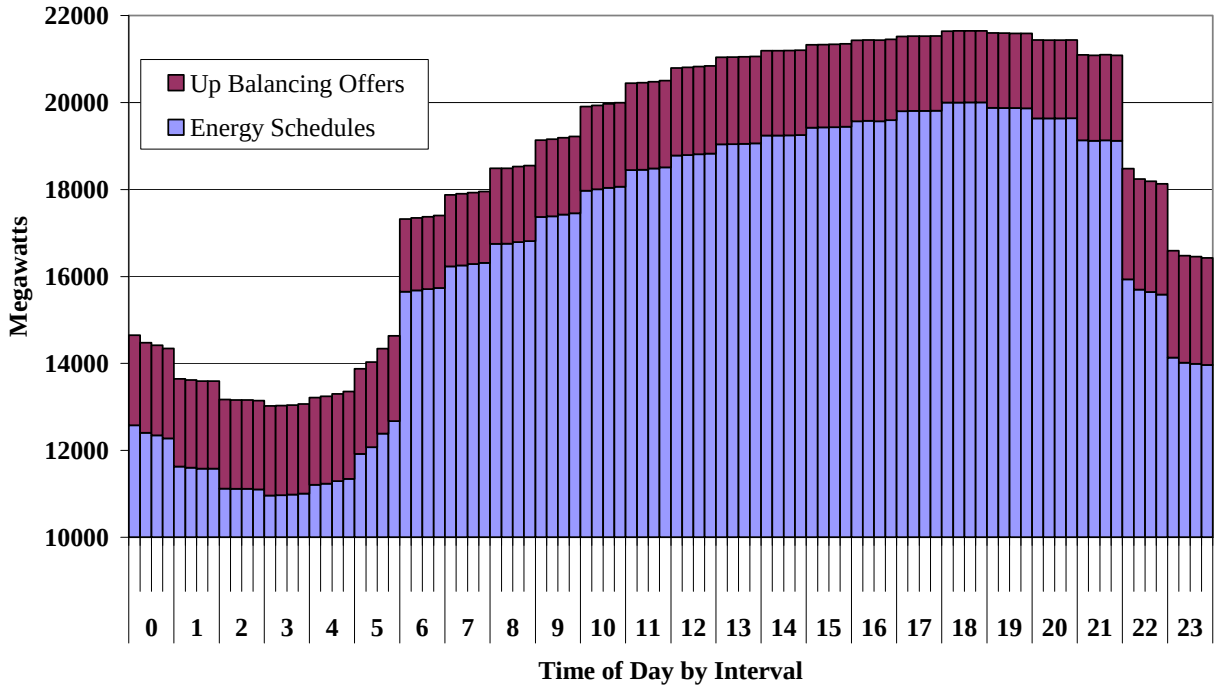
Figure 33: Balancing Energy Prices and Volumes
Ramping-Down Hours – 2004



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), most of the QSEs schedule only on an hourly basis, making little or no changes on a 15-minute basis. We reviewed QSE scheduling data from 2004 and found that the two largest suppliers in ERCOT tend to schedule much more flexibly than other QSEs. The following two figures analyze the scheduling patterns of the two largest QSEs compared to all other QSEs by interval over the entire day. Figure 34 shows the average quantity of energy schedules and balancing-up offers in 2004 for all QSEs in ERCOT except the largest two.

Figure 34: Final Energy Schedules and Balancing-up Offers
All QSEs except the Largest Two



This figure shows that there is almost no change in the energy schedules on a 15-minute basis, but relatively large changes from hour to hour. It is primarily the scheduling patterns by these QSEs that result in the balancing energy deployments and prices shown in Figure 32 and Figure 33. In addition to the fact that these QSEs generally schedule hourly, this figure shows the sharp schedule changes that occur at the beginning and end of the 16 peak hours commonly used in bilateral contracts (hour ending 7 to hour ending 22). Energy schedules increase by 2,975 MW on average from the last interval of hour ending 6 to the first interval of hour ending 7, an increase of nearly 25 percent in just one interval.

The scheduled energy decreases even more abruptly at the end of the peak bilateral contract period. Scheduled energy decreases by 3,190 MW on average from the last interval of hour ending 22 to the first interval of hour ending 23. The next two hours also show relatively large decreases in scheduled energy of 1,450 MW and 1,390 MW, respectively. These large hourly changes in energy schedules are a primary determinant of the balancing energy price fluctuations shown in this section.

Figure 35 shows the energy scheduling patterns of the two largest QSEs, which account for approximately one half of the energy schedules.

Figure 35: Final Energy Schedules and Balancing up Offers
Largest Two QSEs

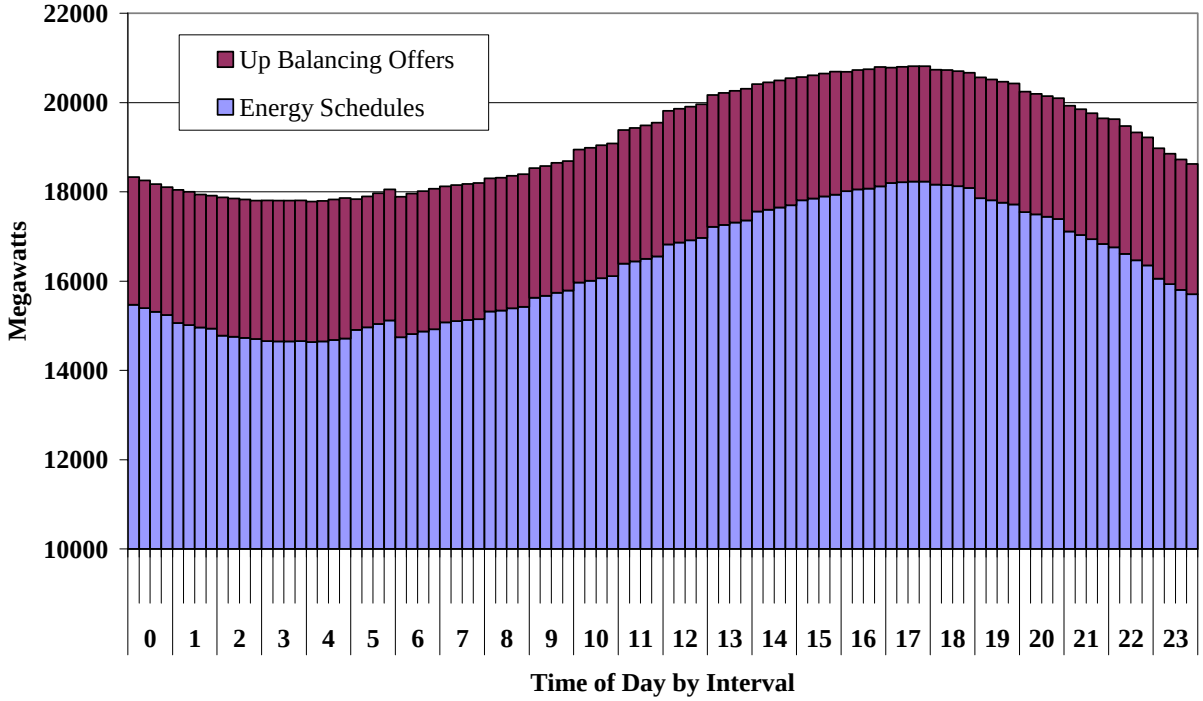


Figure 35 shows that the largest QSEs tend to schedule much more flexibly than the other QSEs. These two QSEs fully utilize the capability to schedule energy on a 15-minute basis. Like the schedules of the other QSEs, these schedules show a shift at the beginning of the peak bilateral contracting period from hour ending 7 to hour ending 22. However, in contrast to other QSEs, the large QSE's energy schedules show a slight *decrease* in the first interval of hour ending 7. While other QSEs decrease their schedules significantly after hour ending 22, these large QSEs show a smooth progression from hour ending 22 to 23.

The large QSEs are in a position to take advantage of profitable arbitrage opportunities that occur at the beginning and end of the peak bilateral contracting period. It allows them to balance up just before and after the peak bilateral contracting period from hour ending 7 to hour ending 22 and balance down just after it starts and before it ends. Although the large QSEs do not fully counter-balance the large changes in schedules by smaller QSEs at the beginning and end of the 16 hour peak period, their scheduling patterns in those ramping hours and the fact that they

generally submit energy schedule changes on a 15-minute basis improve the performance of the balancing energy market.

Finally, the figure shows that the large suppliers tend to offer more energy (in total and as a percentage of their capability) into the balancing energy market throughout the day than the smaller QSEs. This should further improve the performance of the balancing energy market by increasing its liquidity. This is evaluated further in the next subsection.

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 two previous reports,[17] and has become slightly more pronounced in 2004. To address this issue, we have 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. This adjustment would assume that intra-hour increases in energy schedules are supplied from the lowest-cost portion of the QSE's balancing energy offer. This would help ensure that the participant's portfolio energy offer is consistent with its energy schedules when the energy schedule is changing each interval. Furthermore, it would facilitate QSEs offering more of their on-line capacity to the balancing market, which would help address the problem that large amounts of on-line capacity are not offered to the balancing market. Protocol Revision Request ("PRR") 600 was written to address this recommendation but the provision of the PRR that addresses this recommendation has not been retained by the protocol revision subcommittee. We recommend that this provision be adopted in the future.

[17] See "ERCOT State of the Market Report 2003", Potomac Economics, August 2004; "2004 Assessment of the Operation of the ERCOT Wholesale Electricity Markets", Potomac Economics, November 2004.

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). This sub-section analyzes 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.

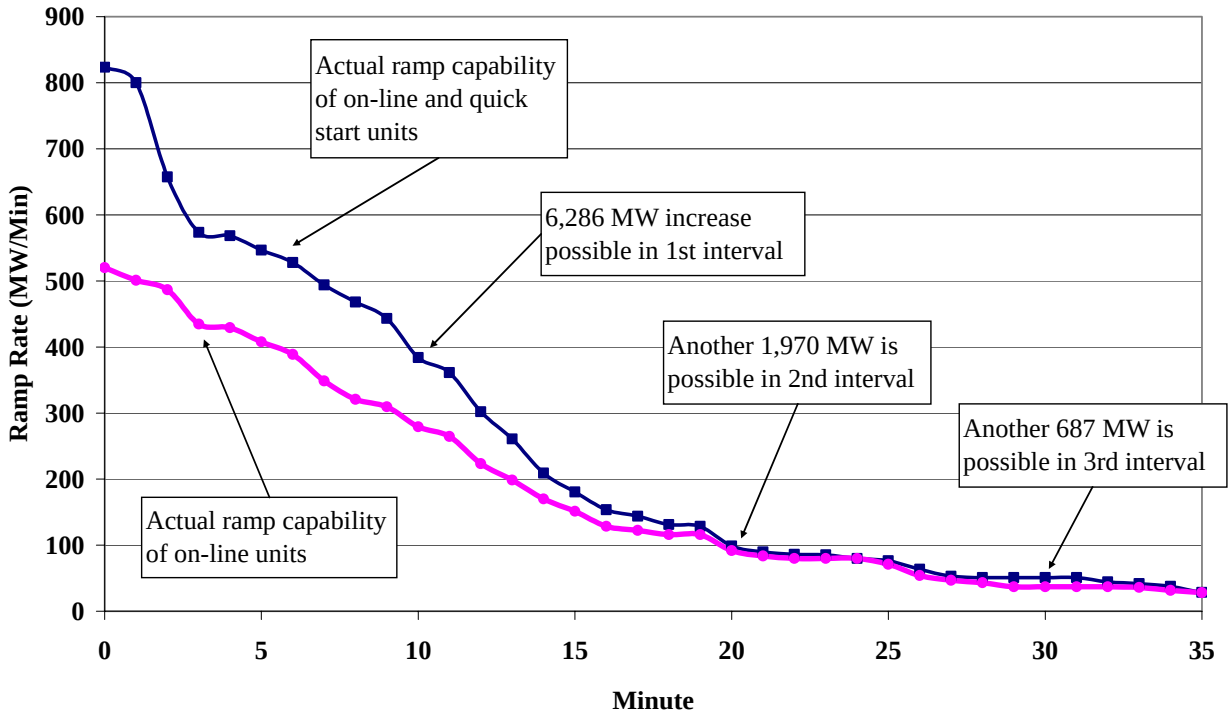
1. Portfolio Ramp Rates

Frequently, market participants are prevented from fully deploying their portfolio by the physical ramp limitations of the resources in their portfolio. To manage this physical limitation, QSEs submit offers with portfolio ramp rates that represent the physical ramp capability of their units on a portfolio basis. Market participants with competitive incentives should submit the highest possible ramp rates that are physically feasible while respecting their ancillary services obligations. Thus, a market participant must estimate the ability of its resources to increase output, which varies as the level of output increases. The following is an analysis that estimates the change in ramp capability for available capacity in ERCOT as resources are deployed upward.

Figure 36 shows two estimates of the physical ramp capability for the dispatchable resources in ERCOT for a sample hour during 2004: (i) including the capability of on-line and quick start resources, and (ii) including the capability of only on-line resources. For this analysis, each unit

is assumed to start out at the planned generation level in the real-time resource plan and increase its output according to the unit's ramp rate specified in the real-time resource plan until the unit reaches its high sustainable limit ("HSL"). Capacity is set aside for each QSE to satisfy its regulation up and responsive reserve obligations.[18]

Figure 36: Physical Ramp Capability of On-Line and Quick Start Resources
March 15, 2004 – Midnight



The figure shows that the ramp capability for all on-line and quick-start resources in ERCOT is 824 MW per minute before any ramping takes place. After one minute of ramping, the ramp capability drops to 800 MW per minute. And after ten minutes of ramping, the ramp capability drops to 384 MW per minute. When an individual resource in ERCOT reaches its HSL, the resource is no longer capable of additional ramp, which implies that the ramp capability of the ERCOT market diminishes as the market increases output. For this reason, both lines in the figure are decreasing over time.

[18] This is done by allocating the obligation to its regulation capable units in proportion to ramp capability and headroom. In addition, a QSE capable of ramping 100 MW in a ten-minute period that has a 40 MW responsive reserves obligation can only ramp 60 MW in a ten-minute period and still provide reserves. Thus, 40 MW of capacity capable of ramping in 10 minutes must be set aside by the QSE. The analysis allocates responsive reserves to the most expensive units in the portfolio, wherever possible. The generic resource costs described in Section 5 of the ERCOT Protocols are used to rank generators in order of cost.

In each market interval, QSEs are supposed to move to their instructed dispatch level over a ten-minute period. Thus, the capacity that is physically deployable in one interval is the sum of what can be ramped in the first ten minutes (as shown in the figure above). This starts at 824 MW per minute but decreases to 384 MW per minute after ten minutes for a total of 6,286 MW in the first interval. However, only 1,970 MW can be deployed in the second interval, and just 687 MW is deployable in the third interval.

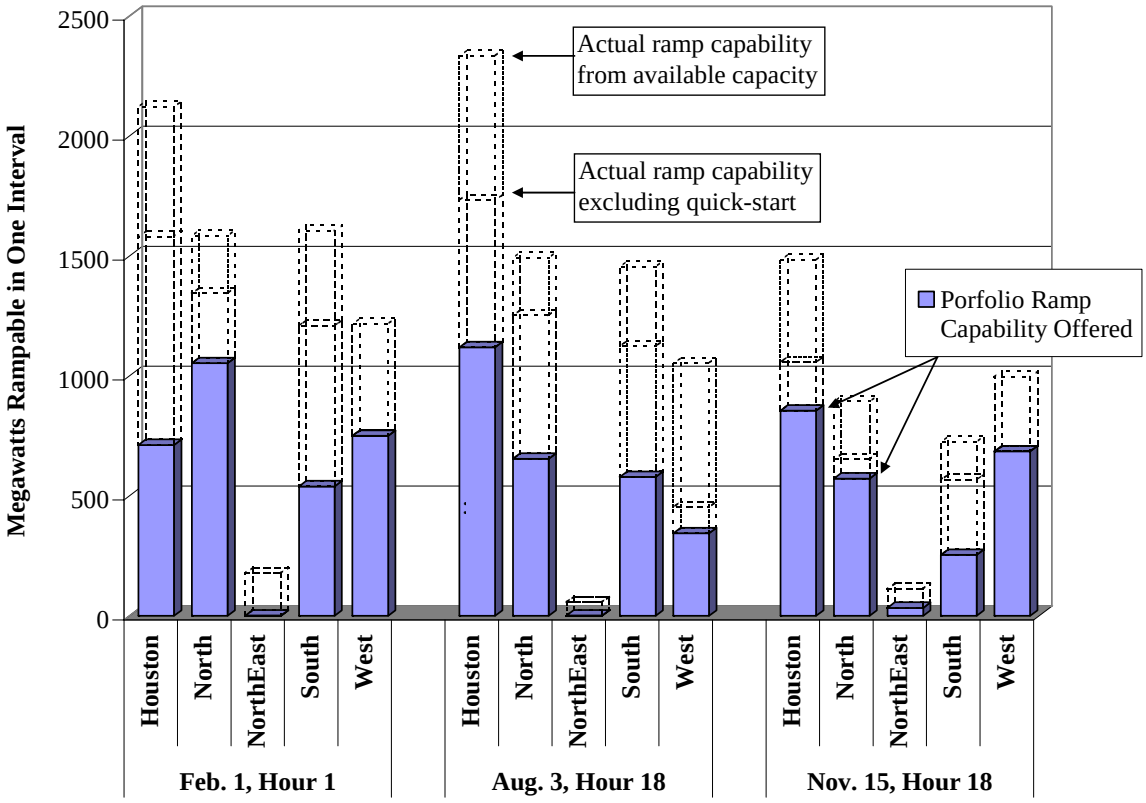
To limit changes in dispatch from one interval to the next, market participants submit a ramp rate for their portfolio in a particular zone. Ideally, market participants could submit a portfolio ramp rate that decreases as production increases to reflect the physical reality shown in Figure 36. However, market participants offer a single constant ramp rate for all capacity above the energy schedule. If they set their ramp rate at a level that allows the maximum feasible deployment in the first interval, it will not be feasible for most QSEs to ramp at that rate in the second interval.

To avoid being over-deployed because of the constant portfolio ramp rate, QSEs have at least three options. First, QSEs can simply lower their portfolio ramp rate to a level that is achievable over all four intervals of the hour. This may allow all capacity to be deployed in one hour, but significantly diminishes the efficient response of the balancing energy market to transient price spikes. Second, QSEs can lower their offer quantities to the amount of capacity that can be physically deployed in two intervals. This allows them to offer a higher ramp rate, but leads to QSEs not offering all of their capacity. Third, QSEs can make multiple portfolio offers, each representing only a fraction of their portfolio. This does not solve the problem, but reduces its magnitude. To address the underlying issue, ERCOT could modify its market software to allow QSEs to offer multiple portfolio ramp rates covering different portions of their offer so that they have the flexibility to offer a larger portion of their available energy into the balancing energy market.

It is not possible to measure how requiring QSEs to offer a constant ramp rate will impact their final offers and/or market outcomes. But it will inevitably lead to smaller offer quantities and lower portfolio ramp rate offers. Portfolio ramp rates are a QSE's means for representing on a portfolio basis the physical ramp capability of the units within its portfolio. In Figure 37, we compare the portfolio ramp rates to actual physical ramp capability at three points in time in

2004. These three hours were selected because they are representative of a variety of conditions during 2004. They occur in the winter, summer, and fall. One is an off-peak hour at night while the other two occur in the afternoon.

Figure 37: Portfolio Ramp Rates versus Ramp Capability Examples from 2004



The bars in this figure show the quantity of unscheduled capacity from each zone that can be ramped in one interval. The bottom portion of the bar represents the balancing up offers that are available in one interval based on the portfolio ramp rate submitted by the QSE. The next portion of the bar is the additional energy that is physically available from online resources given their unit-level physical ramp limitations. The top portion of the bar is the additional energy that can be provided within one interval from offline quick-start resources. Hence, the total height of each bar is the amount of energy that could physically be provided within one interval (i.e., "rampable" capacity) from all available resources while the bottom portion of each bar shows the lower amount of energy available in the balancing energy market due to the tighter portfolio ramp limitations.

This analysis identifies the potential effects of the current portfolio offer structure. The figure shows that:

- In the West zone and in Houston, a substantial portion of the physically rampable capacity was not available based on portfolio offers and ramp rates on all three days. A large share of this difference may be explained by offline quick-start resources not being offered in the balancing energy market. We discuss below why this is likely the case.

- The North Zone showed the highest levels of rampable portfolio offers as a percentage of the physically rampable capacity on all three days. It is notable that the North Zone has less quick-start capacity as a percentage of its total capacity.

While we cannot measure how much of the difference between actual rampable capacity and rampable offers is caused by the lack of flexibility QSEs have in submitting portfolio ramp rates, it is likely that a portion of the difference in Figure 37 is attributable to this factor. Therefore, we recommend that ERCOT consider the costs and feasibility of allowing QSEs to offer multiple ramp rates that vary by output level.

2. Unit-Level Dispatch Would Optimize Ramp Capability

When a QSE receives balancing energy deployments from its portfolio, it will naturally prefer to increase output on its lowest-cost resources to satisfy the deployment. To the extent that a supplier's physical ramp capability is on higher-cost resources (e.g., gas turbines) which it would not prefer to dispatch before its lower-cost resources, it is not rational for the portfolio ramp rate to include that ramp capability. The following example illustrates the problem that QSEs face in offering multiple resources using a single portfolio ramp rate.

Suppose that a QSE has a portfolio consisting of just two units, one coal-fired and one natural gas-fired. The coal unit can produce 100 MW of additional energy at a marginal cost of $16/MWh, while the natural gas unit can produce 100 MW of additional energy at a marginal cost of $50 per MWh. Assume that both units have a ramp rate of 5 MW per minute, allowing each unit to dispatch 50 MW in one interval and to be fully dispatched in two intervals. If the market clearing price in the next interval is between $16 per MWh and $50 per MWh, the balancing energy market will deploy the 100 MW of lower costs energy from the coal unit.

However, to satisfy such a deployment, the generator would have to dispatch 50 MW from the natural gas unit since the coal unit can only ramp 50 MW in one interval (10 minutes times 5 MW per minute). To address this risk, the QSE may submit a portfolio ramp rate of 5 MW per minute, although this introduces potential opportunity costs when prices are higher than $50 per MWh and it is profitable to dispatch the natural gas unit. This simple example illustrates why the portfolio offer structure can generally lead suppliers to submit ramp rates that are substantially less than the maximum physical ramp rate or find other ways to address the difficulties associated with the portfolio bidding framework.

One of the significant benefits of an alternative market design that would allow for unit specific bidding and deployment would be the ability of suppliers to make offers from specific units with unique ramp rates. This would provide increased flexibility and more efficient dispatch of the system. Under the current system, QSEs can submit multiple portfolio offers with independent ramp rates by defining "sub-QSEs". This allows increased flexibility to offer higher-cost, fast-ramping resources without the risk that they may have to be dispatched at a loss. Additionally, this provides a means for QSEs to offer the capability from their gas turbines in the balancing energy market. Thus far, several QSEs have chosen to group coal and wind units under separate sub-QSEs, however, most QSEs have not made use of this flexibility.

3. Ramp Rates Ignore Schedule Changes

An additional factor that limits QSEs ability to submit portfolio ramp rates that make maximum use of their physical resources is that energy schedule changes are not currently considered when the balancing energy market is cleared. QSEs offer balancing-up and balancing-down energy quantities that are constrained by portfolio ramp limitation relative to balancing energy deployment in the prior interval.

For instance, if a QSE has 100 MW of cleared balancing-up energy in the previous interval and 800 MW of balancing up offers in the current interval with a ramp rate of 45 MW per minute, the QSE could sell up to a total of 550 MW (= 100 MW + 45 MW per minute * 10 minutes) in the current interval. If 550 MW is cleared in the current interval, the remaining 250 MW could be available in the subsequent interval. Thus, the ramp rate limits the change in the *deployed* quantity from one interval to the next without respect to changes in the energy schedule. In the

previous example, if the QSE's energy schedule increased by 400 MW in the second interval, the QSE would actually be responsible for increasing output by 850 MW (= 400 MW schedule change + 450 MW change in cleared balancing energy). If 450 MW represents the maximum amount the QSE can ramp in one interval, the QSE would need to submit a ramp rate of 5 MW per minute so that the total change in output is physically feasible and does not exceed 450 MW (= 400 MW schedule change + 50 MW change in cleared balancing energy).

However, this strategy of simply lowering the portfolio ramp rate can prevent the unit from receiving deployments that are fully consistent with its offer prices, reducing the supplier's profits and the efficiency of the overall market in two ways. First, this would reduce the ability of the balancing energy market to produce balancing energy deployments that would offset prior deployments and schedule changes. In the example above, for instance, the 5 MW per minute ramp limitation would constrain the total change in balancing energy deployments to be less than 50 MW even though the QSE could clearly accept a reduction in its balancing-up deployment from the prior interval by 100 MW to zero by simply increasing its output only 300 MW rather than the 400 MW called for by the increase in its energy schedule.

Second, the portfolio offers and ramp rates are set every hour and cannot change each interval. Thus, in the three subsequent intervals of the same hour in the example above, the QSE would be limited to a total change in balancing energy deployments of 50 MW in each interval resulting in a maximum increase in balancing-up energy over the hour of 200 MW. In reality, once the QSE increases its output to satisfy its 400 MW schedule change, it could then accept increasing balancing-up deployments of as much as 450 MW in each interval (and the deployment of its entire 800 MW portfolio offer over two intervals).

To avoid these issues, the QSE could choose to submit a portfolio ramp rate that does not account for its changes in energy schedule (45 MW per minute in our example). However, it would then risk uninstructed deviation penalties in the first interval if it receives a physically infeasible balancing energy deployment. Hence, the current application of the portfolio ramp rate constraints makes it impossible for QSEs to submit an accurate ramp rate for all four intervals when its energy schedule is changing significantly at the top of the hour.

To address this issue, we recommend that ERCOT modify its Scheduling, Pricing, and Dispatch

("SPD") software for the balancing energy market model to account for the ramp capability that is utilized (or created) when the energy schedule increases or decreases. For example, assume that the hypothetical QSE discussed above can ramp up or down by 45 MW per minute or 450 per interval. If its energy schedule changes by 400 MW at the top of the hour, the SPD should recognize that it now has the capability to balance up in the first interval of the hour by only 50 MW (450 MW – 400 MW) and to balance down by 850 MW (450 MW + 400). The recognition that it has an increased capability to balance down by simply not increasing its output consistent with its energy schedule may sometimes provide SPD valuable additional flexibility in making balancing energy deployments.

In the second interval of the hour, SPD would then recognize that the QSE can now increase or decrease its balancing energy deployments by 450 MW without the QSE having to modify its portfolio ramp limitation. Hence, this change can potentially increase the flexibility of the energy offered in the balancing energy market and improve the balancing energy market's performance. This recommendation was made in two previous reports.[19] However, since it would require significant changes to the market software, and the benefits would only be realized under the current market design, the recommendation has been tabled by the Technical Advisory Committee pending the outcome of the PUCT ruling on the Texas Nodal market design.

D. Balancing Energy Market Offer Patterns

In this section, we evaluate balancing energy offer patterns by analyzing the rate at which capacity is offered to supply balancing energy. In Figure 38, we show the average amount of capacity offered to supply balancing up service relative to all available capacity. The offered capacity is divided into that which is ramp-constrained, and would not actually be capable of supplying balancing energy, and that which is non-ramp-constrained, and thus would be available to supply balancing energy. Capacity is considered to be available if it is either physically on-line or it is from an off-line quick start unit, and if it is not scheduled to provide energy, reserves, or up-regulation. This includes the portion of available capacity under the High Sustainable Limit ("HSL"), and therefore does not include emergency ranges.[20] Unused capacity

[19] See "ERCOT State of the Market Report 2003", Potomac Economics, August 2004; "2004 Assessment of the Operation of the ERCOT Wholesale Electricity Markets", Potomac Economics, November 2004.

[20] Although the HSL does not include the emergency range, it may include less flexible operating ranges.

on renewable resources such as wind turbines are excluded from this category. This data is shown for the peak hour of the day on a monthly average basis in 2003 and 2004 in Figure 38.

Figure 38: Balancing Energy Offers versus Available Capacity
Daily Peak Load Hours – 2003 and 2004

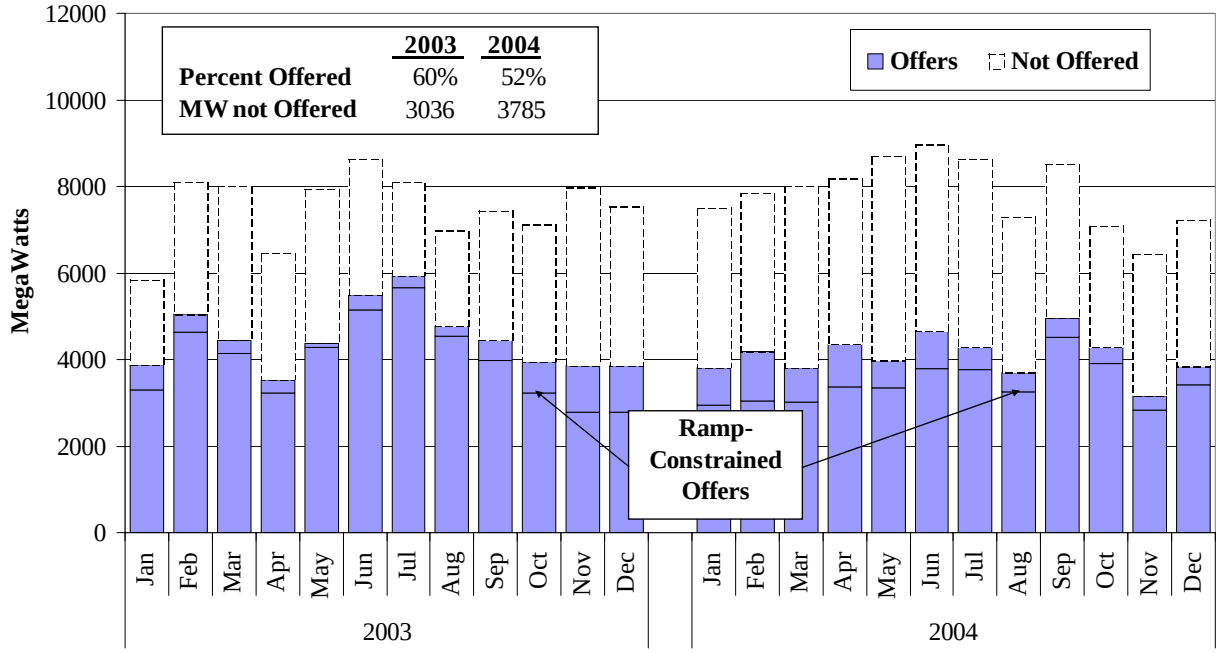


Figure 38 shows the trends over time in quantities of energy available and offered to the balancing energy market. The figure indicates that the amount of available energy has fluctuated in a relatively constant range over the two-year period, with the exception of the significant decrease in available energy during the last three months of 2004. Figure 75 in Section VI indicates that this coincides with significant reductions in Out-of-Merit Capacity ("OOMCs") commitments. The reduced OOMCs have resulted in less on-line capacity overall and, therefore, less available energy.

Figure 38 indicates the average share of the offer quantity that can be ramped up in one interval, indicating that participants generally offer slightly more than this level. On monthly basis, the average share of capacity that can be ramped in one interval ranged between 176 MW to 913 MW. The fraction of offers unavailable due to ramp constraints was relatively constant at various load levels and on a monthly basis. These results support the hypothesis discussed above

For instance, many units are less flexible in their duct firing range. This underscores the benefit of allowing market participants to offer into the balancing market with ramp rates that vary by output level.

that one reason participants may not offer all of their available energy is to manage the ramp capability of their portfolio.

Figure 38 also shows that the portion of the energy offered in 2004 decreased slightly. Hence, the average quantity of available capacity not offered in the balancing energy market increased by 749 MW in 2004. The amount of available non-offered energy increased substantially in the last two months of 2003, and this pattern continues through the early summer of 2004.

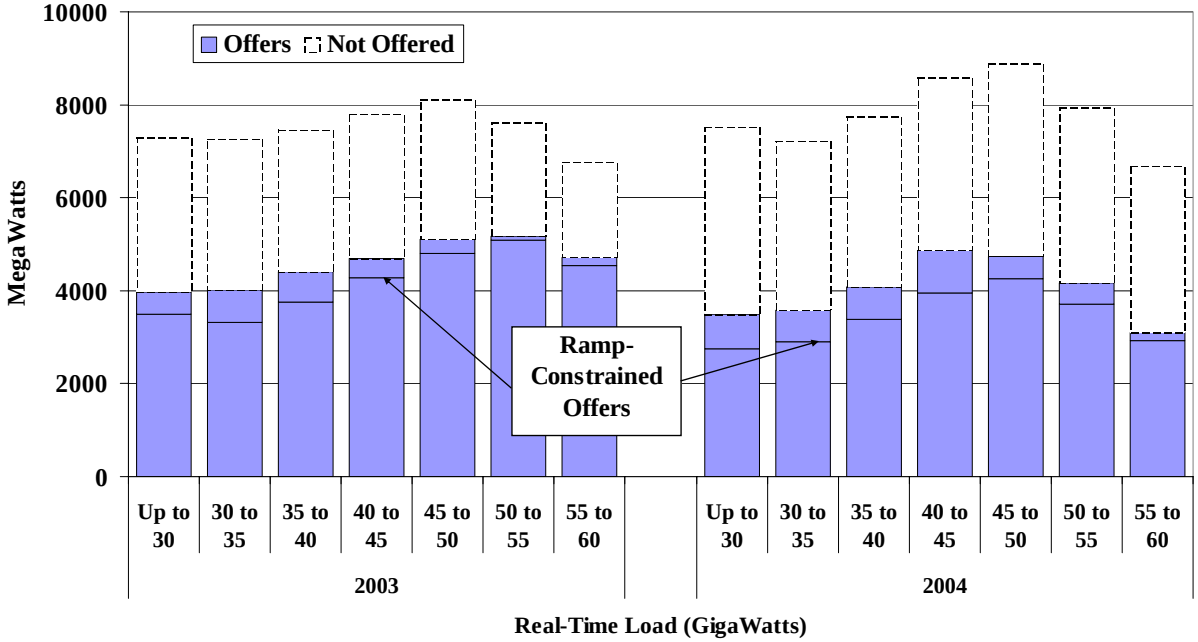
There are many factors that could contribute to additional non-offered energy. First, the fraction of responsive reserves satisfied with demand response has increased from less than one-third of the requirement to nearly one-half. Second, installed wind capacity has also risen over the period which tends to increase the amount of on-line capacity that QSEs set aside for portfolio balancing. Third, on average, a larger share of OOME instructions are made in the down direction rather than up direction. QSEs that receive upward portfolio balancing instructions at the same time as downward OOME instructions may find it difficult to meet both at the same time. This could lead some QSEs to set aside on-line capacity for portfolio balancing purposes.

The increasing trend in non-offered energy is disturbing because it indicates that there is a substantial amount of energy that is not being efficiently utilized. It is of particular concern due to the more frequent price spikes that began to occur toward the end of 2004, which are discussed in Section IV in greater detail. As the surplus capacity in ERCOT dissipates and it becomes more important to fully utilize the existing generating resources, these patterns will have much more serious implications.

Non-offered energy can also raise competitive concerns since a dominant supplier could possibly attempt to exercise market power by strategically withholding this energy from the balancing energy market. To investigate whether this has occurred, Figure 39 shows the same data from the previous figure, but arranged by load level for daily peak hours in each year.[21]

[21] More precisely, available capacity and balancing-up offers were ascertained for the peak hour of each day during 2003 and 2004. This data was then separated by load level and the average capacity and average balancing-up offers were calculated.

Figure 39: Balancing Energy Offers compared to Available Capacity
Daily Peak Load Hours – 2003 and 2004



The figure indicates that in 2003, the average amount of capacity available to the balancing market increased gradually up to 50 GW of load and then declined at higher levels. The pattern was similar in 2004, although the available capacity declined more substantially above 50 GW. In 2003, the fraction of available capacity offered to the balancing market grew as load increased. However, in 2004, the fraction offered was more consistent across load levels.

Overall, the results indicate that there is a substantial amount of available balancing energy that is not offered, which increases the volatility of balancing energy prices in intervals when balancing deployments are relatively large. There could be a number of reasons for these results. First, the issues related to ramp rates discussed in the prior subsection can affect the offer levels. 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, compelling it to dispatch its high-cost resources at a loss.

In addition to this general issue related to the portfolio bidding framework, until June 2004, the SPD software had a specific issue that contributed to the lack of offers. Until June 2004, the SPD software assumed that a supplier could ramp back down to its schedule immediately (i.e.,

the recall ramp rate was not respected).[22] For example, if a portfolio deployment of 300 MW was made over 3 intervals, SPD may instruct the supplier to return to its schedule in the fourth interval. If the supplier can only ramp down 100 MW per interval, it will show an uninstructed deviation for two intervals. By offering what could be ramped in one SPD interval, a participant will minimize its risk of uninstructed deviations caused by infeasible recall instructions. This strategy is effective because it will ensure that the supplier's deployment will not be significantly larger than the amount that can be recalled in one interval. While the change in the software would logically lead to a greater volume of balancing energy offers, the data shown in Figure 38 does not indicate a noticeable change in total offers. However, it does show a decline in the volume of ramp-constrained offers.

Second, it is very difficult to offer gas turbines in the balancing energy market effectively. The available capability in Figure 38 includes quick-start offline resources, primarily comprised of gas turbines. If these quantities were eliminated from the figure, it would show that a much higher portion of the available balancing energy was offered. The current balancing energy market rules present significant challenges for owners of gas turbines due to timing and minimum run-time considerations. With regard to timing, the current balancing energy market rules do not provide adequate advance notice for some suppliers to reliably start gas turbines in response to the balancing energy market instructions. In addition, there is no assurance that prices in subsequent intervals will support the continued operation of the gas turbine.

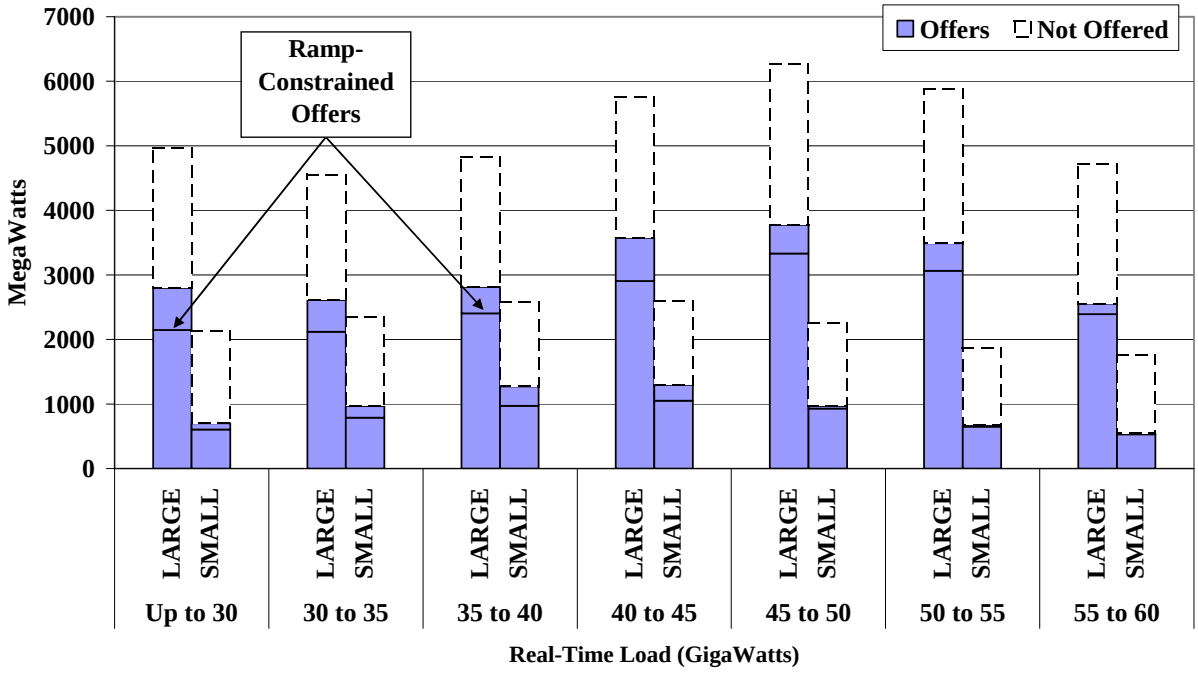
As shown above, balancing energy prices frequently spike in the first or last interval of an hour, before decreasing significantly. This could cause a supplier with a gas turbine that is satisfying its portfolio instruction to have to turn on the gas turbine for the one interval, and then keep it on for the rest of its minimum run time at a loss before it may shut down. Hence, it is understandable that some suppliers would not offer the energy that may be available from their gas turbines in the balancing energy market (or only offer it under higher load conditions when balancing energy prices could be expected to be sustained for multiple intervals).

[22] On June 2, 2004, an upgrade was made to ERCOT's balancing energy market software to enforce recall ramp rates, which should address this issue.

Third, the lack of offers could reflect withholding by ERCOT participants. Although this is possible, if it were the case, we would expect to see the percentage of available capability offered into the market decline in the highest load periods, because the incentive to withhold should be highest under peak demand conditions when the withholding would have the largest effect on prices. Nevertheless, the overall lack of offers is due to multiple factors and further analysis is needed to determine whether the lack of offers is due to withholding.

To provide additional insight regarding the possibility that the offer patterns may raise competitive concerns, we next examine whether large and small suppliers act in systematically different ways in terms of the available capacity they offer into the balancing energy market. If large suppliers offer less of their available capacity, particularly under peak conditions, that could be an indication of market power. Figure 40 shows the balancing up capability relative to the balancing up offers divided between large suppliers and small suppliers. The large suppliers category includes QSEs associated with the largest four owners of generating capacity in ERCOT, whereas all other QSEs are included in the "small" suppliers category.

Figure 40: Balancing Energy Offers versus Available Capacity in 2004
Large and Small Suppliers -- Daily Peak Load Hours



We emphasize that this analysis will not support a definitive finding regarding potential competitive issues related to these offer patterns. Depending on load and contractual obligations

or location, some small suppliers may have more market power than some of the large suppliers. However, the analysis does provide useful information to consider in conjunction with the results of Section VII in this report to determine whether additional investigation is warranted.

Figure 40 shows that available capacity from small suppliers (measured by the sum of the stacked bars) tends to decrease after actual load reaches 45 GW, whereas available capacity from large suppliers decreases after load reaches 50 GW. The amount offered into the balancing energy market by large participants trends upward until load reaches 50 GW while the amount offered by small suppliers starts to decline when load reaches 45 GW.

At the lowest load levels, large participants offered an average of 2,801 MW or 56 percent of their available capacity. At the highest load levels, large participants offered an average of 54 percent of their available capacity. Although this represents a reduction from 2003, it remains substantially higher than the amount offered by small participants and is consistent with the offer patterns at other load levels. The figure also shows that at all load levels, small participants offered an average of 30 percent to 50 percent of their available capacity. Both large and small suppliers offered only slightly more balancing energy than the portfolio ramp constraints allow to be deployed in a single interval.

Large participants are most likely to have market power and would generally have the largest incentive to withhold during periods of high demand. However, we observe from the figure above that large participants offer their generation at a higher rate than small suppliers during the highest demand periods. Since small suppliers generally do not have incentives to withhold, it is possible that the market design problems associated with ramping issues and portfolio bidding discussed earlier in this section affect small suppliers more than large suppliers. A more detailed analysis is needed before drawing conclusions about withholding behavior by either large or small suppliers. While investigating the specific conduct of large or smaller suppliers is beyond the scope of this report, the following figure summarizes the offer patterns of each of the largest four QSEs individually.

Figure 41: Balancing Energy Offers versus Available Capacity in 2004
Four Largest Suppliers -- Daily Peak Load Hours

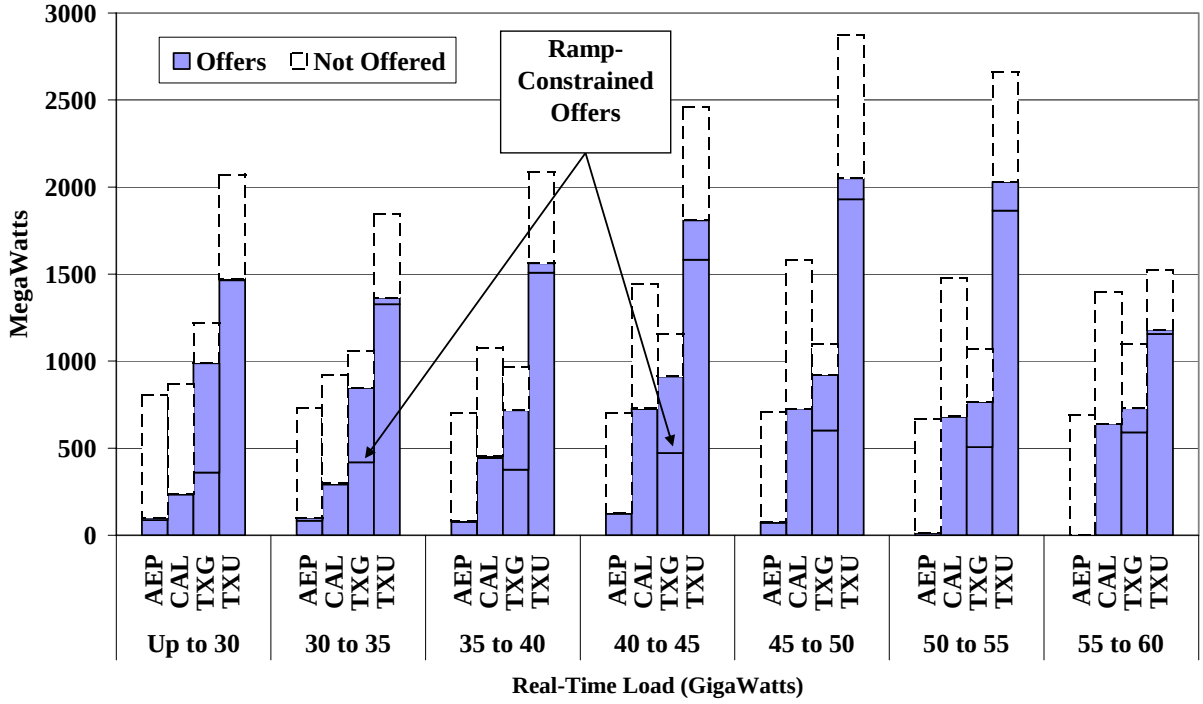


Figure 41 shows the information for each of the large suppliers included in Figure 40. TXU generally has the largest amount of available energy, with approximately 2 GW at low to moderate load levels, rising to 3 GW at high levels, However, its available energy decreases to 1.5 GW when load rises above 55 GW. Calpine exhibits a similar pattern of rising available energy as load rises, but it does not decrease above 55 GW. For TX Genco and AEP, the quantity of available energy is relatively consistent across the load levels.

TXU and TX Genco both generally offer 70 to 80 percent of their available energy to the balancing market, although TX Genco's offers are more significantly restricted by the portfolio ramp rate. Due to ramp constraints, generally only 30 to 40 percent of TX Genco's available energy is dispatchable at low load levels, and close to 50 percent is dispatchable at higher load levels. Calpine offers a smaller fraction of its available energy than TX Genco and TXU, ranging from 27 percent at the lowest load levels to close to 50 percent under most other load conditions.

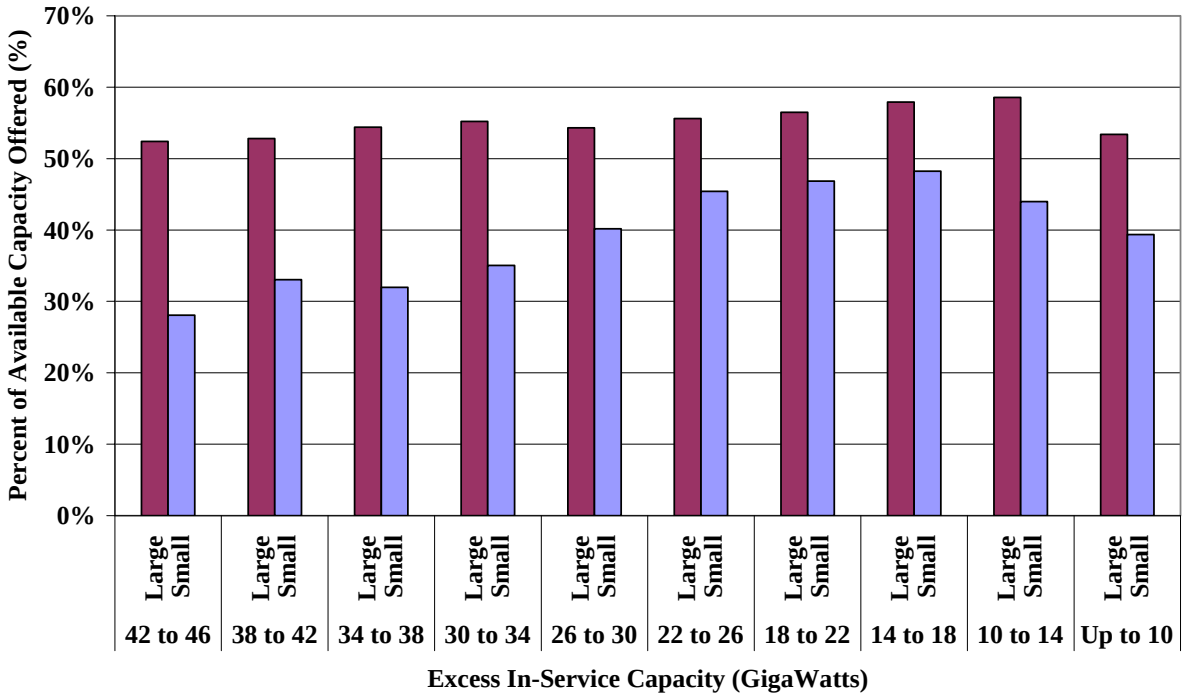
AEP offers a very small portion of its available energy into the balancing market, usually less than 20 percent. The provisions of AEP's Reliability Must Run ("RMR") contracts significantly

discourage them from offering into the balancing market and likely explain why AEP offers such a small amount of its available energy. According to the provisions of the RMR contracts, AEP can sell any excess (over its RMR obligation) into the balancing market, but AEP must return 90 percent of the profits. However, when these units generate unprofitably because of rapid changes in the MCPE, AEP is not shielded from the losses. Thus, the risks of participating in the balancing energy market likely exceed the potential profit for AEP. To improve the performance of the balancing energy market, we recommend that the RMR provision related to balancing energy profit be reconsidered.

If one of the four largest suppliers was withholding by not offering to the balancing market, we would expect them to offer less under the highest load conditions. Figure 41 does not indicate this pattern for any of the large suppliers. However, there are large amounts of unoffered capacity under all conditions by all suppliers that would tend to mask any withholding activity. Thus, without doing a more detailed analysis, it is impossible to draw definitive conclusions regarding whether any QSEs have strategically withheld capacity by not offering it in the balancing energy market.

We performed one final analysis of these offer patterns in which we calculate the percentage of available energy offered in the balancing energy market by large and small suppliers under different excess capacity conditions. The prior three figures examine offer patterns relative to actual load levels. However, volatile prices frequently occur in the spring and fall seasons at lower load levels. Tight conditions can occur during the shoulder months when a relatively large quantity of generation is out of service for planned maintenance. Like the peak conditions during the summer, tight conditions at other times of the year should be characterized by relatively low excess capacity levels. Figure 42 shows the results of this analysis for the largest four QSEs and all other smaller QSEs. The excess capacity metric shown in this figure is equal to the total available in-service capacity as shown in the day-ahead resource plans less the actual load and ancillary services requirements.

Figure 42: Ratio of Balancing Energy Offers to Excess In-Service Capacity Large and Small Suppliers – 2004



These results in Figure 42 confirm the prior results, indicating that the largest suppliers tend to offer a higher share of their available energy in the balancing energy market than smaller suppliers. Additionally, this ratio increases as the quantity of excess capacity in the market decreases (i.e., as system conditions become tighter). While these results do not provide an indication that large suppliers have withheld on-line capacity by not offering it to the balancing energy market, any withholding that occurred would tend to be masked by the large amounts of capacity not offered for other reasons. In addition, large suppliers may withhold in other ways which are examined in Section VII in greater detail.

The ratio of available energy offered by smaller suppliers is between 25 and 50 percent under most excess capacity levels, falling to just below 40 percent under the tightest market conditions. It would be worth investigating the factors that lead to these low offer levels, particularly among the smaller suppliers, because these offer patterns have a significant impact on the performance of the balancing market.

III. ANALYSIS OF RESOURCE PLANS

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.[23] 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. While it is possible for participants to submit unreliable information as part of a gaming strategy, they might also unintentionally submit unreliable information.

This section of the report analyzes the changes in the resource plans between the day ahead and real time, and evaluates the different reasons underlying the resource plan changes. The analysis shows two different types of resource plan changes: changes in the planned operating level and changes in the quantity of committed capacity. Both types of changes can be important for the reasons discussed later in this section.

[23] 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.

A. Summary of Resource Plan Changes

In this section, we summarize patterns in resource plan changes and discuss the impact that certain key factors have on these patterns. In particular, this section evaluates differences between the resource plans QSEs submit day-ahead and the most updated resource plans they submit before the operating period, which we refer to as the real-time resource plan. It is expected that QSEs will make changes to their resource plans that reflect changes in information between the day-ahead and the operating period. Market participants respond to changes in information differently depending on the size and composition of their portfolios. This subsection evaluates the role of the following key factors in explaining changes made to the resource plans.

- Changes in the load forecast
- Out-Of-Merit Commitments by ERCOT for reliability
- Plant technology
- Portfolio size and composition

1. Changes in the Load Forecast

Suppliers make commitment and scheduling decisions based on their bilateral contract obligations and predictions of market conditions. As the forecasted demand increases from the day ahead to real time, suppliers may be willing to commit additional higher cost resources, or purchase additional energy under flexible bilateral contracts.

For these reasons, we expect that changes in the load forecast will be a primary determinant of resource plan changes between day-ahead and real-time. Our first analysis in this subsection examines the correlation between changes in the load forecast and changes in total market-wide planned generation from the day-ahead to real-time on an hourly basis during 2004. The load forecasts used for this analysis are ERCOT's public day-ahead load forecast and the real-time load forecast that is used by ERCOT to balance supply and demand in real-time. While many market participants use other forecasting tools, the change in ERCOT's forecast from day-ahead to real-time will be highly correlated with the change for other forecasting tools. Thus, the change in the ERCOT load forecast provides a useful proxy for how most market participants expect demand to change from the day-ahead to the real-time. Figure 43 shows a scatter plot of

load forecast changes versus aggregate changes in planned generation indicated in the resource plans on an hourly basis.

In Figure 43, the x-axis shows the increase in ERCOT's load forecast from the day-ahead to approximately 30 minutes before real-time. Positive values show hours when the day-ahead forecast was low relative to real-time. The y-axis measures the increase in the total planned generation from the day-ahead resource plans to the final resource plans submitted before real-time. The figure shows a 45-degree line, which maps out the set of points where changes in the load forecast are perfectly matched by corresponding changes in the resource plans. The 45-degree line divides the figure into two areas. Area A, the region above and to the left of the 45-degree line, includes points where the planned generation increased more than the load forecast from day-ahead to real-time. Conversely, Area B shows the set of points where the planned generation increased less than the load forecast.

Figure 43: Change in Planned Generation versus Change in ERCOT's Load Forecast Hourly – 2004

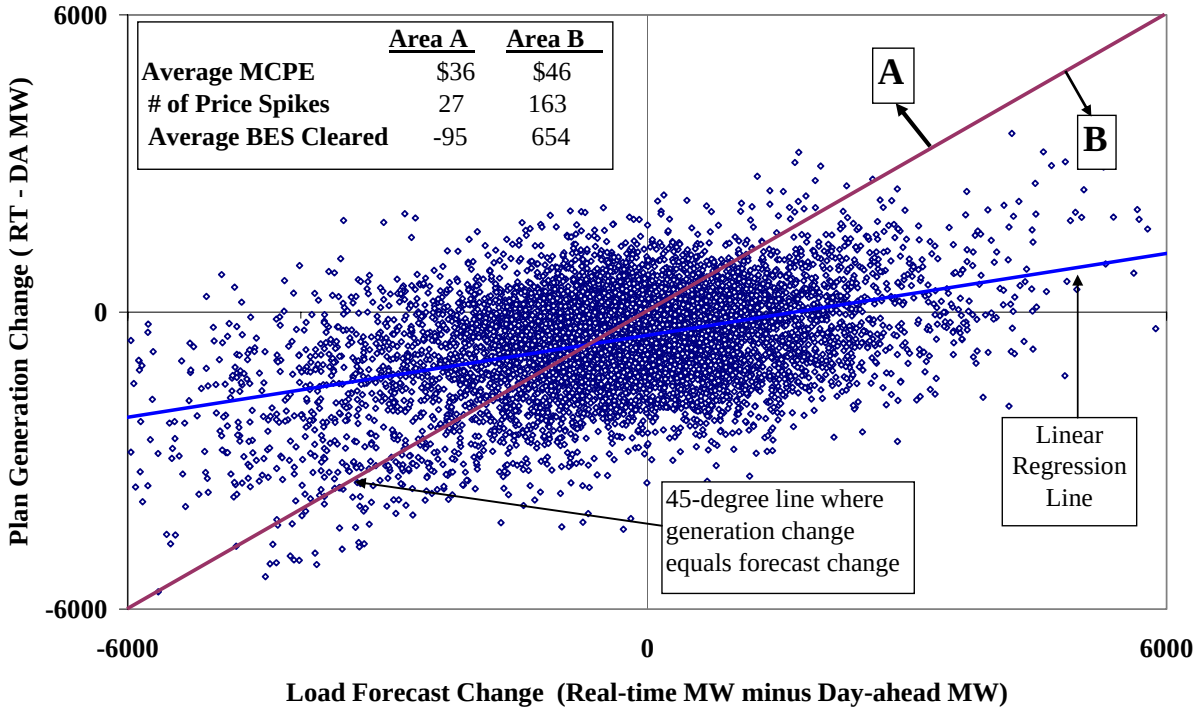


Figure 43 shows a wider dispersion of points in the horizontal direction than the vertical direction. This indicates that the load forecast changes more between the day ahead and real-time than the generation plans of market participants. The trend indicates a positive correlation

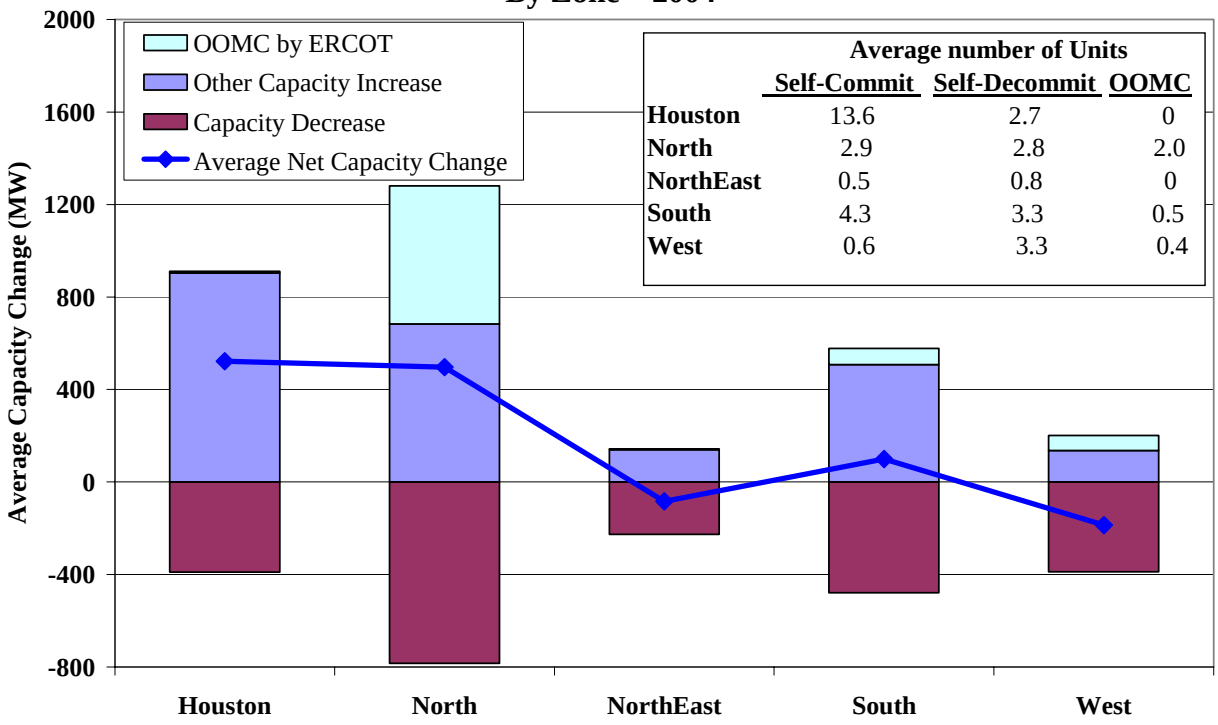
between forecast load changes and planned generation changes. However, the slope of the trend line is significantly lower than the 45-degree line. For each 1,000 MW increase in the load forecast, there is a corresponding 275 MW increase in planned generation. This suggests that QSEs tend to under-react when weather patterns and other load determinants change significantly in the hours leading up to real-time. Hence, when expected load rises significantly in the 24 hours before real-time, suppliers do not bring on enough additional resources to cover the increase. Similarly, when expected load decreases, excess generation tends to remain scheduled. The trend line in Figure 43 generally lies below the x-axis but then crosses at 1,700 MW, indicating that in general, when the planned operating level changes from the day-ahead to real-time, the change is in the downward direction.

Area A shows the set of hours when the planned operating level indicated in the resource plan increased more (or decreased less) than the load forecast. The table inside the figure confirms that in these hours, generation tends to be over-scheduled relative to load. Area B shows the hours when the total planned generation increased less than the load forecast from the day-ahead to the real-time. In these hours, generation was under-scheduled, resulting in an average of 749 MW net up balancing energy to clear the balancing market. The average MCPE for the hours in Area B is $10 per MWh or 28 percent higher than for Area A. This suggests that QSEs are generally not able to adjust their levels of planned generation as rapidly as necessary to adjust to changing expectations of real-time load. This underscores the need for the balancing energy market to operate efficiently since market participants will need to satisfy deviations between energy schedules and actual demand.

2. Commitments by ERCOT for Reliability

ERCOT relies on the QSE's resource plans to assess whether resource commitments are sufficient to maintain reliability. To the extent that additional capacity is necessary to maintain reserves in specific locations, the operator will commit units through the OOMC process, usually in late afternoon in the day ahead. The OOMC unit is required to be on line at its minimum generation level and to offer its remaining capacity into the balancing energy market. While this has a direct impact on the resource plans submitted by QSEs, it also indirectly affects commitment decisions for other resources. The following figure examines the average changes in committed capacity by zone for OOMC and non-OOMC resources.

Figure 44: Change in Committed Capacity from Day-Ahead to Real-Time By Zone – 2004



	Average number of Units		
	Self-Commit	Self-Decommit	OOMC
Houston	13.6	2.7	0
North	2.9	2.8	2.0
NorthEast	0.5	0.8	0
South	4.3	3.3	0.5
West	0.6	3.3	0.4

Figure 44 shows the average gross increases and decreases in committed capacity from day-ahead to real-time in each zone during 2004. The gross increases from OOMCs are shown separately from other commitment changes. The figure shows that Houston has an average of 912 MW commitment additions from the day-ahead to real-time that is partially mitigated by 389 MW being de-commited. The net change in commitment averages 8 percent of planned generation in Houston. The table in the figure above shows the average number of units that change commitment status from day-ahead to real-time: the average number of commitments is 13.6, while the average number of de-commitments is 2.7. This implies that Tx Genco tends to substitute gas turbines (on average, 65 MW units) for larger units (on average, 145 MW units) in its resource plan as it gets close to the operating period.

Figure 44 shows approximately 600 MW committed for OOMC purposes in the North Zone, where the majority of OOMC commitments occur. The North Zone also shows a large amount of de-commitments that are at least partially made by QSEs in response to OOMC commitments. The net change in commitment averages 4 percent of planned generation in the North Zone with

OOMC commitments, -1 percent without OOMC commitments. The West Zone showed a net de-commitment of 186 MW, or 9 percent of the average generation in that Zone.

The following figure summarizes the average changes in commitment on a monthly basis, which allows an examination of seasonal variation. It also shows net changes in planned generation, whereas the previous figure only showed net changes in committed capacity.

Figure 45: Change in Committed Capacity and Planned Generation by Month and Zone – 2004

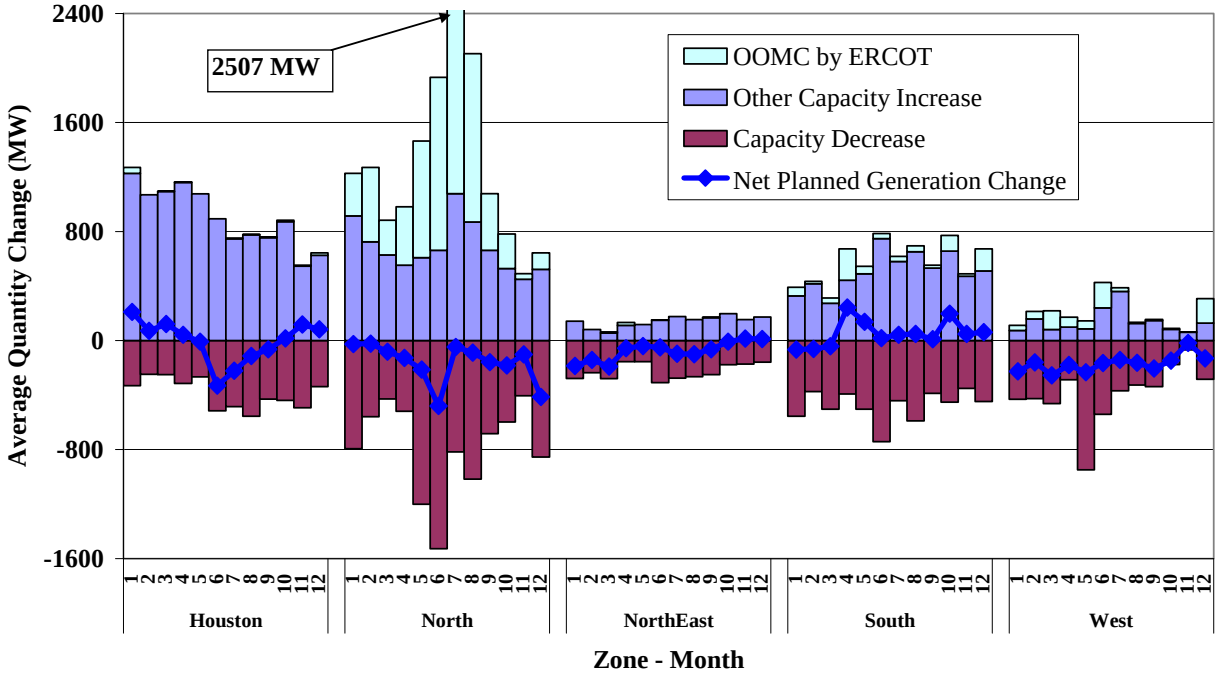


Figure 45 shows that resource plan changes occur in each zone under a variety of seasonal conditions. In Houston, an average of more than 1 GW was newly committed after the day-ahead during the first five months of the year, and this decreased to approximately 800 MW during the summer and less than 600 MW during November and December. There was also a small increase in de-commitments after the first five months of the year. The average change in net generation in Houston was close to zero at the beginning of the year, but dropped to -331 MW in June, and steadily increased thereafter. This generally matches the pattern of load forecasting (not shown here) which generally under-forecasted in the first five months of the year, but then began to systematically over-forecast starting in June. In June, the average daily

peak load forecast was more than 3 percent higher than actual load.[24] The North Zone shows a similar pattern of changes to net planned generation corresponding to the pattern of changes in the load forecast from day-ahead to real-time. However, the North Zone had a 414 MW decrease in net planned generation during December that is not observed in the other zones.

In the North Zone, the volume of commitment changes was significantly higher during the summer and moderately higher during the winter, than the spring and fall. These changes were partly driven by the large amounts of capacity committed through the OOMC process, as well as the de-commitments that are frequently made by QSEs to counter-balance those OOMC commitments. The Northeast Zone contains mostly baseload plants that are committed under most load conditions, and therefore their resource plan information changes less frequently than for generators that are closer to the margin.

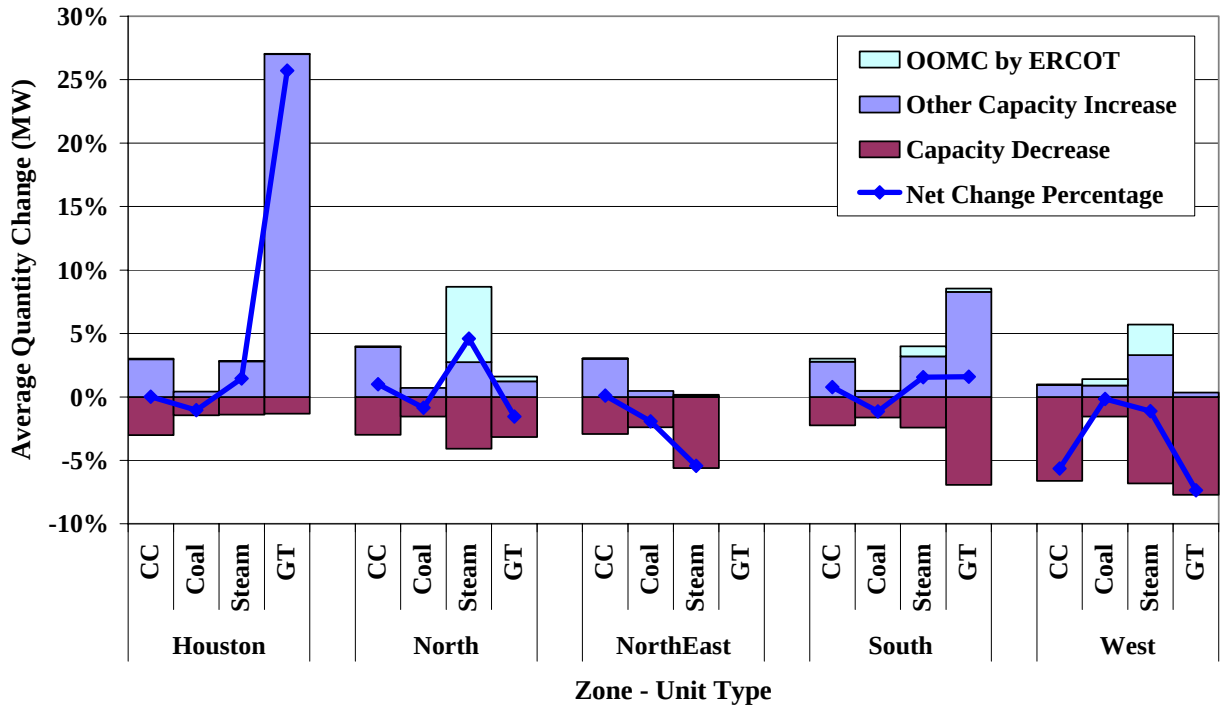
3. Plant Technology

Plant technology plays a large role in how resources are utilized by market participants. Some generators are highly flexible, with short start times and/or fast ramp rates, while other generators have inflexibilities such as narrow dispatchable ranges, long start notification times, and slow ramp rates. Furthermore, plant technology determines whether a unit has high or low operating costs. As market participants forecast market conditions, their response to those conditions is heavily constrained by the technology types in their portfolio.

Based on the demand forecast and expected on-line capacity, QSEs assess which of their units are likely to be economic for commitment. They are likely to cover forward contract obligations with their least expensive units and then commit higher cost units based on assessments of outage risk and balancing energy price forecasts, which can change up until real-time. Therefore, we expect that the pattern of resource plan changes is heavily dependent on the technology of resources. The following figure summarizes changes in commitment status by the technology of the resource, excluding non-fossil units.

[24] "2004 Assessment of the Operation of the ERCOT Wholesale Electricity Markets", Potomac Economics, November 2004.

Figure 46: Change in Committed Capacity versus Resource Technology By Zone – 2004



The figure above indicates that changes in commitment status for resources in Houston were consistent with other zones except for the gas turbine category. On average, 27 percent of the gas turbine capacity in Houston switched from non-committed day-ahead to being committed in the real-time. This explains most of the net increase in capacity that was observed for Houston in the previous two figures. The pattern for other fossil types in Houston is characteristic of ERCOT as a whole. Coal units exhibit the smallest volume of changes between day-ahead and real-time. Given the current level of natural gas prices, coal units are economic and run on virtually every day of the year, barring a forced outage.

Combined cycle units and non-coal burning steam turbines change commitment status more often than coal units, as expected since these units are frequently on the margin. It is more challenging for market participants to predict whether a combined cycle or gas-fired steam turbine will be economic at prevailing spot prices than for coal units.

Gas turbines exhibit the largest variation in commitment status of any fossil technology in Houston. Gas turbines also exhibit large changes in the South Zone relative to other areas and technology types. In the North Zone, the commitment status of gas turbines changes very little,

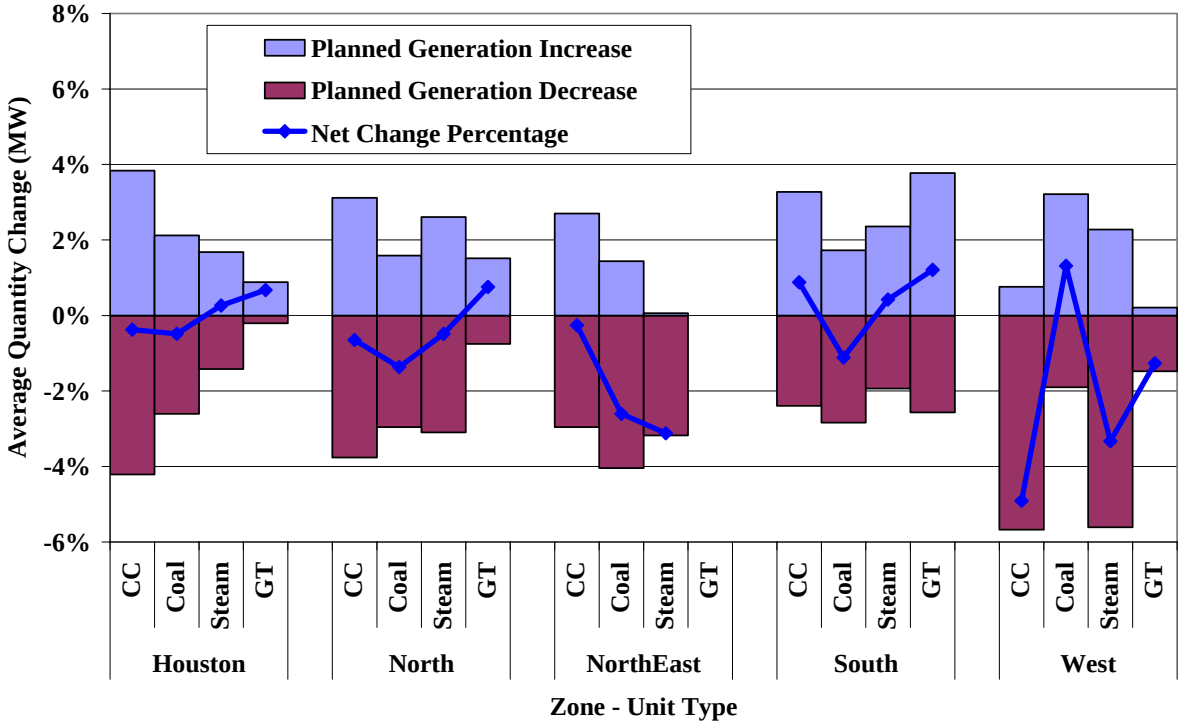
while in the West Zone, more than 7 percent of the gas turbine capacity is de-committed after the day-ahead. The changes in committed capacity associated with the gas turbines are not unexpected, given the costs and operating patterns of these units. They are generally the most expensive units on the system and least likely to be economic to dispatch. In fact, the need to dispatch gas turbines often arises within the day in response to a significant outage or unexpected increases in load. The changes in commitment status for gas turbines have less impact on capacity adequacy than for other technology types, because gas turbines can be started very quickly while other types are less flexible.

Figure 46 shows that nearly all OOMC commitments occur with gas-fired steam turbines. Generally, OOMCs occur when an import-constrained local area is capacity deficient. When a market participant receives an OOMC instruction for one of its gas-fired steam units, it often uses this to replace another unit in its portfolio that it had planned to commit, usually another steam unit. Thus, OOMC instructions lead to commitments and de-commitments of steam units in the North Zone, and to a lesser extent in the West Zone. In some cases, when a supplier receives an OOMC instruction, it may de-commit a resource in another zone if inter-zonal congestion between the two zones is not anticipated, and adjust its energy schedule accordingly. Therefore, the net de-commitment of steam units in the Northeast Zone may be related to OOMC instructions in other zones.

In many cases, it is economically efficient for a QSE to de-commit a resource outside a locally constrained area after receiving an OOMC instruction. However, this pattern raises concerns about reliability to the extent that de-committing units leads to a local or ERCOT-wide capacity insufficiency. For instance, if ERCOT gives an OOMC instruction for a market-wide capacity insufficiency and this is followed by de-commitment, it defeats the purpose of the OOMC. Similarly, an OOMC instruction to maintain reliability in a local area could be followed by a de-commitment that counteracts the purpose of the OOMC. While adverse effects of the de-commitments can occur unintentionally, it is also possible for QSEs to strategically revise their resource plans to cause capacity insufficiencies to compel additional OOMC instructions. This is beyond the scope of this report, although a subsequent report will investigate the market results for this type of strategic resource plan revisions.

The pattern of changes in planned generation is somewhat different from the pattern of commitment changes. The summary of planned generation changes is shown in Figure 47.

Figure 47: Change in Planned Generation versus Resource Technology By Zone – 2004



In Houston, the figure shows only small net changes in planned generation levels for each type of fossil unit. In the case of gas turbines, we have seen that 27 percent of gas turbine capacity switched from non-committed to being committed after the day-ahead. However, this does not translate into more planned generation for these gas turbines because they are being committed with a planned generation level of 0 MW, a convention indicating to the dispatch model that a quick start resource is available to be deployed in the balancing market. Figure 46 shows that the change in planned generation levels is relatively large for combined cycle and coal units. These units are most frequently committed, and they reflect most of the variation in planned generation levels that is correlated with fluctuations in the load forecast in the hours leading up to real-time. Overall, the planned generation level decreases from the day ahead to the real time resource plan. The Northeast and West Zones in particular show significant decreases in planned generation levels after the day-ahead resource plan.

4. Portfolio

In addition to the factors listed above, we expect the size and composition of a supplier's portfolio to affect the pattern of resource plan changes by the supplier. Based on the preceding figures, short-term fluctuations in the load forecast have the strongest impact on resource plan changes. The following figure compares net changes in commitment and planned generation to changes in the ERCOT load forecast from day-ahead to real-time for the two largest QSEs compared with all other QSEs.

Figure 48: Change in Planned Generation and Committed Capacity versus Load Forecast Changes By Supplier – 2004

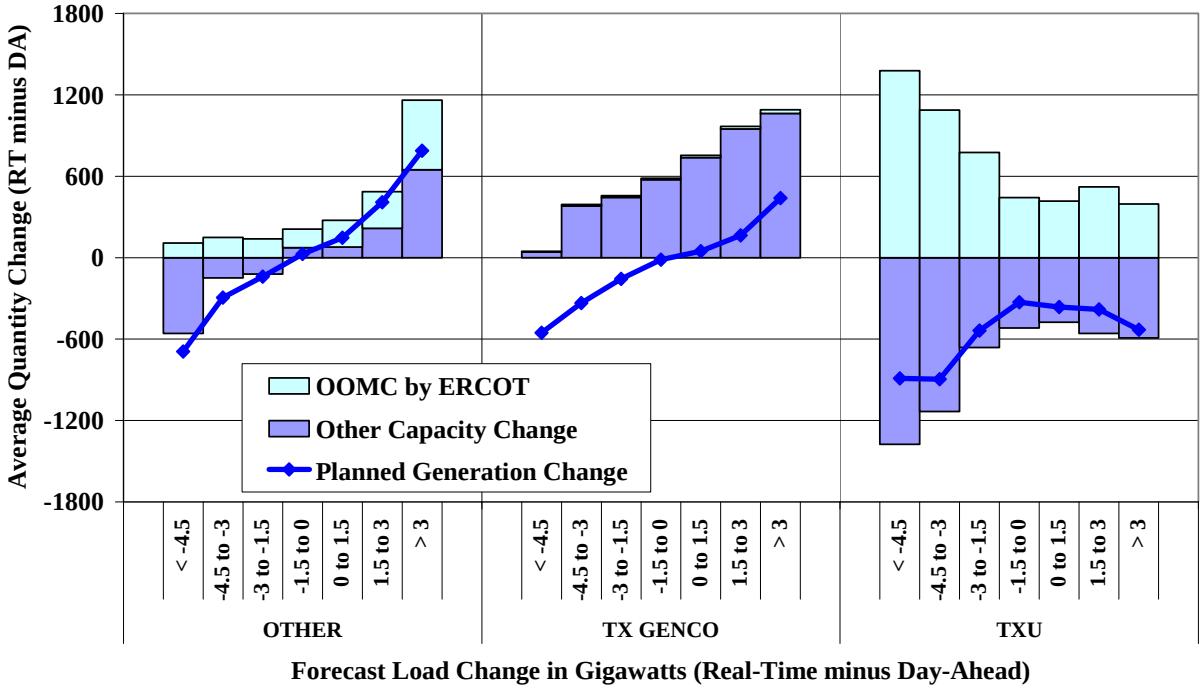


Figure 48 shows the average aggregate resource plan changes for all suppliers other than TXU and TX Genco compared with changes in the load forecast after the day-ahead. Figure 43 indicated a strong positive correlation between changes in planned generation levels and load forecast changes for all suppliers. This is also true for the "other" suppliers in Figure 48. This figure shows that during hours when the load was under-forecasted in the day-ahead by more than 3 GW, the "other" suppliers responded with an average increase in net planned generation of 787 MW. Commitment changes for "other" suppliers also show an increase in real time when the load was under forecasted in the day ahead.

TX Genco adjusts planned generation after the day-ahead in a manner similar to "other" suppliers. During hours when the load was under-forecasted day-ahead, TX Genco showed additional planned generation in the adjustment period, and when load was over-forecasted, they responded with less planned generation. TX Genco also responded to under-forecasted load in the day ahead with additional unit commitment in the adjustment period, in addition to the overall increase in gas turbine commitment mentioned in previous sections. Generally, the resource plan changes for TX Genco and the group of small suppliers are consistent with incorporating new information about demand conditions.

The pattern in Figure 48 for TXU is significantly different from that of "other" suppliers and TxGenco. TXU resources made up the vast majority of OOMCed capacity, and the average de-commitments came close to netting out the OOMCed capacity. The figure suggests that when a resource is brought on line through the OOMC process, TXU de-commits another resource to compensate. The figure shows that TXU significantly decreases planned generation from the day-ahead to the real-time resource plan. This may be because OOMC units come on at their minimum output level while the resources that are de-committed were generally scheduled at a higher, more efficient operating point. It is beyond the scope of this report to examine the adverse affects of these de-commitments, although this will be analyzed in a subsequent report.

The six figures above show that the overall pattern of resource plan changes is consistent with efficient scheduling behavior by market participants, although de-commitments in response to OOMC instructions raise some operational concerns. In some cases, the de-commitment negates the reliability benefits of the OOMC instruction, leading ERCOT to give additional OOMC instructions to maintain reliability. These de-commitments also raise strategic gaming concerns since a market participant could intentionally de-commit units that would later need to be given OOMC instructions. While the analyses in this sub-section point to general conclusions about the revision of resource plans by market participants, we will analyze these patterns in greater detail in a subsequent report.

The current market design has a de-centralized process for unit commitment that leads to reliability problems that must be addressed by OOMC instructions. Furthermore, capacity shortages and surpluses arise when changes in the load forecast are not met by corresponding

changes in committed capacity. These issues could be best addressed through a centralized commitment process, such as the one that has been discussed as a provision of Texas Nodal.

B. Resource Plans and Out-of-Merit Commitments

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, it may do so at no cost since the plans are not binding. In this subsection, we analyze market participants' resource plans to evaluate whether the market protocols may provide incentives for such strategic conduct. Specifically, we evaluate units that are frequently committed out-of-merit or frequently dispatched out-of-merit. Such units receive additional payments from ERCOT and we investigate whether market participants may engage in strategies to increase these payments.

We first analyze the behavior of suppliers that are the primary recipients of payments by ERCOT for out-of-merit capacity. OOMC occurs when ERCOT 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 instructions, they receive payments from ERCOT that are based on an estimate of the cost of starting the unit plus an amount to contribute to the estimated costs of running at minimum level. However, the balancing energy sales revenues are retained by the supplier. Therefore, if a unit is 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 OOMCed.

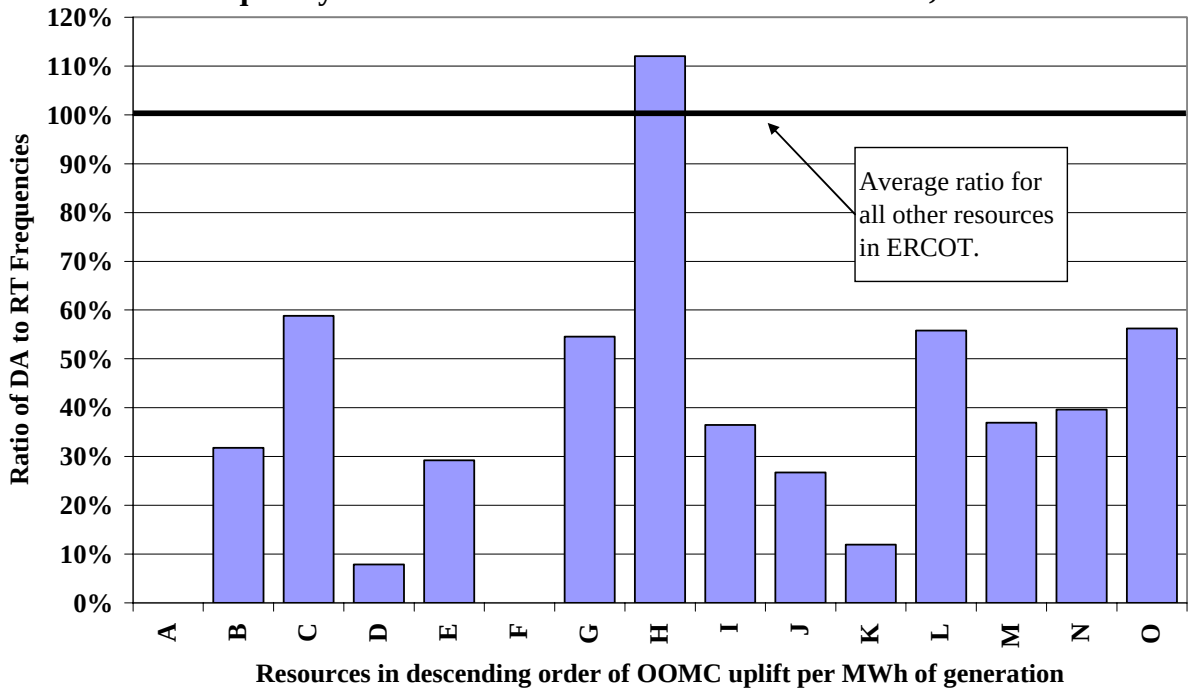
A substantial improvement was made to the OOMC incentive structure at the end of February, 2004. Prior to the improvement, the OOMC payment was the same regardless of any sales made by the supplier in the balancing market. Therefore, market participants always earned more revenue when their units were committed through the OOMC process by ERCOT. If the unit did not receive an OOMC instruction, it did not forego any revenues in the real-time energy market because it could still be committed in a subsequent resource plan before real-time. This was generally a risk-free method to attempt to receive additional revenue through the OOMC process.

However, in February 2004, the compensation formulas were changed so that revenues in the balancing energy market would be used to reduce the OOMC payment. Thus, market participants should no longer have as strong a disincentive to commit units that are needed for reliability.

Because of the incentives presented by the OOMC process, 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 investigated whether this was less prevalent under the new compensation formulas used since February, 2004. We examined the patterns of commitment for units that receive substantial OOMC payments. Figure 49 shows the ratio of day-ahead resource plan commitments to actual real-time commitments under the new compensation rules during 2004 for the 15 resources receiving the largest OOMC payments per MWh of production.[25] This should identify the resources with the largest incentives to engage in this strategy. Hours when the resources are under OOMC 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 on a per MWh basis—from $55 per MWh of generation across all hours for the units on the far left to $11 per MWh for the units on the far right. 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.

[25] We exclude resources that received payments that total less than $10 per kW-year of capacity or averaged less than $10 per MWh of generation for the period where data was available (March to October, 2004).

Figure 49: Ratio of Day-Ahead to Real-Time Resource Plan Commitments*
Frequently OOMCed Resources – March to December, 2004



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

Of the 15 resources shown in Figure 49, 14 have ratios of less than 100 percent, ranging from 0 percent to just under 60 percent.[26] Only one resource had a ratio over 100 percent. In contrast, the average ratio for all other units is 100 percent, reflecting a much higher consistency between the day-ahead and real-time resource plans. The results shown in this figure are consistent with the concern that some QSEs generally wait until after the OOMC process to commit units that are necessary for local reliability, even after the improved compensation formulas were implemented. Furthermore, this is consistent with our findings from 2003.[27]

For the resources shown in Figure 49, uplift payments for OOMC commitments are substantial enough to provide significant incentives to behave in ways that maximize the likelihood of receiving them. Figure 49 suggests that QSEs with resources that frequently receive OOMC instructions regularly delay the decision to commit those units until after ERCOT determines which resources to select for OOMC. This approach to address capacity insufficiency in the

26 Additional information on each of these resources is shown in Appendix A.

27 See 2003 SOM Report, p. 68.

Protocols has several deleterious effects on the market. First, ERCOT incurs OOMC 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 instruction and often does so. Third, this conduct tends to make unreliable 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 a centralized commitment process, such as the one that has been discussed as a provision of Texas Nodal.

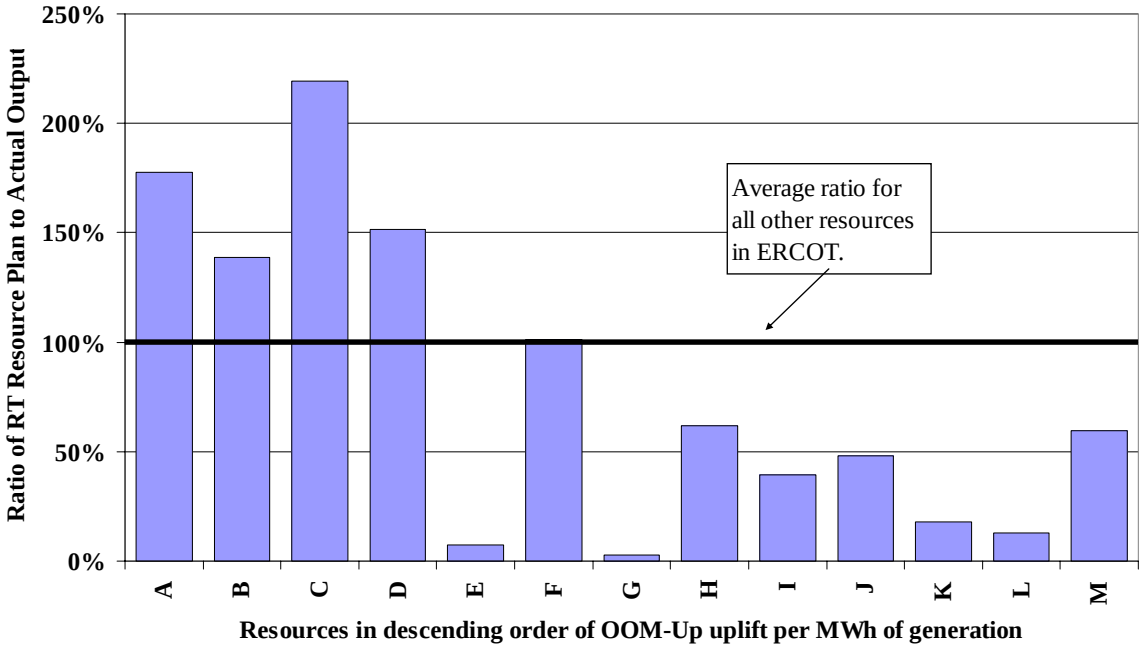
In our next analysis, we evaluate incentive issues associated with out-of-merit dispatch in real-time. In order to resolve intrazonal congestion in real-time, ERCOT will increase or decrease a unit's output (out-of-merit energy or "OOME") to reduce the flow on a constrained transmission facility within 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.[28] 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.

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

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 50 shows the ratio of real-time resource plan scheduled output to actual generation for the 13 units that received the highest average payments (per MWh) for OOME up.[29]

Figure 50: Ratio of Real Time Planned Generation to Actual Generation*
Frequent OOME up Resources – March to December 2004



* 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 13 resources shown in Figure 50 are presented in decreasing order of average payments, from $21 per MWh of generation across all hours for the unit on the far left to $4 per MWh for the unit on the far right.[30] 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 13 resources shown in Figure 50, 8 have ratios of less than 100 percent, ranging between close to zero percent to 55 percent. Five of the resources shown in the figure above operate at lower output levels in real

29 To focus on the most significant units, the analysis excludes resources where total uplift was less than $2 per kW-year of capacity or the average was less than $2 per MWh of generation for the period where data was available (March to October, 2004).

30 Additional information on each of these resources is shown in Appendix A.

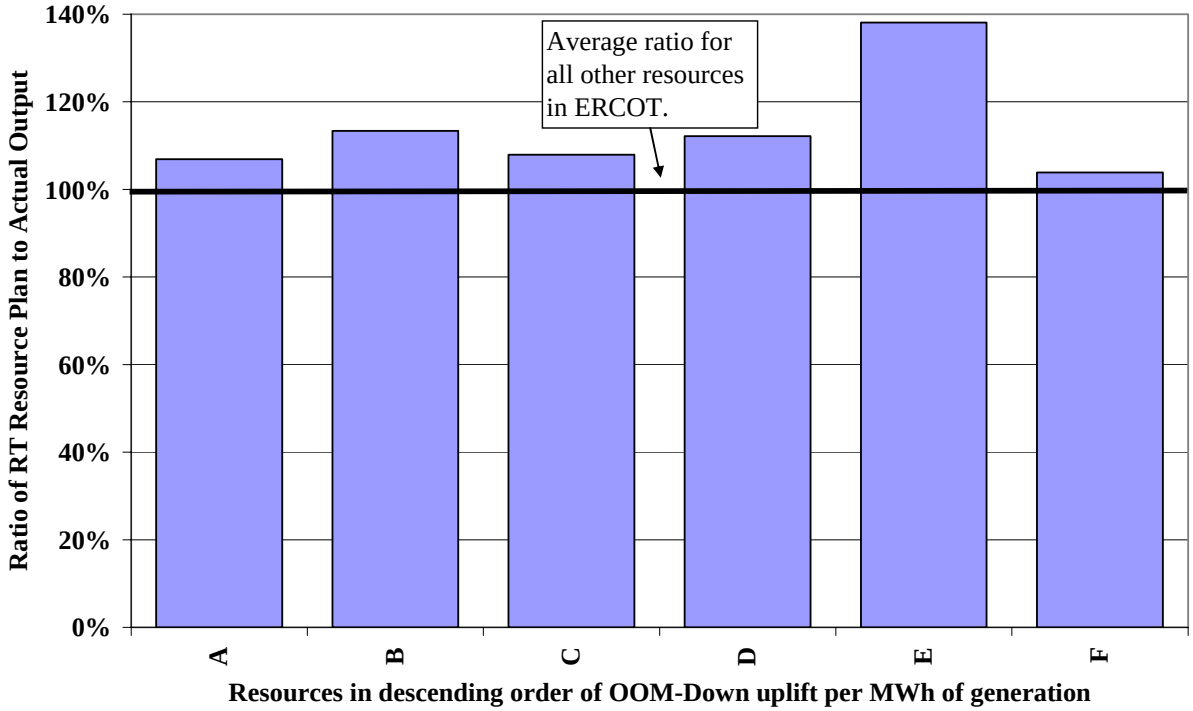
time than they were scheduled to run in the real-time resource plan, although four of them, units A through D, are located at one plant and operated by a single entity. Thus, A through D may be anomalous, suggesting that the OOME process still provides significant incentives to submit systematically low planned generation in real-time resource plans. The other units in ERCOT had a ratio of 100 percent during the period, reflecting, on average, consistency between the scheduled output and actual generation. The data suggests that resources frequently providing OOME up are regularly included by the QSEs in the real-time resource plans at output levels that are significantly 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 51 shows the ratio of real-time resource plan output to actual generation for nine select resources that earned the highest average payments for providing OOME down (on per MWh basis) in 2004.[31]

[31] This analysis excludes resources with uplift payments totaling less than $1 per kW-year of capacity or averaging less than $1 per MWh of generation for the period where data was available (March to October, 2004). This analysis also excludes cogeneration and renewable resources.

Figure 51: Ratio of Real-Time Planned Generation to Actual Generation*
Frequent OOME down Resources – March to December, 2004



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

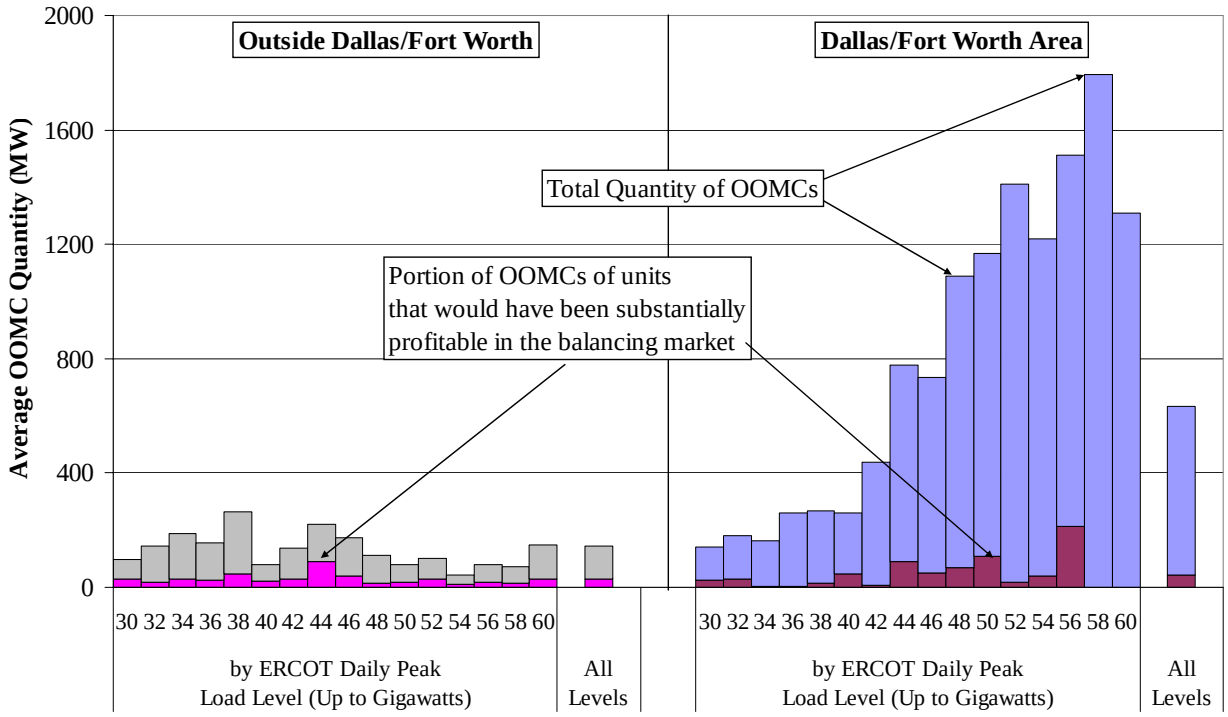
Figure 51 shows the six units that received the highest OOME down payments for their total production. The six resources are shown in decreasing order of the average OOME down payments received per MWh of output, ranging from $2.45 per MWh on the far left to $1.18 per MWh on the far right.[32] For comparison purposes, the figure also shows the generation-weighted average ratio of real-time resource plan output to actual generation for all other units.

All six of the resources shown in Figure 51 have ratios above 100 percent, ranging from 107 percent to 138 percent. This is in contrast to the average ratio exhibited by other units in ERCOT of 100 percent during the same period. While this reflects much better consistency between the planned output level and actual generation for OOME down than OOME up units, the average resource plan planned output level is still clearly higher than the average actual output for these six units. This is consistent with frequent OOME down units systematically over scheduling their resources.

[32] Additional information on each of these resources is shown in Appendix A.

Finally, we conducted a further analysis of the local congestion and out-of-merit patterns in the Dallas/Ft. Worth area. The transmission constraints into the Dallas/Ft. Worth area are the most significant local constraints in ERCOT by most measures. Figure 52 shows two panels, one for Dallas/Ft. Worth and one for all other areas in ERCOT. Each panel shows the average quantity of OOMC relative to the peak demand levels. The figure also reports the portion of OOMC that would have been substantially profitable to self-commit based on estimated start-up costs, minimum generation costs, incremental costs, and minimum run times.[33]

Figure 52: OOMC Supplied vs. ERCOT Load Level
Dallas/Fort Worth and Other Areas, March to December, 2004



This figure shows that on average ERCOT commits approximately four times more capacity out-of-merit in Dallas/Ft. Worth than all other areas. The figure also shows that as the demand in Dallas/Ft. Worth rises, operators must take more out-of-merit actions to maintain reliability. In

33 Profits are considered to be substantial if they would exceed the estimated minimum commitment costs of the unit by a margin of at least 50 percent. Continuous Emissions Monitoring (CEMS) data, collected by the Environmental Protection Agency, is used to estimate incremental heat rates and heat input at minimum generation levels. We also assume $4 per MWh variable operating and maintenance expenses. Whenever CEMS data is unavailable, minimum generation and incremental costs are estimated from a sample of balancing energy prices that coincide with each resource's production over the previous 90 days.

contrast, there is no clear relationship between OOMC quantities and demand levels outside Dallas Ft. Worth.

Our previous analysis of resource plan changes between the day-ahead and real-time shown in Figure 49 indicates that units frequently committed out of merit are often voluntarily committed when ERCOT does not provide an OOMC instruction. This raises concerns about QSEs having the incentive to delay commitment decisions in order to garner OOMC payments. However, Figure 52 indicates that both inside and outside Dallas/Ft. Worth, a very small percentage of resources receiving OOMC instructions would clearly have been economic for the QSEs to self-commit. This suggests that the incentive to delay commitment decisions may be limited to periods where the resource would have been only marginally profitable.

These analyses indicate that the current procedures for OOME and OOMC provide incentives for participants to submit resource plans that do not reflect anticipated real-time operations. One change we recommend to the current markets that would mitigate these issues would be to create a zone for Dallas/Ft. Worth. This would allow a large share of the congestion that is currently managed with out of merit actions to be priced more efficiently and transparently. It would also provide superior economic signals to guide investment in generation and transmission in that area. Lastly, if ERCOT were to move to a nodal market design, creating this zone would ease the transition to nodal markets where all congestion would be reflected in locational clearing prices.

We understand, however, that there would be significant issues to consider in forming such a zone, including the effect on current bilateral contracts, the need for measures to effectively mitigate market power in the area, and the equity implications of such a change. In addition, the benefits described above assume that CSCs between Dallas-Fort Worth and adjacent areas could be defined that include the key transmission constraints that currently result in OOME and OOMC actions by ERCOT. This would need to be analyzed and validated.

A comprehensive solution for all of these issues would be to implement a properly structured nodal electricity markets. Such nodal markets would virtually eliminate the need to commit and dispatch resources out of merit. Such markets would substantially improve the efficiency of the management of local congestion, as well as the management of interzonal congestion as

discussed in detail in Section VI below. Hence, we strongly encourage the continued development and adoption of the Texas Nodal markets that are currently under consideration.

IV. SHORTAGES IN THE BALANCING ENERGY MARKET

In this section, we analyze the balancing energy market outcomes during periods when ramp-capable balancing energy offers are not sufficient to meet balancing energy demand. We refer to these instances as balancing energy shortages. The performance of the balancing energy market during shortages is a critical aspect of the overall efficiency of the ERCOT market. When there is insufficient supply to serve the energy and ancillary services demand in the system, the value of all energy produced by suppliers in ERCOT (or imported) is extremely high. Accordingly, the market should produce economic signals during these periods that reflect this value.

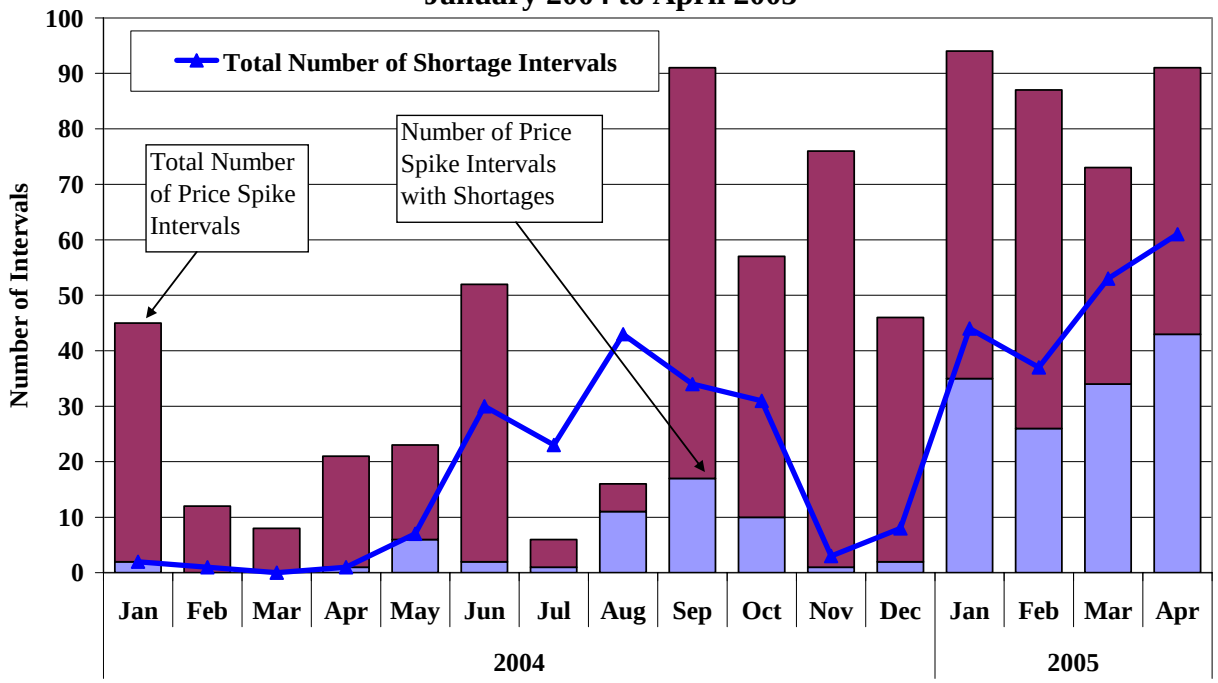
In general, spot markets like the ERCOT balancing energy market experience a shortage when the total supply offered to the market is insufficient to meet system demand for energy, responsive reserves, and regulation. Unexpected outages during peak load hours often contribute to these shortages. Shortages can also occur when the quantity of committed resources is insufficient due to unexpectedly high real-time load (i.e., when the load forecast error is large). In such situations, market operators frequently choose to hold less reserves or regulation in order to make more energy available to the energy market and "keep the lights on".

Other markets have shortage pricing provisions that cause energy prices to reflect the economic value of the reserves or regulation that are sacrificed to supply energy. These types of pricing provisions can improve the efficiency of the economic signals provided by the market. In ERCOT, however, shortages can also occur due to insufficient offers when there is more than enough on-line capacity because suppliers are not obligated to offer this energy in the balancing energy market. This characteristic of the ERCOT market makes it difficult from a pricing perspective to accurately discern true shortages.

A. Price Spikes and Shortages in the Balancing Market

The following analysis summarizes the coincidence of shortage conditions and price spikes in the balancing energy market. We define price spikes as balancing energy prices that are higher than 18 MMbtu per MWh times the spot price for natural gas (i.e., usually greater than $100 during 2004). This analysis is shown in Figure 53.

Figure 53: Total Number of Price Spike Intervals and Shortage Intervals January 2004 to April 2005



The bars in the figure show that the number of price spikes has increased significantly since September 2004. Indeed, price spikes occurred in approximately 2.5 percent of the balancing market intervals during the period from September 2004 to April 2005. The bottom portions of the bars in the figure indicate intervals with shortages in the balancing energy market. During 2004, a relatively small share of the price spikes occurred during shortage intervals. However, nearly half of the price spikes occurred during balancing market shortages in early 2005.

The line in the figure above shows the number of shortages in each month, whereas the bottom bar shows the portion of these intervals when a price spike occurred. Thus, the difference between the line and bottom bar is the number of shortages when no price spike occurred. Balancing energy shortages frequently did not cause price spikes during the summer of 2004. However, price spikes occurred during the majority of balancing energy shortages in 2005.

In order to further analyze this issue, Figure 54 shows the available capacity that could have been offered, but that was not offered during intervals when a shortage occurred. This includes capacity that is flagged in the resource plan as on-line or capable of being started quickly[34] that is

[34] During the study period for this analysis, QSEs could indicate that specific off-line resources were

not already scheduled for energy or ancillary services. However, this analysis excludes excess capacity on wind turbines and other renewable resources, units that are constrained down for local congestion, and resources in zones that are dispatched down for inter-zonal congestion.

Figure 54: Excess Unoffered Capacity During Shortages versus the Number of Shortages January 2004 to April 2005

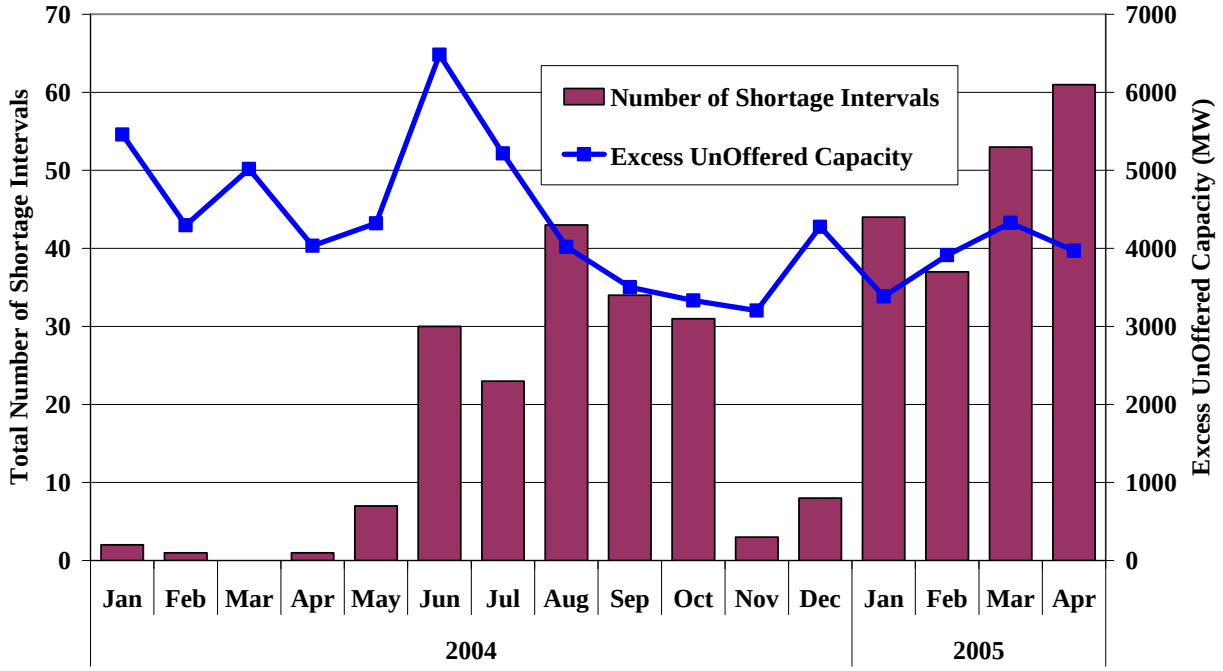


Figure 54 shows that during balancing energy market shortages, the amount of unoffered available capacity in the balancing energy market averaged more than 3 GW in each month. Even as the number of shortage intervals increased in early 2005, the amount of unoffered capacity during these intervals did not decrease. If all of the available capacity were offered to the balancing energy market, the number of shortages in the balancing energy market would have been reduced substantially and eliminated in most months.

When supply is not sufficient to meet demand in real-time, this can lead to emergency operating conditions such as load shedding, significant frequency deviations, and responsive reserves deployments. However, the frequent balancing energy market shortages have not compelled ERCOT to take emergency actions. Instead, when there is a shortage in the balancing energy market, ERCOT typically deploys unutilized resources through Verbal Dispatch Instructions

available to the balancing market by setting their status flag to “on-line” and their planned generation levels to 0 MW.

("VDIs") and deploys regulating units up to make up the difference. While these actions keep the lights on, they occur outside the market and are not economically efficient. In addition, such out-of-market actions will not be reflected in the market prices.

There are persistently large amounts of excess capacity not offered to the market during shortage intervals. This indicates that resources are not being deployed efficiently in real-time to meet demand for energy and ancillary services. This also suggests that the price signals generated in the balancing energy market are not efficient. Therefore, it is particularly important, given the recent growth of shortages and price spikes, to address the aspects of the current market design that discourage participation in the balancing market and inhibit full utilization of available capacity.

ERCOT is scheduled to implement a market for replacement reserves during June, 2005. A replacement reserves market procures additional capacity when insufficient capacity is anticipated. This type of market is designed to address certain types of shortages, and the following sub-section discusses the likely impact that it will have on market outcomes. In particular, it discusses the likely impact on the frequency of balancing market shortages.

B. Replacement Reserves Market

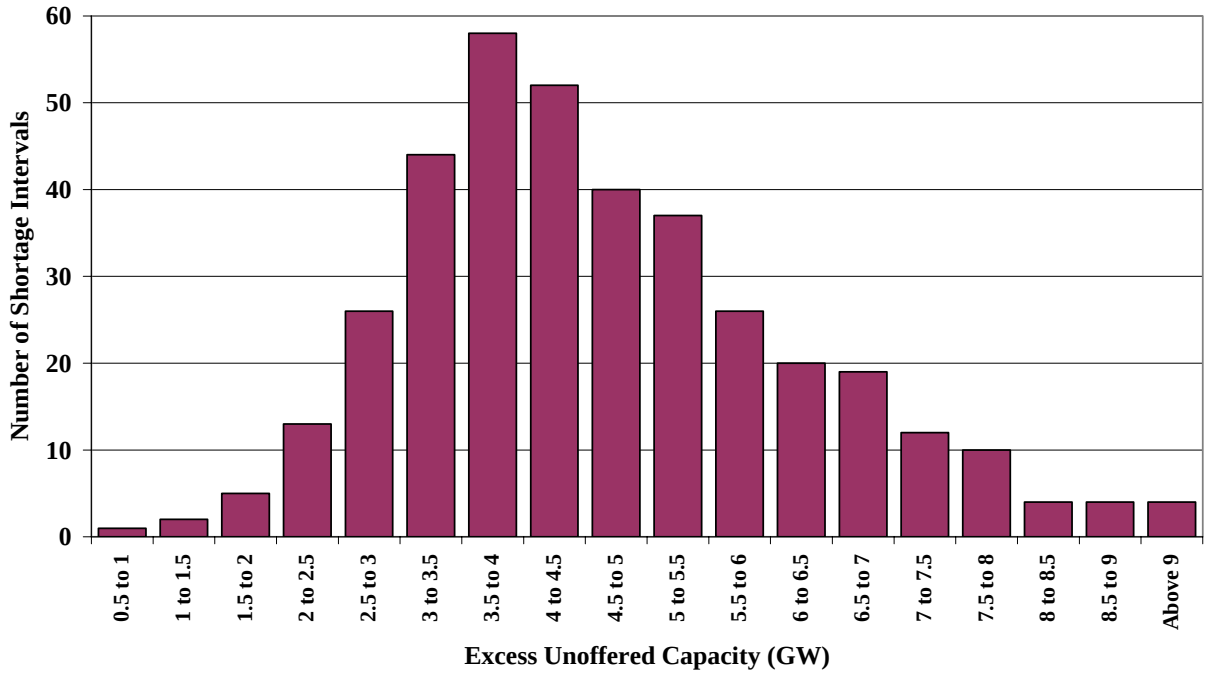
During 2005, a market for replacement reserves will be reinstituted to address capacity shortages anticipated on the day before real-time. After the Resource Plan validation in the afternoon of the day ahead, the replacement reserves market model will evaluate whether additional capacity is necessary. It does this by comparing the total capacity in the resource plans submitted by market participants with the forecasted load and regulation and reserves requirements. If the capacity in resource plans of generators that plan to be on-line is insufficient, the market will procure sufficient capacity to cover the anticipated needs from the least expensive resource(s) available.

Likewise, if capacity is not sufficient in certain areas because of transmission constraints, the model will procure capacity in specific zones when a binding CSC is anticipated or from specific resources to address local constraints. Being selected for replacement reserves requires the resource to be committed and available to be dispatched up in the balancing market. It is

important to recognize that the replacement reserve market addresses shortages of *capacity*, not shortages of balancing energy caused by a lack of offers.

In the previous section, we identified that during shortage intervals, there is generally a significant quantity of excess un-offered capacity. The following analysis examines the distribution of available un-offered capacity not being used to satisfy energy and/or ancillary services needs to assess the frequency of capacity insufficiencies.

Figure 55: Excess Un-offered Capacity Compared to Number of Shortage Hours January 2004 to April 2005



The figure above shows the frequency of balancing market shortages over a 16 month period according to level of un-offered capacity. Approximately 88 percent of the shortages occurred when at least 3 GW of on-line capacity was not offered to the market, and there was only one interval where less than 1 GW was un-offered. If the un-offered capacity were made available to the balancing market, it is possible that some of it would have been unutilized due to congestion or ramp constraints. However, it is unlikely that shortages would have occurred in many of the intervals shown above. This suggests that there were very few, if any, periods of authentic shortage during the 16 months shown above, and that these were likely due to a shortage of ramp capability and other transient conditions rather than inadequate on-line capacity.

If the replacement reserves market is initialized to procure capacity when more is needed on-line to meet energy and ancillary services requirements for the next day, the market will not anticipate the shortages that occur from a lack of offers to the balancing energy market. Thus, we would not anticipate a substantial reduction in balancing market shortages after the implementation of the replacement reserves market. However, the replacement reserves market could be initialized to ensure sufficient capacity to meet energy and ancillary services needs plus the estimated amount of un-offered capacity. While this would reduce the quantity of balancing market shortages, a more efficient solution would address the market design elements that discourage full participation in the balancing market. While some aspects of the current market design could be improved incrementally, a comprehensive proposal that addresses a range of flaws is currently being considered under the Texas Nodal design.

Although it is not yet clear how the replacement reserves market will affect the frequency of shortages that occur in the balancing market, it will significantly impact the way in which OOMC commitments are made. The operators will have the ability to incorporate local capacity constraints into the replacement reserve market model which will resolve the local need by committing the most economic resource(s) available. This has the potential to make OOMC decisions more economically efficient, and thereby lower the uplift costs for these commitments.

There are two potential negative impacts from the replacement reserves market that should be monitored. First, the model determines whether additional capacity is necessary based on ancillary services requirements and a forecast of load on the following day. On days when the day-ahead load forecast is significantly higher than real-time load, it may result in unnecessary purchases of replacement reserves. Currently, QSEs are relied upon to start more capacity when real-time load is higher than anticipated. However, they are able to do this closer to real-time when load forecasts are likely to be more accurate, rather than in the day-ahead timeframe when the replacement reserves market will clear.

Second, the replacement reserves market compensates units that are started through this market, but not units that are already scheduled to be on-line. This may create an incentive for market participants to delay committing certain resources until after the replacement reserves market. This would allow them to receive a capacity payment in addition to payments for sales through

the balancing market. If these units are not selected in the replacement reserves market, the QSE can still commit them after the market clears. If QSEs wait until after the replacement reserves market to commit large amounts of capacity, it may result in unnecessary capacity purchases and higher costs to consumers.

Therefore, we recommend that the PUC assess the impact of the replacement reserves market on an on-going basis to ensure that it improves the overall efficiency of the wholesale market in ERCOT.

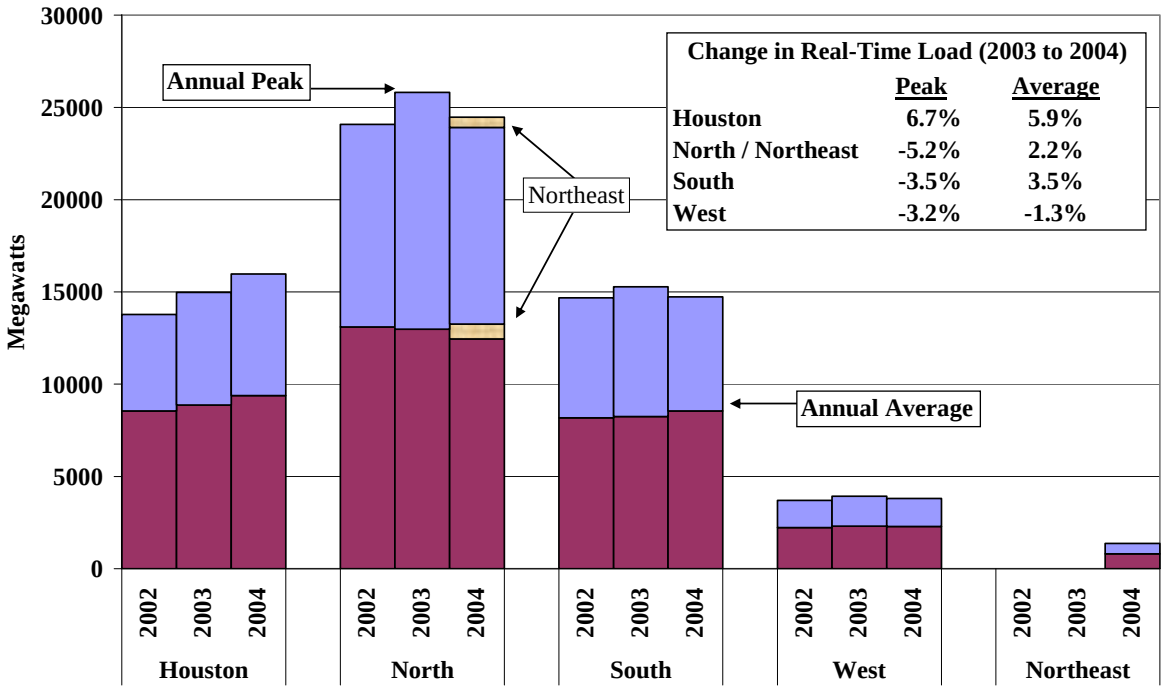
V. DEMAND AND RESOURCE ADEQUACY

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 2004 and the existing generating capacity available to satisfy the load and operating reserve requirements.

A. ERCOT Loads in 2004

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 are very important because they determine the probability and frequency of shortage conditions. More broadly, the peak demand levels and capability of the transmission network are the primary factors that determine whether the existing generating resources are adequate to maintain reliability. Hence, both of these dimensions of load during 2004 are examined in this subsection and summarized in Figure 56.

Figure 56: Annual Load Statistics by Zone
2002 to 2004



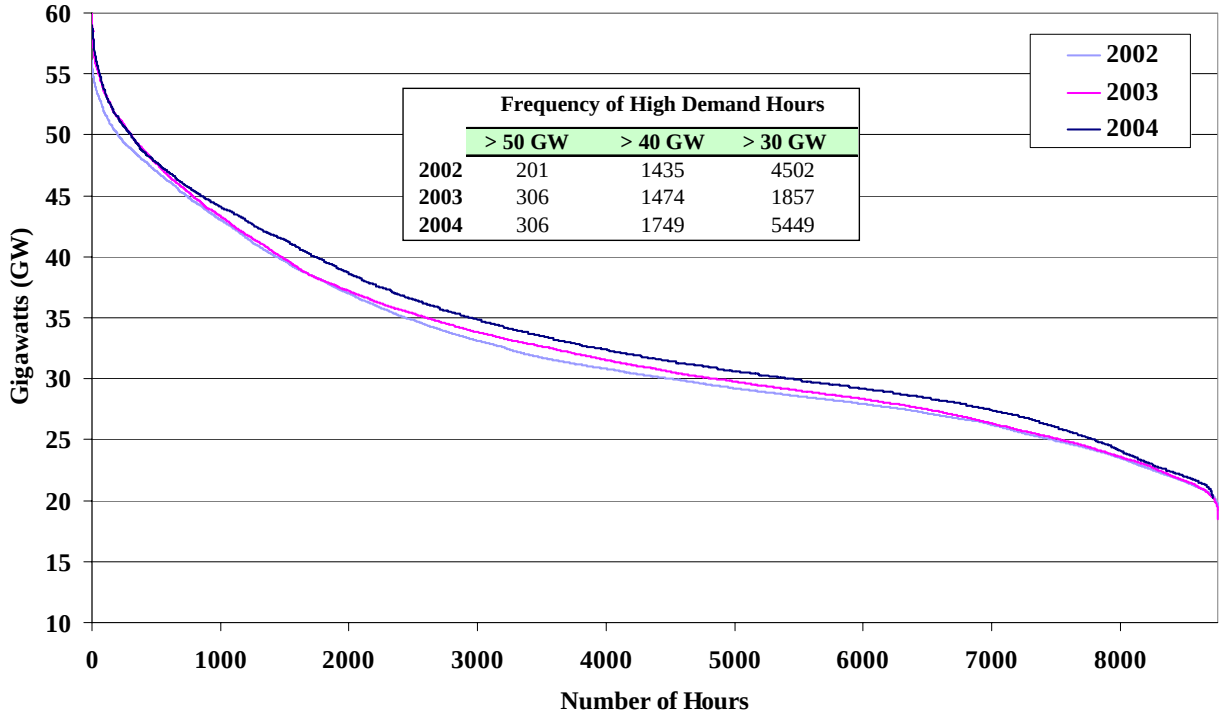
This figure shows peak and average loads in each of the four ERCOT zones from 2002 to 2004. Figure 56 indicates that in each zone, as in most electrical systems, peak demand significantly exceeds average demand. The North Zone is the largest zone (about 37 percent of the total ERCOT load); the South and Houston Zones are comparable (with about 27 percent each) while the West Zone and Northeast Zone are the smallest (with about 7 percent and 2 percent of the total ERCOT load).

No load statistics are shown for the Northeast Zone before 2004 because it was created from the North Zone at the beginning of 2004. For comparison purposes, the Northeast Zone is also shown stacked with the North Zone in 2004. ERCOT's peak load was 1,500 MW lower in 2004 than in 2003, while the ERCOT average load rose 3 percent. Houston showed the most significant load growth with 6.7 percent at the peak and 5.9 percent on average. Figure 56 shows that average loads in each zone were comparable between 2002 and 2003. This was due in part to the fact that the temperatures, with the exception of the hottest days, were relatively moderate in 2003.

The average load factor across the state in 2004 (defined as the ratio of average demand to peak demand) increased in 2004 from 54 percent to 57 percent. Similar improvements occurred at the zonal level, except for Houston where the load factor remained close to what it was in 2003. The highest load factors were in Houston (59 percent) and the West (60 percent). Houston has a higher load factor because the Gulf of Mexico moderates peak temperatures and the city's large manufacturing base provides a larger proportion of non-weather related demand.

To provide a more detailed analysis of load at the hourly level, Figure 57 compares load duration curves for 2002, 2003, and 2004. 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. In 2004, the highest load hours occurred in the summer months, particularly in August.

Figure 57: ERCOT Load Duration Curve*
All Hours – 2002 to 2004



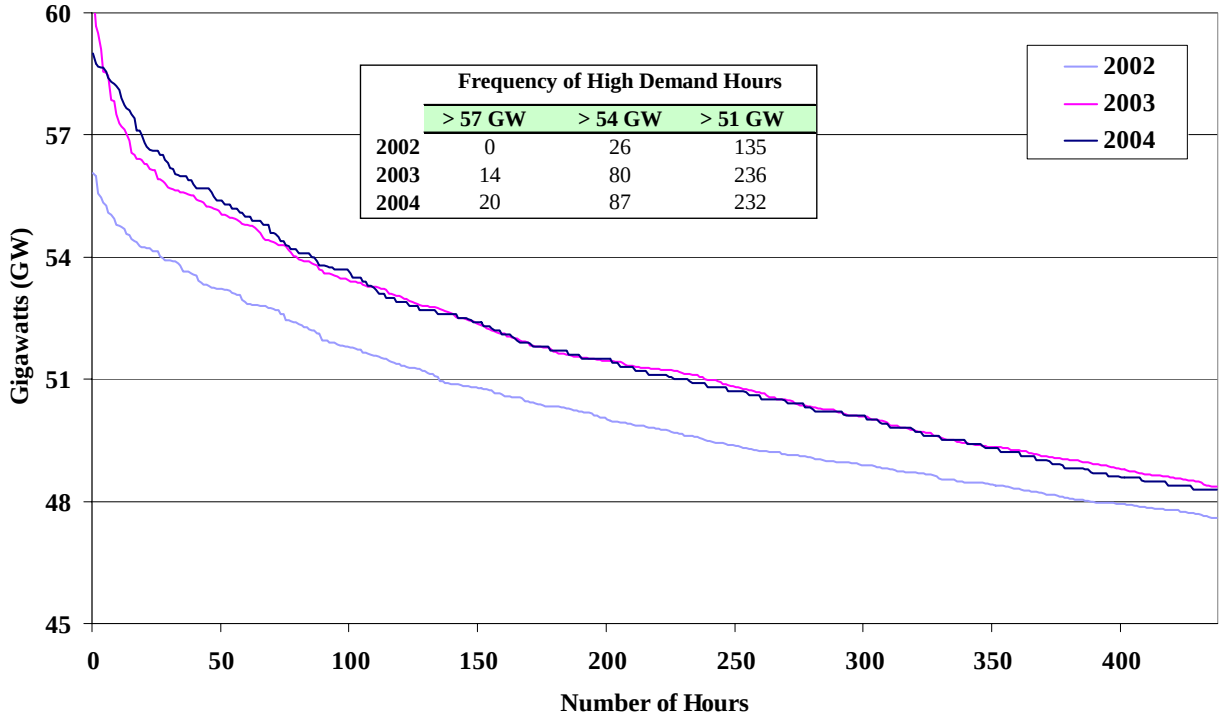
Frequency of High Demand Hours			
	> 50 GW	> 40 GW	> 30 GW
2002	201	1435	4502
2003	306	1474	1857
2004	306	1749	5449

* This is the load that the dispatch model uses to dispatch supply resources in the balancing market. This can differ slightly from actual metered load.

As Figure 57 shows, the load duration curve for 2004 lies above the ones for 2003 and 2002. Although the peak day demand in 2004 was lower than in 2003, overall demand was 3 percent higher in 2004 than in 2003. This indicates that demand at the mid-load levels was higher in 2004 than in 2003. This is not surprising since the loads in off-peak and mid-load periods are much less affected by random differences in weather patterns from year to year than are the peak load levels.

To better evaluate the differences in the highest-demand periods between the two years, Figure 58 shows the load duration curve for the top 5 percent of hours with the highest loads. This figure shows that differences in demand in the peak hours between 2002 and 2003 was significant but that the highest demand hours in 2003 and 2004 were comparable.

Figure 58: ERCOT Load Duration Curve*
Top Five Percent of Hours – 2002 to 2004



* This is the load that the dispatch model uses to dispatch supply resources in the balancing market. This can differ slightly from actual metered load.

Figure 58 shows that demand exceeded 57 GW in 20 hours in 2004 and 14 hours in 2003. In 2002, demand was not higher than 57 GW in any hour. The same pattern prevailed at lower load levels. Demand exceeded 54 GW in 87 hours in 2004 and 80 hours in 2003, compared to only 26 hours in 2002. Although peak demand conditions were more severe in 2004 and 2003 compared to 2002, this did not tend to cause sharp increases in electricity prices because the ERCOT market continues to enjoy substantial excess capacity.

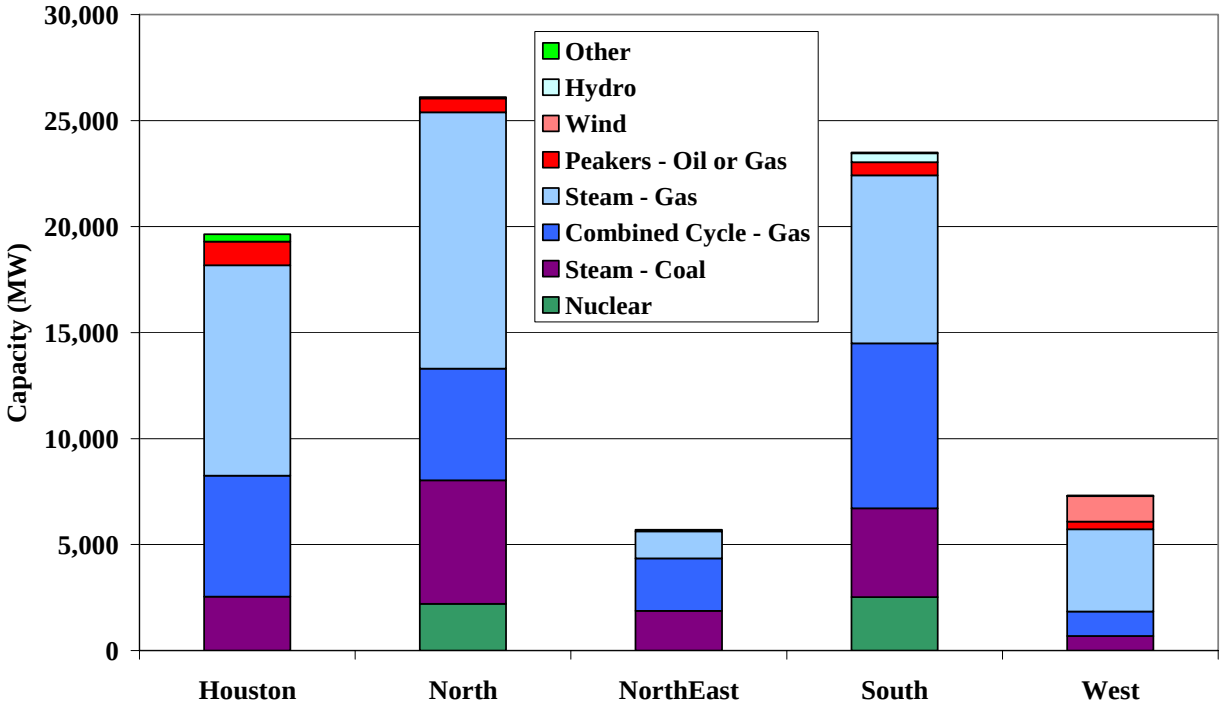
This figure also shows that the 58 GW peak load in 2004 was roughly 25 percent greater than the load at the 95th percentile of hourly load (approximately 48 GW). This is typical of the load patterns in an electricity market. Given that an additional 3 GW to 4 GW are needed to supply operating reserves and regulation, this implies that in long-run equilibrium with no surplus capacity, almost one-third of the generating resources are needed to supply energy in less than 5

percent of the hours while maintaining required regulation and operating reserves.[35] 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 59 shows the generating capacity by type in each of the ERCOT zones.

Figure 59: Installed Capacity by Technology for each Zone 2004



This figure shows that there is some nuclear capacity in both the North and South Zones, while lignite coal is also a major contributor in the North Zone. However, the primary fuel in all five zones is natural gas -- accounting for 73 percent of generation capacity in ERCOT as a whole,

[35] The range in the operating reserve and regulation requirements is based on the variable nature of the non-spinning reserves requirements.

and 85 percent in the Houston Zone. Much of this natural gas-fired capacity represents relatively new combined-cycle units than have been installed throughout ERCOT over the past few years.

ERCOT's reliance on natural gas resources makes it vulnerable to natural gas price spikes because coal and nuclear plants are primarily base load units. There is approximately 20,000 MW of coal and nuclear generation in ERCOT. Because there are very few hours where 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 more than half of the energy in ERCOT, they play a much less significant role in setting spot electricity prices due to their relatively low marginal production costs.

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 integrated in independent geographic areas. The North Zone accounts for 32 percent of capacity, the South Zone 29 percent, the Houston Zone 24 percent, the West Zone 9 percent, and the Northeast Zone 7 percent. The North Zone is an importer of power, while the Northeast Zone exports significant quantities because it has approximately three times more generation than its peak zonal load. Because large amounts of power flow from the Northeast to the North Zone, ERCOT created the Northeast Zone and associated Commercially Significant Constraint to manage these flows. The ratio of generating resources to load is slightly higher in the South and lower in Houston than the ERCOT average. This helps explain the patterns of exports from the South to Houston, as discussed below.

1. ERCOT Resource Margins

In this subsection, we estimate the resource margin in ERCOT based on the actual peak demand and installed capacity over the past two years. A resource margin indicates the amount of generating resources (including imports) that are available in excess of the peak load as a percentage of the peak load. Table 3 provides a detailed breakdown of generation capacity by technology type and resource margins in ERCOT.

This table shows that ERCOT had substantial excess capacity in 2003 and 2004. Excluding mothballed capacity and import capability, resource margins for ERCOT as a whole have remained above 20 percent the last two years. When import capability from external ties and

switchable resources are included, the resource margin is 31 percent in both 2003 and 2004. When including total potential response from loads acting as resources, the resource margin is 33 percent in both years.

Table 3: Generation Capacity and Resource Margins in ERCOT 2003 & 2004

Category	Formula	2003	2004
Installed Capability by Type (MW)			
Nuclear		4,737	4,737
Steam - Coal		15,133	15,133
Combined Cycle - Gas		17,111	19,398
Steam - Gas		35,943	35,072
Peakers - Oil or Gas		3,026	2,763
Wind(10% included here)		94	119
Hydro		552	552
Other		413	438
Total Capacity	(1)	77,009	78,212
Out-of-Service Capacity (MW)			
Mothballed Capacity	(2)	2,420	5,644
In-Service Capacity	(3) = (1) - (2)	74,589	72,568
Imported Capacity (MW)			
Switchable Capacity	(4)	3,068	2,988
DC Tie Import Capacity	(5)	856	856
In-Service Capacity Incl. Imports	(6) = (3) + (4) + (5)	78,513	76,412
LaaRs - Loads Acting as Resource	(7)	1,200	1,478
In-Service Capacity, Imports, LaaRs	(8) = (6) + (7)	79,713	77,890
Actual Peak Demand (MW)	(9)	59,996	58,528
Ratio of Resources to Actual Peak Demand:			
No Imports, Switchable, LaaRs	(10) = (3) / (9) - 1	**24%**	**24%**
Plus Switchable*	(11) = (4) / (9) + (10)	**29%**	**29%**
Plus DC-Tie Imports	(12) = (6) / (9) - 1	**31%**	**31%**
Plus LaaRs**	(13) = (8) / (9) - 1	**33%**	**33%**

* Most comparable to ERCOT methodology for calculating resource margin.

** This resource margin is over-estimated to the extent that the peak demand was reduced by the deployment of LaaRs (since the true peak would have been higher and LaaRs are already counted as resources).

Although these resource margins are sizable, it is important to consider that electricity demand in Texas has been growing at a rapid pace. From 1994 to 2004 the coincident peak grew at an annual rate of 3.0 percent.[36] At this rate, it will take little more than four years to reduce the ERCOT resource margin to 15 percent with no new generation. It is also important to consider that a significant number of generating units in Texas will soon be reaching or are already exceeding their expected operating lives. Over 8,300 MW of generation capacity is at least 40 years old, and another 18,600 MW of generation is between 30 and 40 years old.[37] Hence, it is important to ensure that the ERCOT markets are designed to send efficient economic signals so that investment occurs to maintain adequate resources as load grows and older resources retire.

2. Generation Outages and Deratings

The prior subsection shows substantial resource margins, indicating that the adequacy of resources is not a concern in ERCOT in the near-term. However, resource adequacy must be evaluated in light of the resources that are actually available on a daily basis to satisfy the energy and operating reserve requirements in ERCOT. A substantial portion of the installed capability is frequently unavailable due to generator deratings. A derating is the difference between a generating resource's installed capability 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 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).

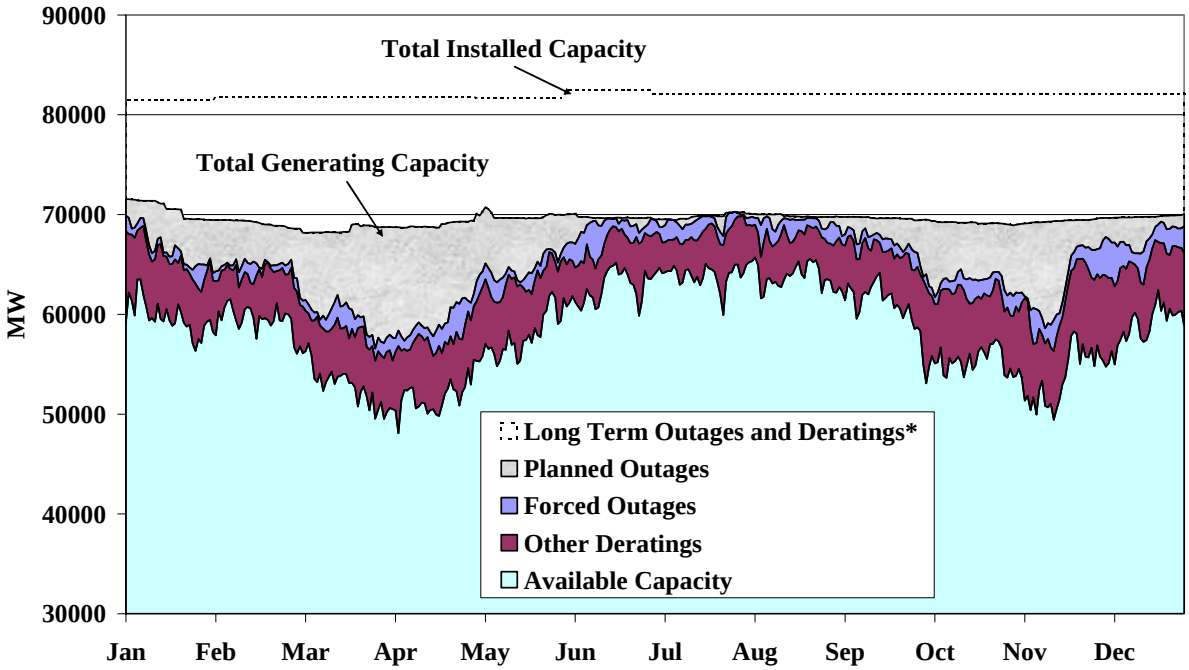
In this subsection, we evaluate long-term and short-term deratings to inform our evaluation of ERCOT capacity levels. Figure 60 below shows a breakdown of total installed capability for ERCOT on a daily basis during 2004. This analysis includes all in-service and switchable capacity. The capacity in this analysis is separated into five categories: (a) long-term outages and deratings, (b) short-term planned outages, (b) short-term forced outages, (c) other short-term deratings, and (d) available and in-service capability.

[36] ERCOT Transmission Study, 2003, p. 56.

[37] ERCOT Transmission Study, 2003, p. 69.

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. Some of this capacity may be out-of-service for extended periods due to maintenance requirements or may be out-of-service during the spring and fall months for economic reasons. 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 60: Short and Long-Term Deratings of Installed Capability
2004



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

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

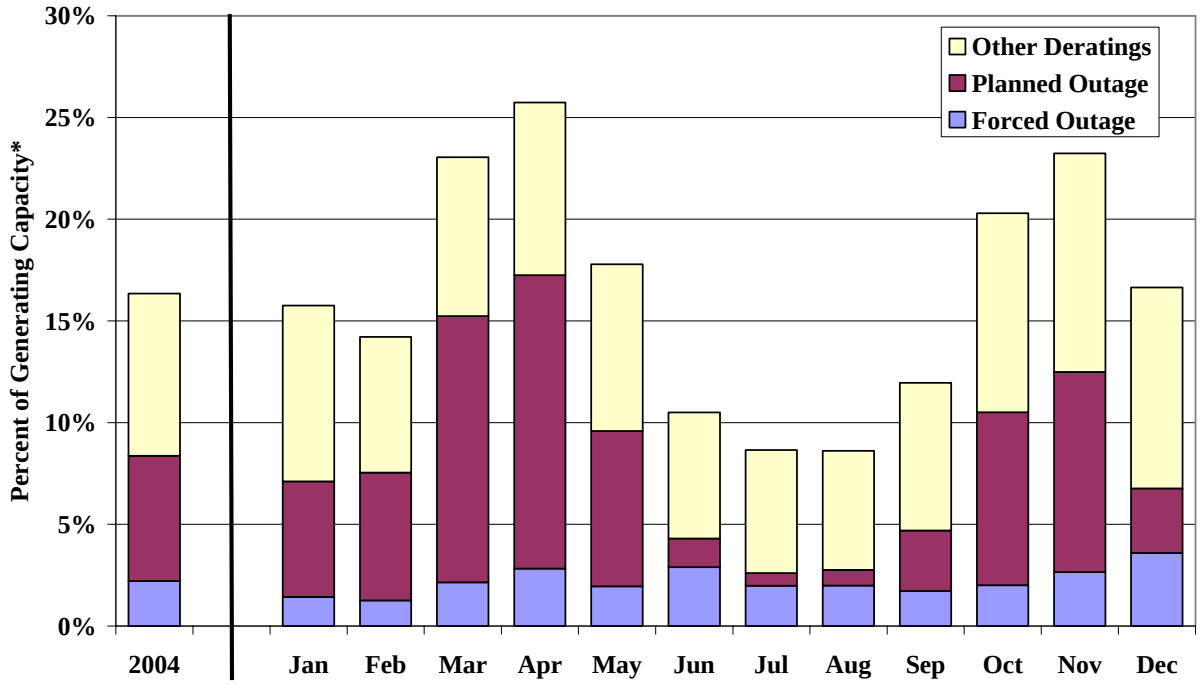
Figure 60 shows that installed capacity, including mothballed and switchable capacity, rose from 81 GW at the beginning of 2004 to 82 GW at the end of 2004. This increase is due to several new generators coming on-line although it was diminished by several retirements. The figure shows that the long-term outages and deratings fluctuated somewhat but generally grew from 10 GW at the beginning of 2004 to 12 GW at the end of the year. The long-term outages and

deratings also include over 5 GW of mothballed capacity. These classes of capacity can be made available if market conditions become tighter as load rises.

As expected, planned outages are relatively large in the spring and fall, decreasing to close to zero during the summer. Available in-service capacity fluctuated between 48 GW in March and April and 65 GW in August. The peak hour for the year required less than 59 GW to satisfy ERCOT's energy requirements and an additional 3 GW for operating reserves and regulation-up requirements, resulting in surplus capacity of less than 4 GW. This surplus is much smaller than the resource margin statistics would imply.

The next analysis focuses specifically on the short-term outages and deratings. To more clearly show the outages and deratings lasting less than 60 days, Figure 61 shows the average magnitude of the outages and deratings lasting less than 60 days for the year and for each month during 2004.

Figure 61: Monthly Average Outages and Deratings*
2004



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

Figure 61 shows that short-term deratings and outages were as large as 26 percent of installed capacity in the spring, dropping below 9 percent for most of the summer. Most of this

fluctuation was due to anticipated planned outages, which ranged from approximately 10 to 15 percent of installed capacity during March, April, October, and November. Short-term forced outages occurred more randomly, as expected, ranging between 1 percent and 3 percent of total capacity on a monthly average basis during 2004. 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 61 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. Lastly, 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 natural deratings due to ambient conditions and other factors described above. Because these natural 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 5 percent on average during the summer in 2004 and as high as 10 percent in other months.

3. 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 units minus the demand for energy, operating reserves, and up regulation. 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 62 plots the excess capacity in ERCOT during 2004. The figure shows the excess capacity in only the peak hour of each day because the commitments of generating resources are intended to cover the forecasted peak for the following day. Hence, one would expect larger quantities of excess capacity in other hours.

Figure 62: Excess On-Line and Quick Start Capacity During Daily Peaks on Weekdays -- 2004

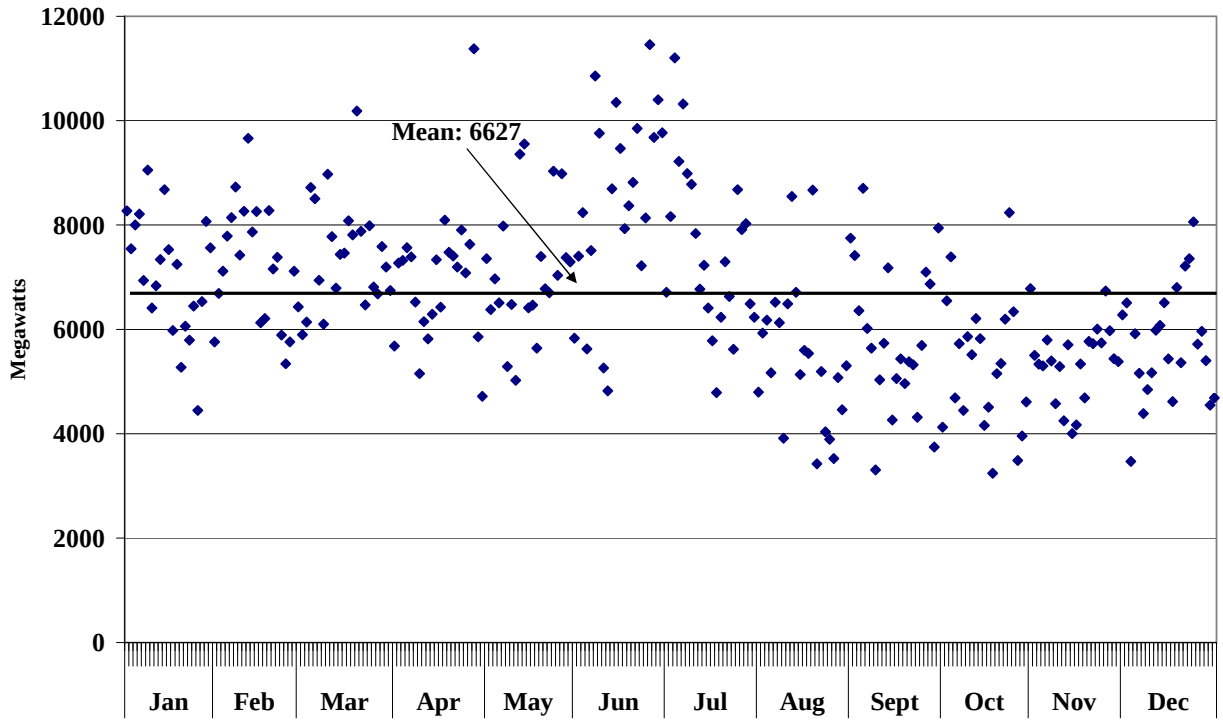


Figure 62 shows that the excess capacity in ERCOT was significant during 2004. The levels rarely fell below 4,000 MW on any day and sometimes exceeded 10 GW. During the peak load day in 2004 (on August 3), there were 5,930 MW available. The excess capacity averaged 6,627 MW, which is approximately 20 percent of the average load in ERCOT. As explained above, some of this excess capacity reflects the fact that it can be economic to commit steam units or combined cycle units to serve the peak load even when quick-start peaking resources are available. However, the off-line quick-start resources reflect less than half of the excess shown in the figure. The fact that the quantity of capacity committed exceeds the energy and ancillary services requirements by such a wide margin indicates that the current ERCOT market design

tends to result in an over-commitment of resources. While this assists in ensuring reliability, this level of committed capacity is not efficient because these sizable excess resource commitments result in higher than necessary production costs.

The figure shows that the average level of excess capacity fluctuated significantly over 2004, but clearly shifted downward from August through the end of the year. This may reflect an attempt by market participants to efficiently utilize the resources in their portfolio. There was also substantial reduction in OOMCs for local capacity needs after August which has likely contributed to the reduction in excess capacity.

The tendency to over-commit capacity can be attributed in large part to the lack of a centralized day-ahead commitment process in ERCOT. Without a centralized commitment mechanism, each participant makes independent generator commitment decisions that, taken together, are not likely to be optimal. Hence, the introduction of day-ahead energy and operating reserves markets 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 conditions. The ERCOT market allows participants with demand-response capability to provide the energy, reserves, and regulation 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 LaaRs that are qualified to offer responsive reserves and non-spinning reserves into the day-ahead ancillary services markets. Those that are qualified 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. LaaRs that are capable of controllably reducing or

increasing consumption under dispatch control (similar to AGC) are not currently able to provide regulation service.

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. Unlike some LaaRs, however, they are not qualified to provide reserves or regulation service.

During 2004, 67 resources totaling 1657 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 or the non-spinning reserves market. There were no BULs registered with ERCOT in 2004. Figure 63 shows the amount of responsive reserves provided from LaaRs on a daily basis in 2004.

Figure 63: Provision of Responsive Reserves by LaaRs
Daily Average – 2004

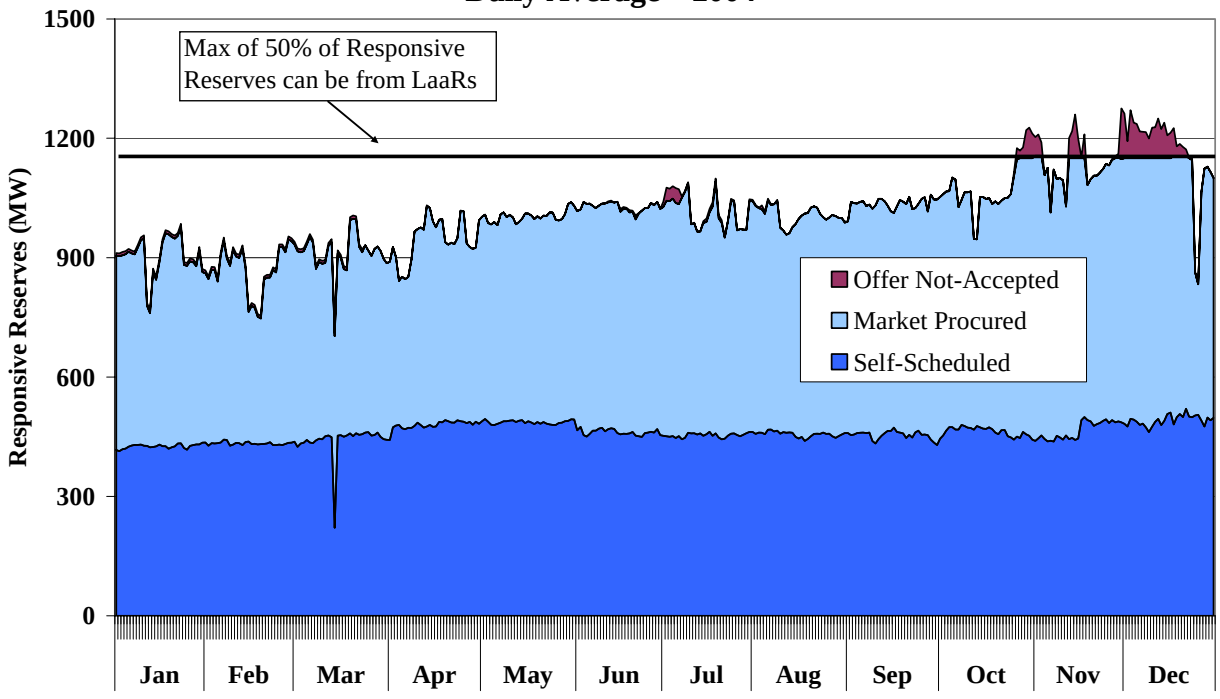


Figure 63 shows that the amount of responsive reserves provided by LaaRs gradually increased from about 900 MW at the beginning of 2004 to as much as 1,150 MW at the end of 2004. The majority of this increase was procured through the ERCOT administered auction rather than self-provision and bilateral agreements. Currently, LaaRs are permitted to supply up to 1,150 MW of

the responsive reserves requirement. Virtually all of the responsive reserves offered by LaaRs are procured, indicating that the LaaRs are offered at prices that are relatively low. In fact, as the figure shows, the unaccepted offers were generally not accepted because of the 1,150 MW limit. The total quantity of responsive reserves supplied by LaaRs represented 44 percent of the total 2,300 MW requirement for responsive reserves in 2004, and 49 percent of this requirement during November and December. The high level of participation by demand response sets ERCOT apart from other operating electricity markets.

Although LaaRs are active participants in the responsive reserves market, they did not provide offers in the balancing energy, non-spinning reserves, or regulation services markets in 2004. 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, providing non-spinning reserves introduces a much higher probability of being curtailed. Participation in the regulation services market requires technical abilities that LaaRs cannot meet at this point. Finally, 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 regulation services, non-spinning reserves, or balancing energy.

VI. 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 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

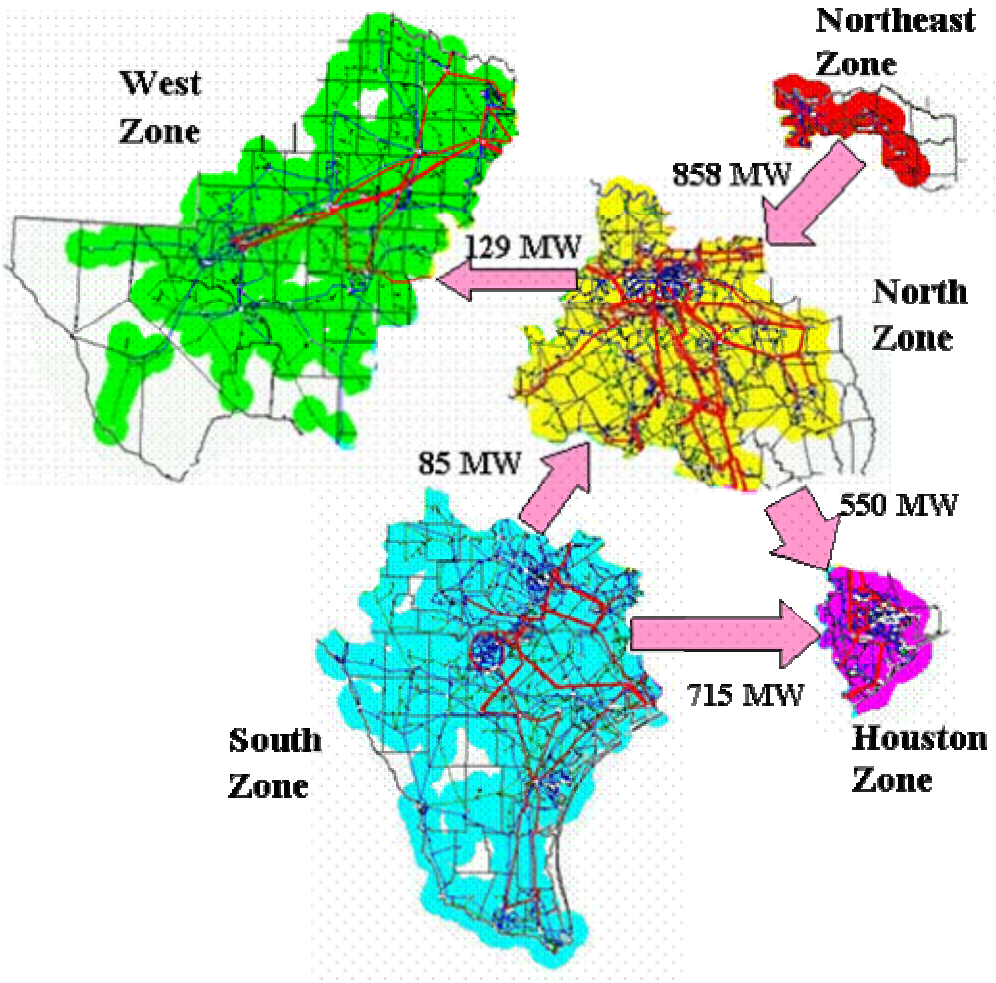
In 2004, there were five commercial pricing zones in ERCOT: (a) the North Zone, (b) the West Zone, (c) the South Zone, (d) the Houston Zone, and (e) the Northeast Zone, which was created in 2004 by carving up the North Zone. The balancing energy market uses the SPD software that dispatches balancing 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 five transmission interfaces. These five transmission interfaces, referred to as Commercially Significant Constraints ("CSCs"), are simplified representations of groups of transmission elements. ERCOT operators use planning studies and real-time information to set limits for each CSC that are intended to utilize the total transfer capability of the CSC. In this subsection of the report, we describe the SPD model's simplified representations of flows between zones and analyze actual flows in 2004.

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. In particular, it discusses the impact on congestion management of adding one new zone and two new CSCs. Figure 64 shows the average SPD-modeled flows over CSCs between zones during 2004.

Figure 64: Average SPD-Modeled Flows on Commercially Significant Constraints 2004



Note: In the figure above, CSC flows are averaged taking the direction into account. For instance, if one hour has a North to West flow of 100 MW, and a second hour has a West to North flow of 200 MW, the average would be 50 MW from the West to North. This treats the North to West flows in the first hour as negative for averaging purposes.

Figure 64 shows the five ERCOT geographic zones as well as the five CSCs that interconnect the zones: (a) the West to North interface, (b) the South to North interface, (c) the South to Houston

interface, (d) the Northeast to North interface, and (e) the North to Houston interface. The Northeast to North and North to Houston CSCs were defined before 2004 to address large amounts of uplift generated by managing these pathways using local congestion management procedures. Based on SPD modeled flows, Houston is a significant importer while the Northeast Zone and the South Zone export significant amounts of power. It is interesting to note that SPD calculated net flows from the North Zone to the West Zone on average, while the West to North CSC was defined to only limit flows in the opposite direction. Not surprisingly, a new North to West CSC was defined for 2005 because ERCOT has found that congestion occurs in both directions.

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")[38] in relation to the CSC. In reality, the generators within each zone can have widely varying effects on the flows over a CSC. In order to illustrate this, we calculated flows that would occur over the CSC using actual generation and actual generation shift factors and compared this to transmission flows calculated using actual generation and zonal average shift factors. Table 4 shows this analysis.

Table 4: Average Calculated Flows on Commercially Significant Constraints Zonal-Average vs. Unit-Specific GSFs – 2004

CSC 2004	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	-129	-158	*-29*	-217	*-59*
South-North	85	99	*14*	49	*-50*
South-Houston	715	699	*-16*	1072	*373*
North-Houston	550	512	*-38*	358	*-154*
NorthEast-North	858	849	*-9*	758	*-91*

[38] 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 4 shows the average flows over each CSC calculated by SPD. The second column shows the average flows over each CSC we 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 slightly lower calculated flows on each CSC except the South to North interface, where calculated flows are slightly higher. The fourth column in Table 4 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 measures the inaccuracy caused by treating each unit within a particular zone as having identical impacts 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 4 shows that the unit-specific GSFs increased the calculated flows on the South-Houston interface by 373 MW and reduced the calculated flows on the other four CSCs by 50 MW to 154 MW each. These differences are sizable and are significantly 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. By using generation-weighted shift factors for load rather than load-weighted shift factors, it can cause large differences between SPD flows and actual flows. For instance, SPD flows for the Northeast to North interface will be approximately 400 MW higher than actual flows due this simplification.[39] However, this does not raise

[39] The annual planning study used by ERCOT to forecast transmission capability prior to the 2004 annual Transmission Congestion Rights auction calculates this effect to be 418 MW during summer peak conditions.

concerns that are as significant as for generators since 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 re-dispatching to manage congestion, while this is not a concern for loads in the balancing market since they are not re-dispatched by the model anyway. 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 (e.g. the Northeast Zone), and others pay too little (e.g. the North Zone). [40]

In order to effectively manage interzonal congestion, it is important for SPD to accurately model the major constrained transmission interfaces between zones. In 2004, the five CSCs modeled by SPD did not include all significant interfaces between zones, although five was a substantial improvement over 2003 when only three CSCs were modeled by SPD. Even with the new CSCs, sizeable quantities of power were transported on transmission facilities not modeled by SPD. Table 5 summarizes the actual net imports into each zone compared to SPD modeled flows in 2003 and 2004.

Table 5: Actual Net Imports vs. SPD-Calculated Flows on CSCs 2003 & 2004

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

[40] For instance, the generation-weighted shift factor of the Northeast Zone with respect to the Northeast to North CSC is generally about 25 percent, whereas the load-weighted shift factor is generally about 35 percent. On April 29, 2004, interval-ending 16:00, the Northeast Zone price was $60 and the North Zone price was $39 due to a shadow price of $50 on the Northeast to North CSC. If the load-weighted shift factor was used to calculate the price to load, the price would be $34.

Table 5 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 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 causes systematic differences between SPD flows and actual flows. For instance, SPD generally underestimated flows on the Northeast-North CSC, accounting for a significant chunk of the difference between SPD flows and net exports from the Northeast Zone. However, these reasons do not explain all of the difference between actual net interchange and interchange modeled on CSCs.

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 more than ten CREs ("Closely Related Elements") which can 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 5 shows significant changes in the levels of net imports into each zone between 2003 and 2004. The West Zone shifted from being a net exporter in 2003 to importing substantial quantities in 2004. From 2003 to 2004, net exports increased from the South Zone as well as the combined area of the North and Northeast zones. In every case, the flows on CSCs were significantly less than the actual interchange in both years. In 2003, the Houston Zone showed the largest difference, importing an average of 1,796 MW while SPD modeled an average CSC import of 565 MW.

Part of this difference occurred because the Houston Zone imported large quantities from the North Zone on four 345 kV transmission lines that were not managed by zonal balancing deployments in 2003. When these additional flows do not cause transmission constraints to bind,

they raise no significant market issues. However, if transmission constraints between zones that are not defined as part of a CSC do become binding, ERCOT's means for managing the constraints can result in inefficiencies. To address this, ERCOT introduced the North to Houston CSC in 2004 to allow it to better manage interzonal congestion.[41] Table 5 indicates that SPD flows on CSCs into Houston more than doubled because of the addition of the North to Houston CSC. In 2004, actual net interchange also increased by 683 MW, largely because Houston experienced the most substantial load growth of any zone from 2003 to 2004.

B. Interzonal Congestion

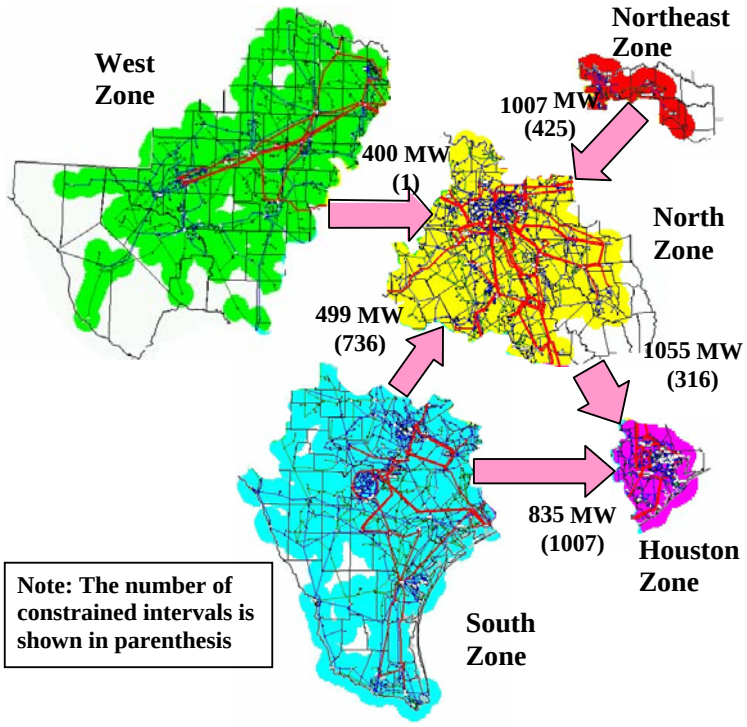
The prior subsection showed the average interzonal flows calculated by SPD compared to actual flows in all hours. This subsection focuses on those intervals when the interzonal constraints were binding. Although this is a small subset of intervals, it is in these constrained intervals that the performance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

Figure 65 shows the average SPD-calculated flows between the five ERCOT zones during constrained periods for the five CSCs. The arrows show the average magnitude and direction of the SPD-calculated flows during constrained intervals. The frequency with which these constraints arise is shown in parentheses.

Figure 65 shows that the SPD-calculated flows averaged 835 MW on the South to Houston interface during 1,007 constrained intervals in 2004. The average SPD-calculated flows on the South to North interface was 499 MW during the 736 intervals the interface was constrained. The South to Houston and South to North interfaces exhibited higher SPD-calculated flows in these constrained intervals than the average flows in the other intervals. Similarly, the two new CSCs exhibited significantly higher flows during constrained intervals than in unconstrained intervals. Although both of the new CSCs were constrained less frequently than the South to North and South to Houston interfaces.

[41] On an interim basis, the North to Houston CSC was managed using zonal OOME deployments from June to December 2003.

Figure 65: Average Modeled Flows in Transmission Constrained Intervals[42] 2004



Note: In the figure above, CSC flows are averaged taking the direction into account. For instance, if one hour has a North to West flow of 100 MW, and a second hour has a West to North flow of 200 MW, the average would be 50 MW from the West to North. This treats the North to West flows in the first hour as negative for averaging purposes.

1. Congestion Rights

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. In order 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 difference in these prices.

Market participants in ERCOT can hedge congestion charges in the balancing energy market by acquiring Transmission Congestion Rights ("TCRs") or Pre-assigned Congestion Rights

42 There were four additional intervals where SPD re-dispatched the system for congestion on the West to North interface. These are not shown in Figure 56 because they were the result of limits that were erroneously entered.

("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 and/or PCR payments that fully 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.

In order to analyze the congestion rights in ERCOT, we first review the TCRs and PCRs that were allocated for each CSC in 2004. Figure 66 shows the average number of TCRs and PCRs that were allocated for each of the CSCs in 2004, as well as the average SPD-modeled flows during the constrained intervals.

Figure 66: Transmission Rights vs. Real-Time SPD-Calculated Flows Constrained Intervals – 2004

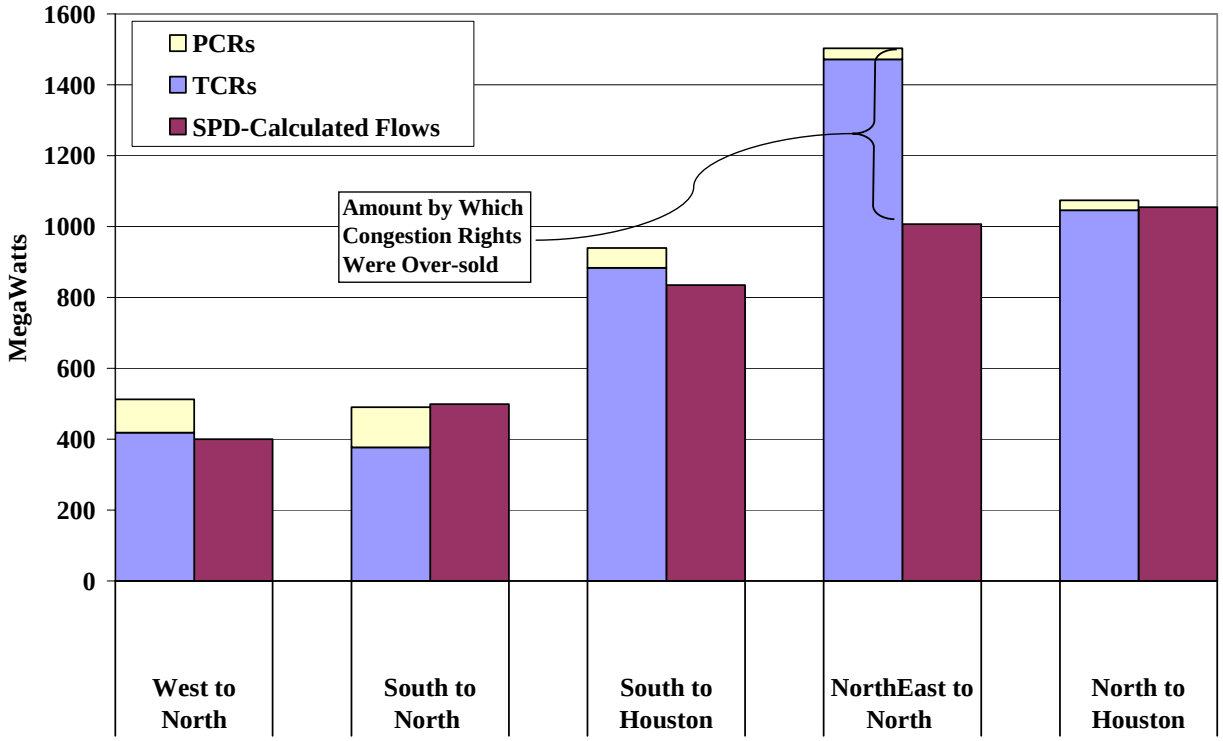


Figure 66 shows that total congestion rights (the sum of PCRs and TCRs) on the West to North, South to Houston, and Northeast to North interfaces exceeded the average real-time SPD-calculated flows during constrained intervals. These results indicate that the congestion rights

were oversold in relation to the SPD-calculated limits. For instance, congestion rights for the Northeast to North interface were oversold by an average of 496 MW.

The largest divergence between the SPD-calculated limits and the limits implied by the congestion rights was on the Northeast to North interface where 1,503 MW of congestion rights were allocated, but the average SPD-calculated flow during constrained intervals was 1,007 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 496 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 66, we compare the SPD-calculated flows and congestion rights quantities for each of the constrained intervals by CSC.

2. South to North Interface

The first CSC we analyze at the interval level is the South to North CSC. Figure 67 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 2004. Because only congested intervals are shown, some months will have significantly more observations than other months. Indeed, the figure shows that congestion occurred with moderate frequency in May and September, while January, February, and December accounted for 73 percent of all constrained intervals during 2004.

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 67 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 67: Congestion Rights Allocated vs. SPD Flows during Constrained Intervals South to North – 2004

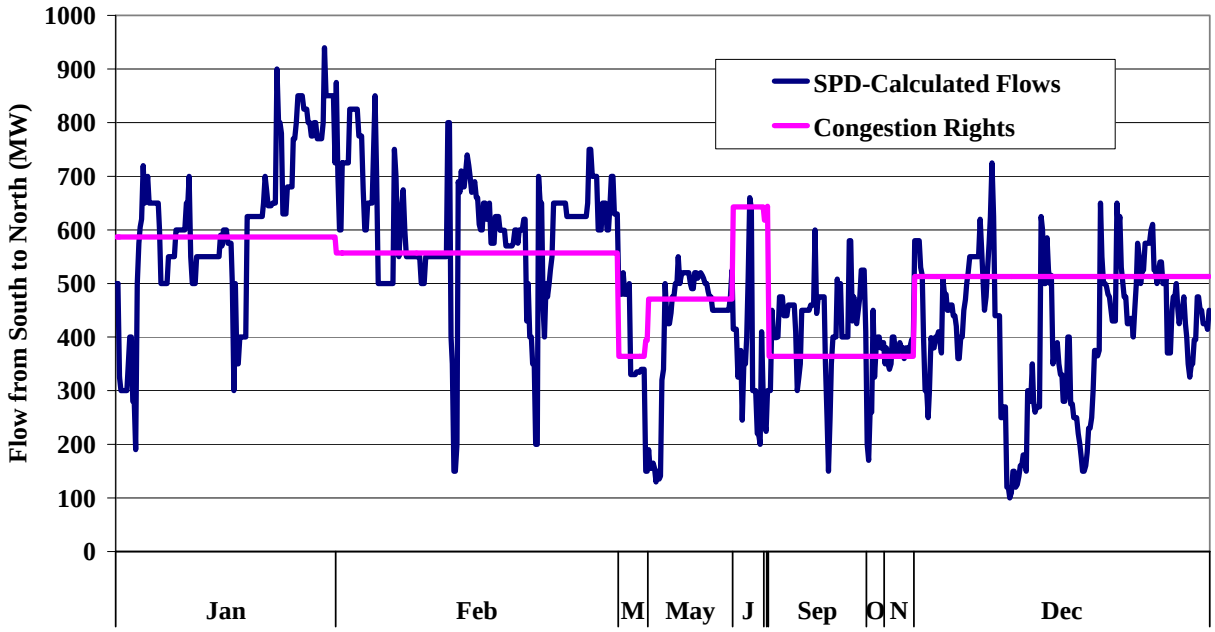


Figure 67 exhibits periods where SPD-calculated flows are both above and below the quantity of congestion rights. Congestion rights exceeded SPD flows during the majority of December, but they were smaller than SPD flows during most of February. The early part of January showed

flows generally below the quantity of rights, but this was reversed at the end of January. Figure 66 indicates that these generally averaged out so that SPD flows exceeded congestion rights by an average of 9 MW in 2004. The figure does not show any instances where the SPD-calculated flows were negative during constrained intervals, although it fell as low as 100 MW in December.

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.

3. South to Houston Interface

Figure 68 shows the total quantity of congestion rights allocated by ERCOT for the South to Houston interface relative to the SPD-calculated flows over the interface in congested intervals during 2004.

Figure 68: Congestion Rights Allocated vs. SPD Flows during Constrained Intervals South to Houston – 2004

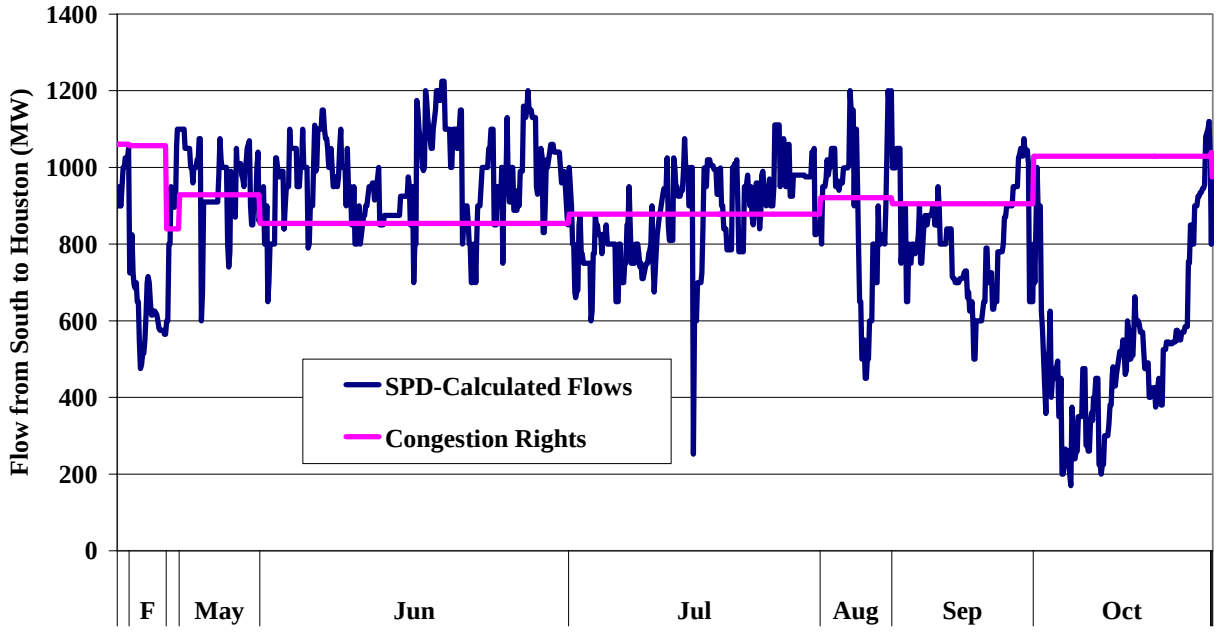


Figure 68 shows that the quantity of congestion rights for the South to Houston interface exceeded 1,000 MW at the beginning and end of 2004, but was reduced to approximately 900 MW from March to September. This is different from the pattern of SPD-calculated flows during constrained intervals, which were lowest during February and October, but usually higher than the quantity of congestion rights from March to July. During August and September, the SPD-calculated flows were significantly lower than the quantity of congestion rights.

Finally, while the West to North and South to North interfaces frequently bind when SPD flows are very low or negative, the South to Houston interface exhibits fewer periods where the SPD flows reached extremely low levels.

4. North to Houston Interface

Figure 69 shows the total quantity of congestion rights allocated by ERCOT for the North to Houston interface relative to the SPD-calculated flows over the interface in congested intervals during 2004.

Figure 69: Congestion Rights Allocated vs. SPD Flows During Constrained Intervals North to Houston – 2004

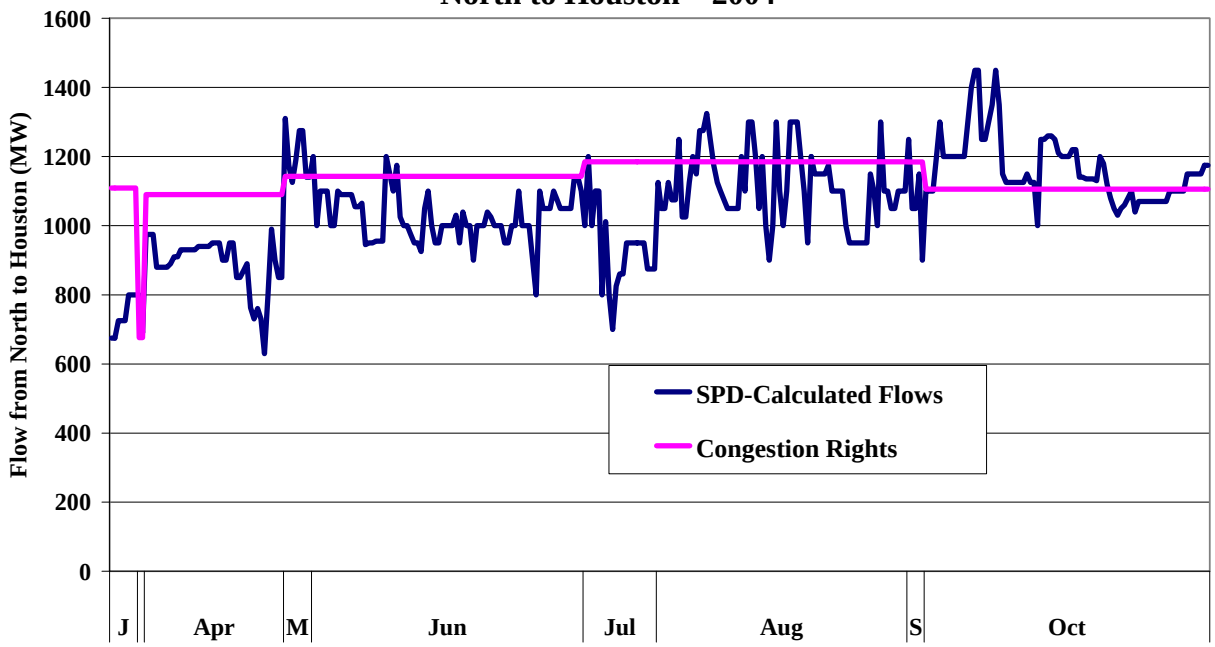


Figure 69 indicates that consistency between the quantity of congestion rights allocated and SPD-calculated flows was better for the North to Houston interface than for other interfaces. This is consistent with the overall conclusion of Figure 66, which showed that the average amount of congestion rights exceeded average flows during 2004 by just 19 MW. While the annual difference shown in Figure 66 was also small for the South to North interface, the North to Houston interface exhibits much better consistency during individual intervals.

5. Northeast to North Interface

Figure 70 shows the total quantity of congestion rights allocated by ERCOT for the Northeast to North interface relative to the SPD-calculated flows over the interface in congested intervals during 2004.

Figure 70: Congestion Rights Allocated vs. SPD Flows During Constrained Intervals Northeast to North – 2004

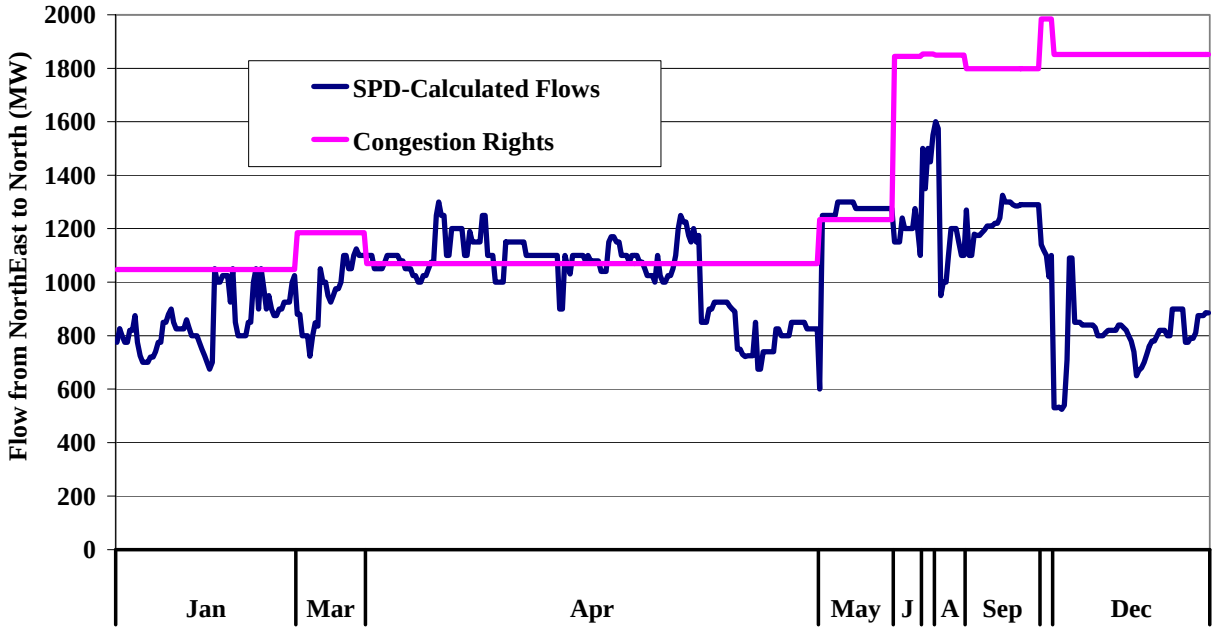


Figure 70 shows that the quantity of congestion rights for the Northeast to North interface ranged below 1,300 MW from January through May, and then increased above 1,800 MW for the remainder of the year. Constraints were significantly more common during the early period, which is consistent with there being less transmission capability. However, after the upgrade in transmission capability, SPD-calculated flows were lower than the quantity of congestion rights by an average of 828 MW during constrained intervals.

The increase in transmission capability resulted from a Special Protection Scheme ("SPS") that ERCOT implemented June 1, 2004. Ordinarily, interface limits are set so that in the event of a sudden contingency, the grid would still be reliable. The SPS sets in place procedures and/or equipment that allow the interface to carry more flow under normal conditions by limiting the impact that a large contingency would have on reliability. When these can be implemented reliably, they greatly enhance the capability of the transmission system to carry power from low-cost areas to higher-cost areas.

Transmission outages can have a significant effect on these results by reducing the flows that will be allowed by SPD. When the outage is recognized prior to when the monthly congestion rights are sold, ERCOT will reduce the quantity of rights that are made available to participants,

which would prevent the outage from causing a significant shortfall associated with a large divergence between the congestion rights and the SPD flows. However, short-term outages that are not recognized in the monthly auctions can contribute to such divergences and result in revenue shortfalls. The next section describes ERCOT's process for selling congestion rights and reviews the results of these sales for 2004.

In conclusion, the SPD-calculated flows can vary substantially and frequently they are not close to the actual flows or limits for the CSC. Because transmission rights are generally sold based on the actual CSC transfer capability, this can result in substantial surplus congestion revenue or congestion revenue shortfall that results in uplift charges. Under the current market design, it is extremely difficult to develop procedures for selling transmission rights that fully subscribe (without overselling) the available transmission capability.

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 60 percent of the congestion rights sold in the annual auction. These rights are auctioned in December for the coming year. The remaining 40 percent of the rights are designated based on monthly updates of the summer study.[43] Since the monthly studies tend to more accurately reflect conditions that will prevail in the coming month, the monthly designations tend to more closely reflect actual transmission limits.

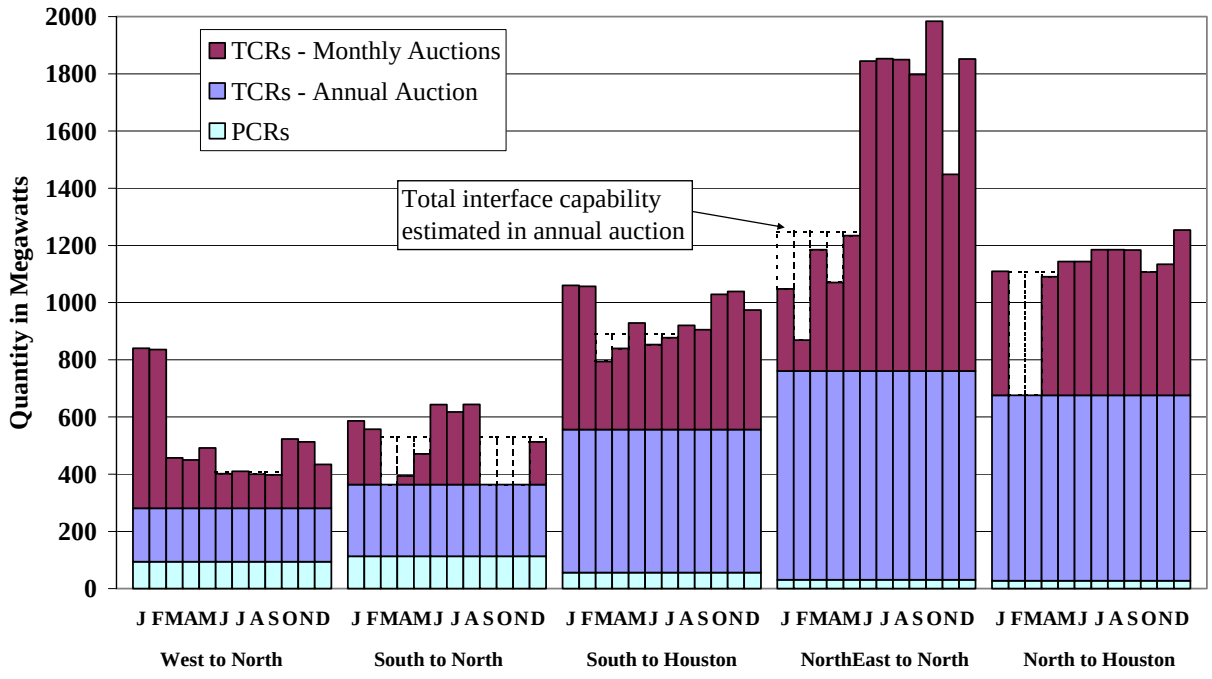
However, the summer and monthly studies used to designate the TCRs do not reflect transmission conditions that can arise in real-time. This happens for two main reasons. First, transmission and generator outages can occur unexpectedly, and significantly reduce the transfer capability of a CSC. Second, 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

[43] Starting in 2005, only 40 percent of estimated capability is sold in the annual auction, while the remaining 60 percent is sold in the monthly auctions.

transmission capability in SPD. This is one potential source of divergence for the West-North interface shown above.

To examine how these processes have together determined the total quantity of rights sold over each interface, Figure 71 shows the quantity of each category of congestion rights for each month during 2004. The quantities of PCRs and annual TCRs are constant across months and were determined before the beginning of 2004, while monthly TCR quantities can be adjusted monthly.

Figure 71: Quantity of Congestion Rights Sold by Type
2004



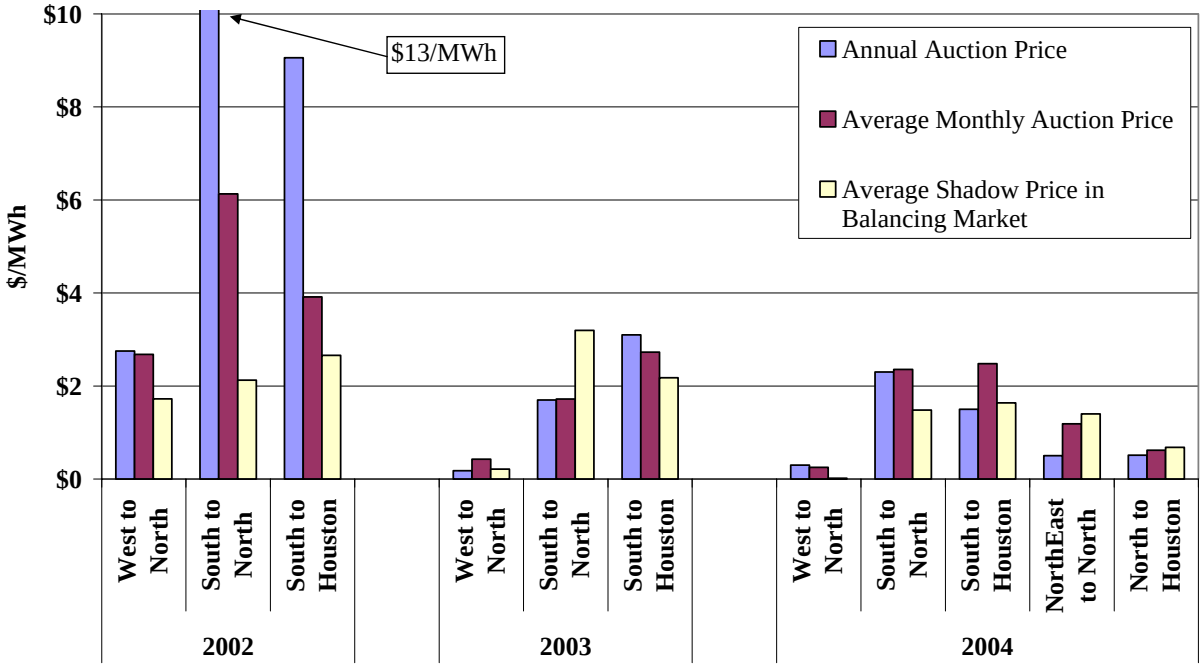
When the monthly planning studies indicate changes from the summer study, revisions are often made to the estimated transmission capability. Therefore, the auctioned congestion rights may increase or decrease relative to the amount estimated in the summer study. The shadow boxes in the figure represent the capability estimated in the summer study that is not ultimately sold in the monthly auction. When there is no shadow box in Figure 71, 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 Northeast to North 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 March or from September to November in the monthly auctions. The large increase in capability for the Northeast to North interface can be attributed to the implementation of a Special Protection Scheme ("SPS") to control the impact of a large contingency. This increase in TCR sales underscores the economic benefit of the SPS that was implemented by ERCOT. The divergence between annual and monthly estimates of transmission capacity on the other interfaces was smaller.

Market participants who are active in congestion rights auctions are subject to substantial uncertainty. Outages and other contingencies occur randomly 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 72 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 72 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 72: TCR Auction Prices versus Balancing Market Congestion Prices 2002 to 2004



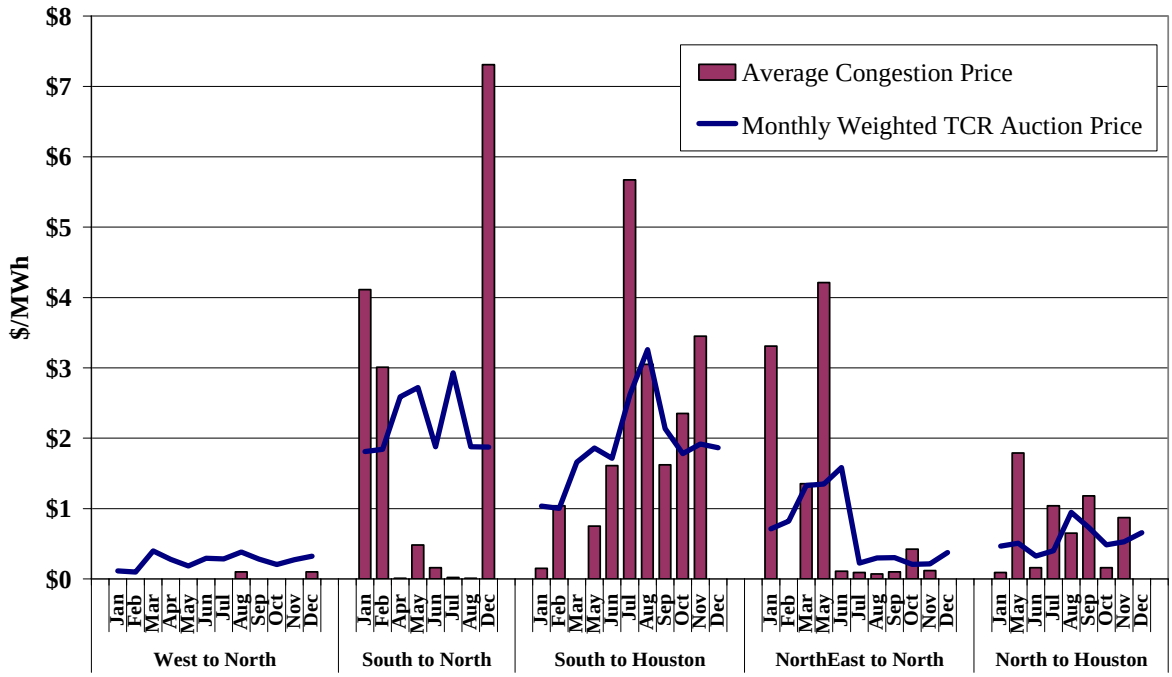
In 2003, the TCR prices for all of the interfaces decreased considerably, causing the prices to converge more closely with the actual value of the congestion rights. It is noteworthy that the TCRs for the South to North and South to Houston interfaces settled at prices in 2004 that were closer to the previous year's value than in 2003. This indicates that participants have improved in their ability to forecast interzonal congestion and to value the TCRs, in part by observing historical outcomes. This improvement is 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 CSCs. In both cases, the annual TCR price was below the monthly average TCR price, which was slightly below the average value of congestion. This reflects cautiousness on the part of market participants when purchasing a TCR for a CSC that did not exist before 2004.

Figure 73 compares monthly TCR auction prices with monthly average real-time CSC shadow prices from SPD for 2004. To compare these more easily, 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 interface and the North-Houston interface.[44]

Figure 73: Monthly TCR Auction Price and Average Congestion Value 2004



Although congestion in the balancing market can be sporadic and inherently difficult to predict, the monthly TCR prices for the South to Houston, Northeast to North, and North to Houston interfaces exhibited patterns that were correlated to balancing market congestion prices. For example, it seems that the monthly TCR market anticipated the decrease in congestion on the Northeast to North interface following its upgrade in capability. The TCR prices for the two interfaces into Houston indicate that market participants correctly believed congestion on those interfaces would rise during summer.

Based on Figure 73, market participants did a poor job predicting congestion on the South to North interface during 2004. Balancing market congestion was highest during January, February, and December, far exceeding the TCR prices in those months. However, from April through August, there was virtually no congestion on the South to North CSC. This drop in

[44] Notice that these missing months correspond to the missing monthly auction values in Figure 71.

congestion was unanticipated by the TCR market where participants paid higher prices for TCRs than in the three months where congestion was significant in the balancing energy market.

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 $3,300 (600 MWh * $55/MWh) while suppliers in the West Zone will receive $2,400 (600 MWh * $40/MWh). The net result is that ERCOT collects $900 in congestion rent ($3,300 – $2,400) and uses it to fund payments to holders of TCRs.[45] 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 74 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.

[45] 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 74: TCR Auction Revenues, Credit Payments, and Congestion Rent[46] 2002 to 2004

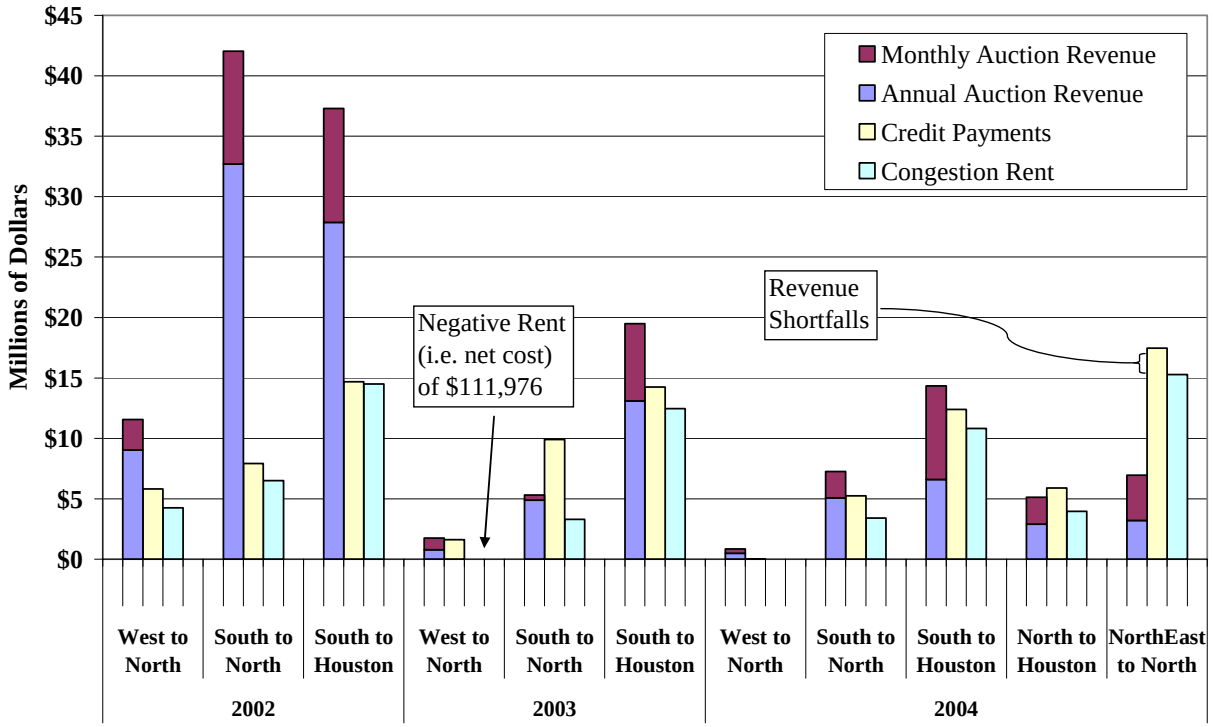


Figure 74 shows that in 2002, the total auction revenues were far greater than credit payments to TCR holders. This is the result of the auction prices being much greater than the average shadow prices that occurred in the balancing energy market (as was shown in Figure 73 above). The figure also shows 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

46 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.

Figure 74 also shows that the congestion rents exhibited better convergence with payments to congestion rights holders in 2004 than in 2003. In 2004, congestion rents were only moderately lower on each interface than the credit payments. The better convergence between congestion rents and credit payments can be attributed to better convergence between the amount of TCRs sold and the SPD-calculated flow during the constrained intervals.

In 2003, the convergence of congestion rents and credit payments were much worse on certain interfaces:

- Congestion rents from the West to North interface were *negative* $111,976.
- Rents from the South to North interface were less than half of the total credits payments.

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.[47] 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 costs 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. Hence, these constraints are not managed by the zonal market model. ERCOT manages local congestion by requesting that generating units adjust their output

[47] 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.

quantities (either up or down). When not enough capacity is committed to meet reliability, then 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. Also as explained above, a unit that ERCOT commits to meet its reliability requirements is an out-of-merit commitment or OOMC. The payments made by ERCOT when it takes OOME, OOMC, or RMR actions are recovered through uplift charges to the loads. The payments for each class of action are described below.

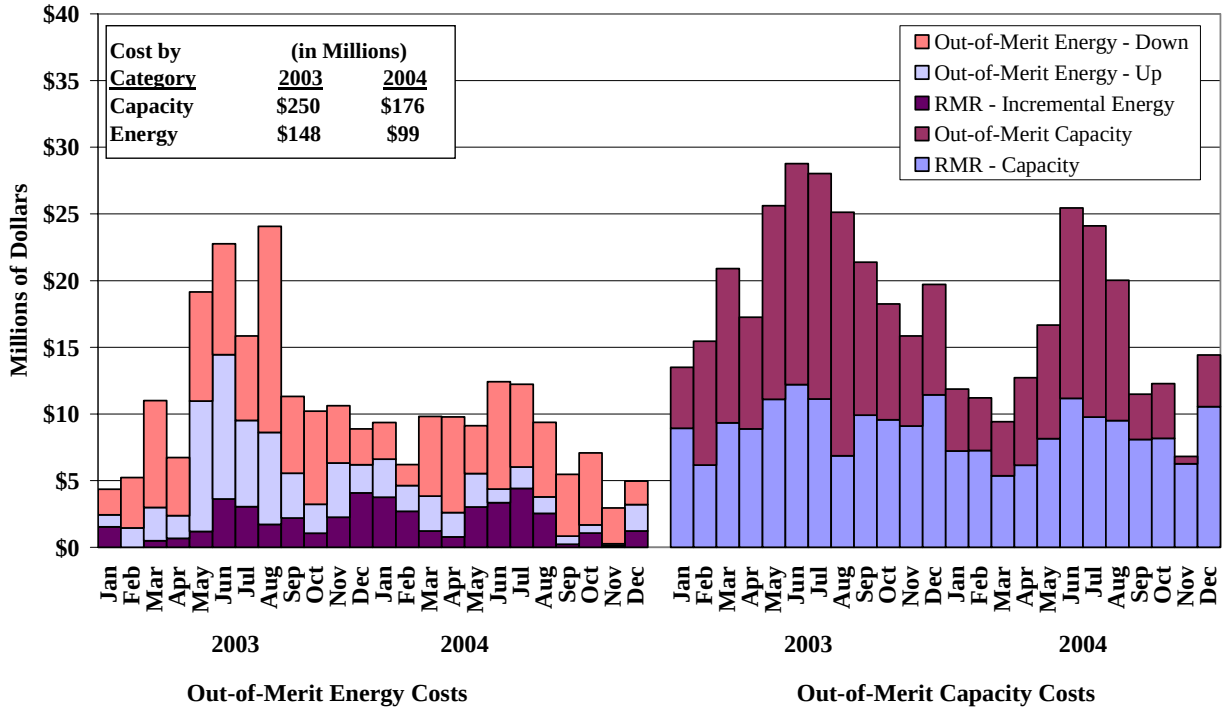
When a unit is dispatched out of merit (OOME up or OOME down), the unit is paid for a quantity equal to the difference between the scheduled output based on the unit's resource plan and the actual output resulting from the OOME instruction from ERCOT. The payment per MWh for OOME is a pre-determined 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. There were no RMR contracts in ERCOT prior to October of 2002. In response to AEP's announcement that they would place out-of-service all of its gas fired plants in ERCOT because it could buy power at a lower cost than operating the plants, ERCOT contracted with AEP for seven plants to provide RMR service beginning in October 2002. One unit at the Frontera plant in the Rio Grand Valley was also contracted to provide RMR service. During the spring of 2004, a unit at the Eagle Mountain plant was added to RMR contract status. Units contracted to provide RMR service to ERCOT are compensated for start-up costs, energy costs, and are also paid a standby fee. The analyses in this section separate RMR uplift into two categories: (a) capacity costs, which include start-up costs, standby fees, and energy costs up to the minimum dispatch level, and (b) incremental energy costs, which are the costs associated with output above the minimum dispatch level. Figure 75 shows each of the five categories of uplift costs by month for 2003 and 2004.

Figure 75: Expenses for Out-of-Merit Capacity and Energy 2003 to 2004



The left side of Figure 75 shows costs of OOME (up and down) and incremental energy from RMR units, while the right side shows the net costs of RMR units and OOMC units. Net cost for RMR units includes only the portion of RMR payments that exceeds the value of energy produced from RMR units at the balancing energy price.

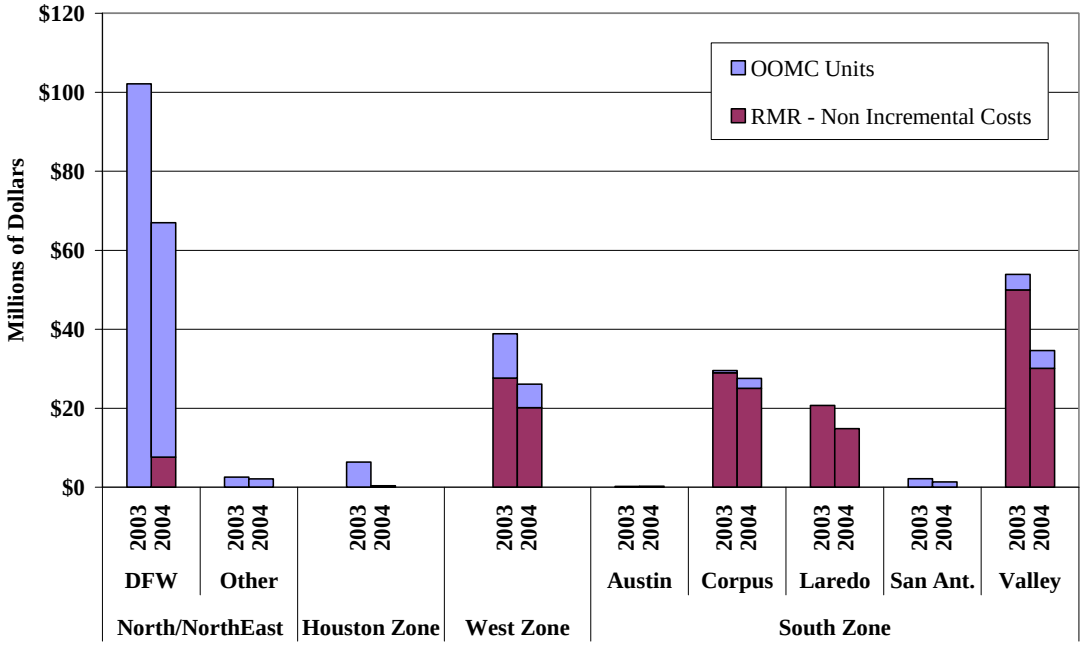
The results in Figure 75 show that OOME costs and incremental energy costs from RMR units declined from $148 million to $99 million from 2003 to 2004, a decrease of 33 percent. Likewise, the costs of OOMC and the capacity costs from RMR units declined 30 percent in 2004. The most substantial percentage decrease in these costs between 2003 and 2004 was associated with payments for OOME-Up which declined 64 percent. Out-of-merit costs are greater during the summer when higher loads increase the need for ERCOT operators to take out-of-merit actions to manage local congestion and reliability needs. However, RMR costs do vary substantially during the year because RMR payments are primarily designed to recover fixed costs, which are constant throughout the year.

Although the costs are borne by load throughout ERCOT, the costs are caused in specific locations because most of these actions 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 from 2003 to 2004. The first of these analyses focuses on the payments made for commitment of capacity, which include OOMC payments and net RMR payments excluding the portion for incremental energy dispatch. Figure 76 shows these payments by location.

Commitment-related uplift costs decreased or stayed even from 2003 to 2004 in each of the areas shown in the figure above. The total commitment-related uplift costs, including OOMC and RMR payments for capacity costs, decreased from $250 million in 2003 to $176 million in 2004, a decrease of nearly 30 percent. This was led by a decrease in OOMC costs of 41 percent. The largest source of OOMC uplift costs is the Dallas/Fort Worth area, accounting for 75 to 80 percent of the OOMC costs in 2003 and 2004.

Figure 76: Expenses for OOMC and RMR by Region
2003 & 2004



Significant transmission upgrades in the Dallas/Fort Worth area have led to significant reductions in OOMC costs associated with that area. Changes to the compensation formulas for OOMC units in February 2004 also contributed to reductions in OOMC costs. According to the ERCOT Protocols, ERCOT pays OOMC units for (a) starting-up and (b) staying on-line. Previously, payment formulas for starting-up and staying on-line were not dependent on the balancing energy price, which caused problems for two reasons. First, there was no guarantee

that the sum of the start-up payment, operating payment, and revenue from the balancing energy market would be sufficient for a unit to recover its costs.

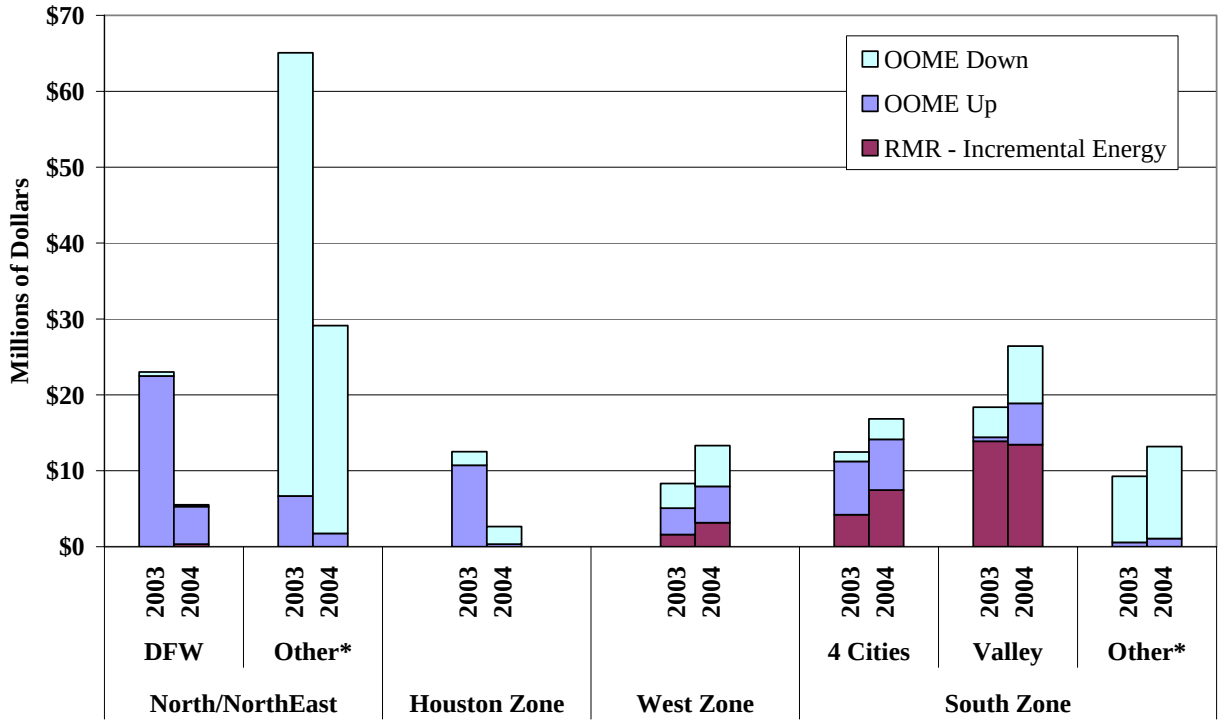
Second, units would receive the same uplift payment regardless of whether the balancing energy market revenue at the prevailing price was compensatory. This created a disincentive for QSEs to voluntary commit resources that were frequently needed for local reliability. It was often more profitable to wait for the resources to be committed through the OOMC process. We believe the current formulas have mitigated this incentive problem by making it less profitable to have ERCOT commit resources through the OOMC process when prices are expected to be high enough to cover the resources' commitment costs. This change has likely contributed to the lower OOMC commitment costs in 2004.

The next analysis reviews the costs incurred by ERCOT to dispatch generating resources out of merit to resolve local congestion. The costs are incurred in the form of OOME up and OOME down payments, as well as payments to RMR resources for incremental energy above minimum generation.[48] Figure 77 shows annual uplift costs for units providing OOME by region and by zone.

The figure shows that uplift for OOME decreased significantly in the North, Northeast, and Houston zones, and increased substantially in the West and South zones relative to 2003. In these three zones, uplift for OOME Up deployments decreased from $40 million to $7 million and uplift for OOME Down deployments decreased from $61 million to $30 million. The uplift payments for out-of-merit dispatch to resources in the Houston zone decreased 79 percent from 2003 to 2004. The dramatic reduction in out-of-merit dispatch is related to the creation of a new CSC in 2004. The North-to-Houston CSC allows the balancing energy market to resolve the congestion on the 345 kV lines directly connecting the North zone to Houston.

[48] Local balancing energy is included in the OOME costs as described above.

Figure 77: Expenses for OOME by Region
2003 to 2004



* The "Other" category includes portfolio deployments where the operator does not specify a single resource for deployment.

In the West zone and South zone, the portion of uplift paid to RMR units for incremental energy rose from $20 million to $24 million. In addition, uplift for OOME Up deployments increased from $12 million to $18 million and uplift for OOME Down deployments increased from $17 million to $28 million.

Several changes were adopted in ERCOT during 2003 and 2004 that led to dramatic reductions in out-of-merit dispatch in the North and Northeast Zones. Dallas/Fort Worth is a load pocket where ERCOT dispatches units up to relieve congestion, while the Northeast is a generation pocket where ERCOT dispatches units down to relieve congestion. Before 2004, these areas were within the same pricing region and so congestion between the areas was managed as local congestion. The addition of the Northeast Zone has contributed to the reduction in OOME Down dispatch within the Northeast and OOME Up in DFW and other areas within the North zone.

We also attribute some of the reduction in local congestion costs in 2004 to the suspension of the "Market Solution" method used to solve local transmission constraints. Prior to July 18, 2003, a

Market Solution would exist whenever three or more unaffiliated suppliers were capable of relieving a local constraint. When a Market Solution existed, SPD would select the most cost effective resource(s) based on the shift factors and offer premiums for each resource. Incremented resources were paid the energy clearing price plus their offer premium. Likewise during this period, ERCOT paid decremented resources their offer premium as compensation for the lost opportunity of producing at the market price.

ERCOT discontinued the Market Solution process because, in practice, the outcomes often were non-competitive and ERCOT frequently dispatched resources with offer premiums approaching $1,000/MWh. Market Solutions accounted for approximately $22.8 million in uplift payments from January to July 2003 to relieve relatively small amounts of congestion in real-time. While use of the Market Solution only occurred in 7.5 percent of intervals, they accounted for approximately 30 percent of the uplift for dispatch to manage local constraints during this period.

In summary, there have been significant reductions in expenses for out-of-merit commitment and dispatch actions in 2004. In particular, the formation of the new zone and CSCs has directly assigned congestion rents on these CSCs, shifting the burden of relieving congestion on these lines to the participants using these CSCs rather than uplifting the costs to all load in ERCOT.

E. Conclusions and Recommendations: Interzonal and Intrazonal Congestion

Consistent with the conclusions from the 2003 State of the Market Report and the Market Operations Report, the results in this section highlight significant opportunities for improvements in the operation of the ERCOT markets. These results indicate that in 2004, the vast majority of the congestion costs are associated with intrazonal congestion. This process results in uplift that is difficult to hedge and that is inefficiently allocated to the load in ERCOT. The process also results in economic signals that are not transparent. In addition, the intrazonal congestion management procedures appear to provide incentives for some suppliers to submit inaccurate resource plans to increase the frequency of out of merit commitment and dispatch actions by ERCOT.

With regard to interzonal congestion, the report highlights significant issues related to the zonal assumptions used in the ERCOT market. These assumptions and the operation of the current markets in Texas have been evaluated in greater detail in the Market Operations Report issued

last November, which identified some significant issues related to congestion management processes in ERCOT [49] In addition, the results in this section of the report continue to indicate that:

- The current zonal market can result in large inconsistencies between the interzonal flows calculated by SPD and the actual flows over the CSC interfaces; and
- These inconsistencies can result in under-utilized transmission capability and difficulties in defining transmission rights whose obligations can be fully satisfied.

The most complete long-run remedy for both the interzonal and intrazonal issues identified in this report would be to implement nodal markets, an option that is currently being evaluated in ERCOT. These markets would provide transparent prices for both generators and loads that would fully reflect all transmission constraints on the ERCOT network. Hence, we strongly recommend the continued development and implementation of such markets.

Absent implementation of nodal markets, we continue to recommend the following changes from the Market Operations Report to improve the management of interzonal and local congestion.[50]

- Improve the process for designating zones to minimize the effects of the simplifying zonal assumptions.
- Improve the process for evaluating and revising CSC definitions.
- Modify the calculation methodology of the zonal average shift factor to exclude generation whose output is generally fixed (e.g., nuclear units).
- Provide ERCOT the operational flexibility to temporarily modify the definition of a CSC associated with topology changes
- Modify the multi-step balancing energy market optimization to recognize the interactions between its local congestion management and zonal balancing energy deployments to minimize the costs of both classes of deployments.

---

49 See "2004 Assessment of the Operation of the ERCOT Wholesale Electricity Markets", Potomac Economics, November 2004.

50 The Commission has opened Project No. 30634, *Activities Related to Implementation of Recommendations from the Potomac Economics 2004 Report on the Operation of the ERCOT Wholesale Electricity Markets*, to address these recommendations.

The Protocols have been revised to address the first four recommendations.[51] However, a decision on the last recommendation has been deferred pending a decision on whether ERCOT will move to a nodal market design. The last recommendation is particularly important because local congestion management can have large indirect effects on portfolio energy deployments and the balancing energy prices. In the Market Operations Report, we concluded that current multi-step process does not efficiently consider the interaction between actions taken to resolve local congestion versus those taken to resolve interzonal congestion, resulting in inefficient market results and artificial price spikes in the balancing energy market. The last recommendation addresses this concern.

[51] See PRRs 587 and 589, effective July 1, 2005, and PRR 592, effective November 1, 2005.

VII. 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 2004.

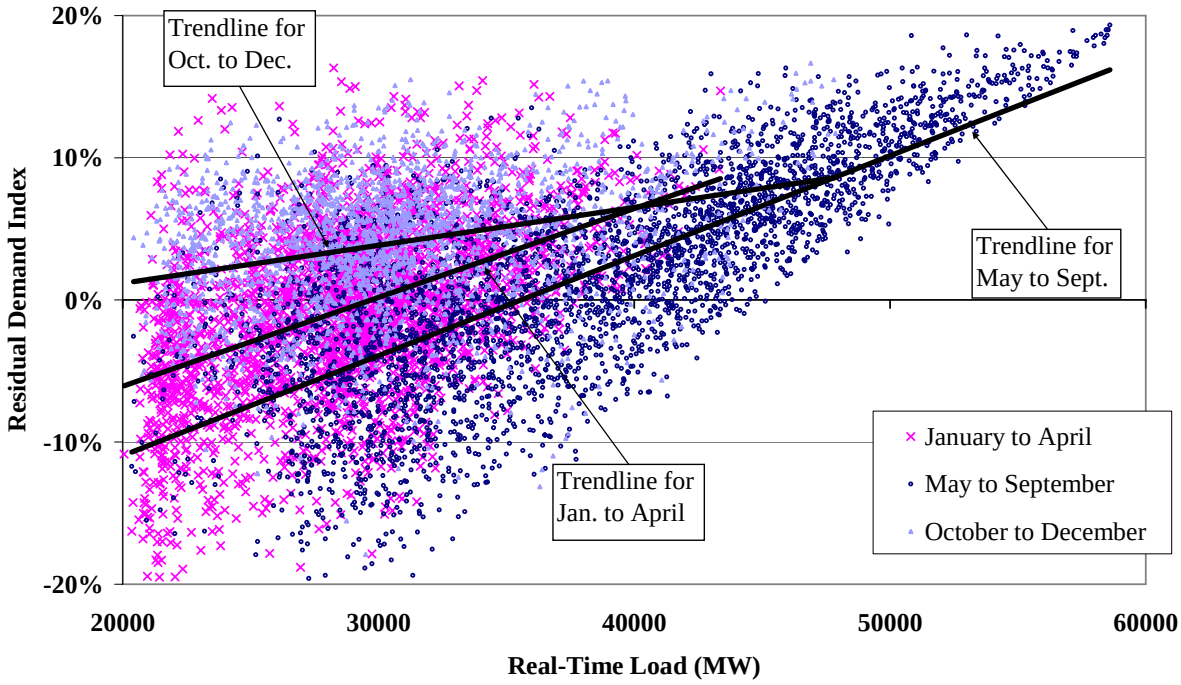
A. Structural Market Power Indicators

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 78 shows the RDI in 2004 for three separate time periods relative to load. The three periods are: (i) the spring period from January to April, (ii) the summer months from May to September, and (iii) the fall period from October to December. The trend lines for each data series are also shown and indicate a strong positive relationship between load and the RDI. This relationship is expected since the quantity of resources available from competing QSEs would have to increase as load increases to keep the RDI from increasing. This analysis is done at the QSE level because the largest suppliers that determine the RDI values shown below own roughly 90 percent of the resources they are scheduling or offering. They may also control the remaining 10 percent 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 78: Residual Demand Index[52]
Spring, Summer, and Fall Hours - 2004



The figure shows that the RDI for May to September generally begins to be positive in many hours when load exceeds 35,000 MW. During the entire summer, the RDI was greater than zero in almost 60 percent of hours. For the January to April period, the RDI is generally positive when the load rises above 29,000 MW. The RDI is typically positive at lower load levels during the spring due to the large number of generation planned outages. Hence, although the load is lower during the spring, our analysis shows that a QSE is pivotal in almost 50 percent of hours during that period. The effects of the planned outages in the spring period are reflected in the difference between the trend lines for the two periods: the trend line for the spring hours is 3 percent to 4 percent higher than in the summer hours, indicating that the RDI was higher in the Spring.

During the fall months (October to December), demand levels are comparable to those of the spring period. However, Figure 78 indicates differences in the RDI values during the fall period.

[52] A similar analysis was shown in the 2003 SOM Report with RDI values that were generally lower than in Figure 78. The methodology used for Figure 78 is different because it only includes on-line and quick start capacity. In contrast, the analysis in the 2003 SOM Report included all in-service capacity. Using a more restrictive set of resources leads to higher RDI values in this report.

First, the trend line indicates that the RDI was generally much higher in the fall than in the spring. A QSE was pivotal in approximately 82 percent of hours between October and December. Furthermore, the flatter slope of the trend line indicates a weaker relationship between the RDI and demand level in the fall. This suggests that less capacity was available during the latter part of the year. The reduction in supplemental commitment through OOMC that occurred during these months contributed significantly to the reduction in available capacity.

It is important to recognize that inferences regarding market power cannot be made solely from this data. Some of the largest suppliers also serve substantial load, which causes them to be a much smaller *net* seller than the analysis above would indicate. 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 long-term contracts and load obligations. To account for this factor, we also calculated a load-adjusted RDI.

The "load-adjusted" RDI is adjusted for the load served by each supplier. Thus, a supplier with 3,000 MW of capacity and 2,000 MW of load would have a "load-adjusted capacity" of 1,000 MW and only the load-adjusted capacity is used in calculating the RDI. The supplier would not have the incentive to withhold more than 1,000 MW because it would have to purchase that additional amount from the balancing market to serve its load (assuming that the supplier has no other physical or financial contracts to purchase energy, which may or may not be the case). Because many suppliers may have substantial contractual positions and because some of the load may be served on a relatively short-term basis, the true RDI for the largest suppliers is likely to lie between unadjusted values shown in Figure 78 and the load-adjusted RDI values shown in Figure 79. Figure 79 shows the load-adjusted RDI for ERCOT as a function of the actual load level for spring, summer, and fall hours.

Figure 79: Load-Adjusted Residual Demand Index vs. Actual Load[53]
Spring, Summer, and Fall Hours -- 2004

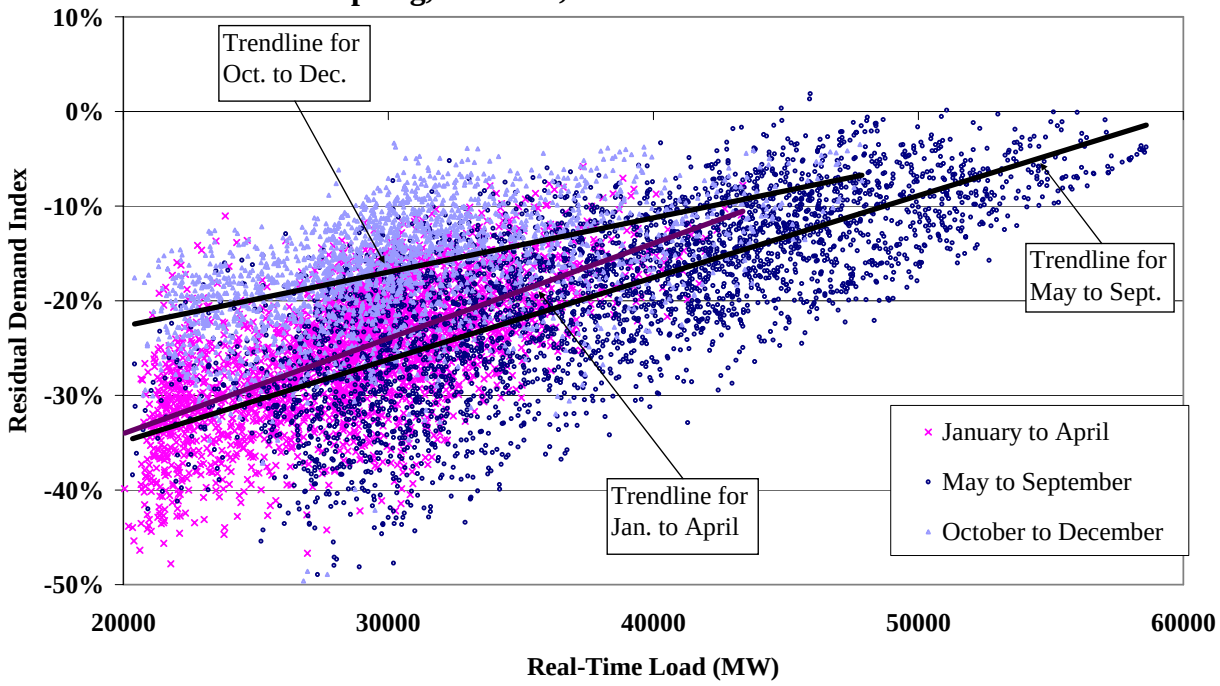


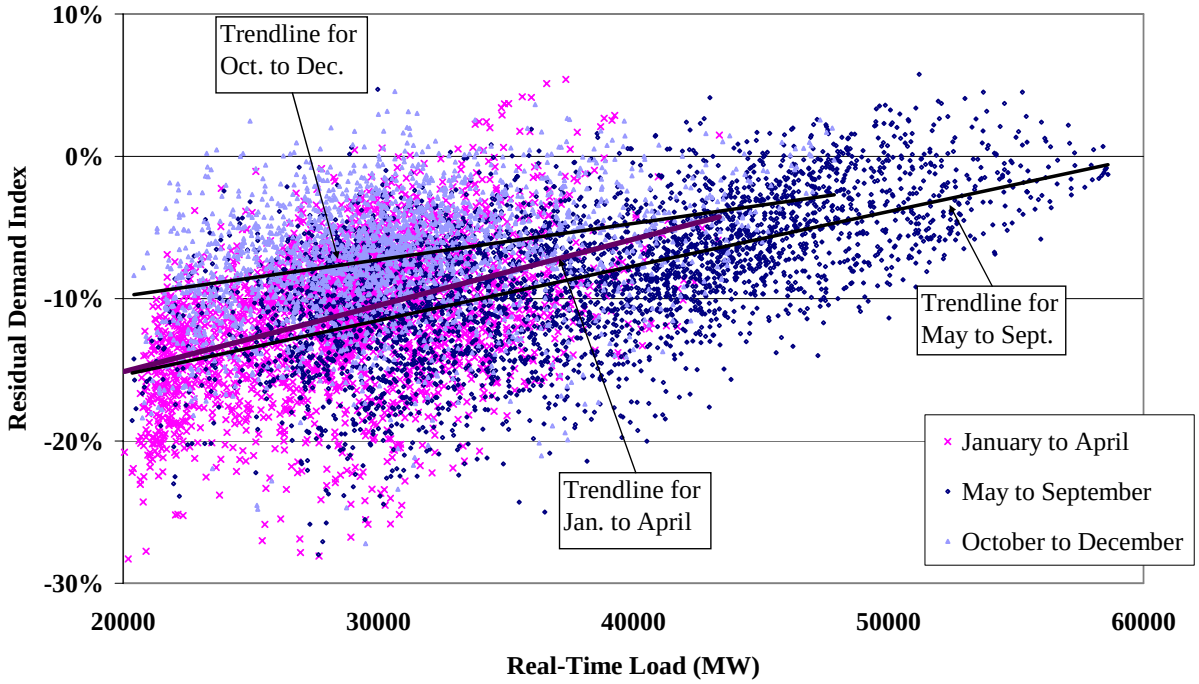
Figure 79 shows that there were only four hours in the summer with a positive RDI and none during the other times. Hence, for practical purposes, no suppliers were pivotal by this measure. This RDI measure does not consider the contractual position of the supplier, which can increase a supplier's incentive to exercise market power compared to the load-adjusted capacity assumption made in this analysis. The PUCT is now collecting bilateral contract information that could potentially be used to improve the accuracy of this measure. The load-adjusted RDI is significantly higher from October to December compared with the other portions of the year due to the reduced levels of on-line and quick start capacity during the latter portion of the year.

In addition, a supplier's ability to exercise market power in the current ERCOT balancing energy market will generally be higher than indicated by the load-adjusted RDI because a significant share of the available energy resources in real time are not offered in the ERCOT balancing market (as shown in prior sections of this report). Hence, a supplier may be pivotal in the balancing energy market when it would not have been pivotal more generally. To account for

53 A similar analysis was shown in the 2003 SOM Report with RDI values that were generally lower than in Figure 79. The methodology used for Figure 79 is different because it only includes on-line and quick start capacity. In contrast, the analysis in the 2003 SOM Report included all in-service capacity. Using a more restrictive set of resources leads to higher RDI values in this report.

this, we developed RDI statistics for the balancing energy market. Figure 80 shows the RDI in the balancing energy market relative to the actual load level.

Figure 80: Balancing Energy Market Residual Demand Index vs. Actual Load Spring, Summer, and Fall – 2004



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 owned by other suppliers. Figure 80 limits the other supplier's capacity to the energy 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 price spike in the balancing energy market.

Under the load adjusted scenario, while the RDI was negative in the majority of hours, it was positive in 3 percent of hours. The instances when the RDI was positive occurred over a wide range of load levels, from 25 GW to 60 GW. The RDI results for the balancing energy market shown in Figure 80 help explain how transient price spikes can occur under mild demand while large amounts of capacity are available in ERCOT. These results also show how QSEs offering only part of their available energy in the balancing energy market can cause the balancing energy market to be vulnerable to withholding and other forms of market abuses even when no suppliers

are fundamentally pivotal (i.e., the load-adjusted RDI is negative). This highlights the importance of modifying the current market rules and procedures to minimize any barriers or disincentives to full participation in the balancing energy market.

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 identify evidence of attempts 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 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

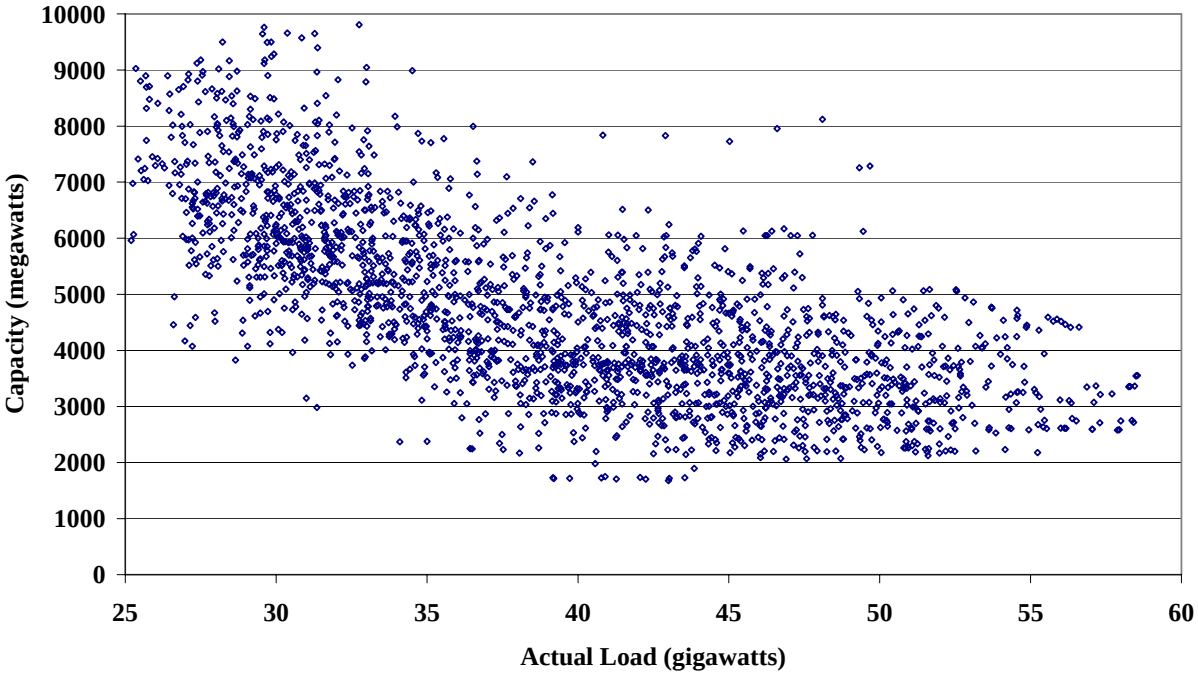
Physical withholding occurs when a participant makes resources unavailable for dispatch that are otherwise physically capable of providing energy and that are economic at prevailing market prices. This can be done by derating a unit or designating it as a forced outage. In any electricity market, deratings 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 78 and Figure 79 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 81 shows the relationship of short-term deratings and forced outages to real-time load levels in each hour during the summer months. We focus on these 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 resources and cogeneration resources are also excluded from this analysis given the high variation in the availability of these classes of resources.

Figure 81: Short-Term Deratings and Forced Outages vs. Actual Load
June to August, 2004



As the figure shows, short-term deratings and outages varied between just under 2 GW and 10 GW. Since the figure includes data from only three summer months, the lower load levels generally represent shoulder hours during weekends and nighttime. It is common for QSEs to

submit resource plans for some units with status "unavailable" during shoulder hours and status "available" during the daytime. This causes the data to show an inverse relationship between deratings and outages and real-time demand levels. At demand levels above 56 GW, the sum of deratings and outages were generally near or less than 4 GW. This is notable because at the highest demand levels, resources that are seldom dispatched and generally less reliable must be called on to satisfy the market's energy requirements. The practice of making certain units "unavailable" during shoulder hours became more common in 2004 and explains why similar data for 2003 shows a flatter downward trend in the shoulder hours. The results in Figure 81 are consistent with the conclusion that most suppliers have competitive incentives to increase their resource availability under peak demand conditions when energy sales are most profitable.

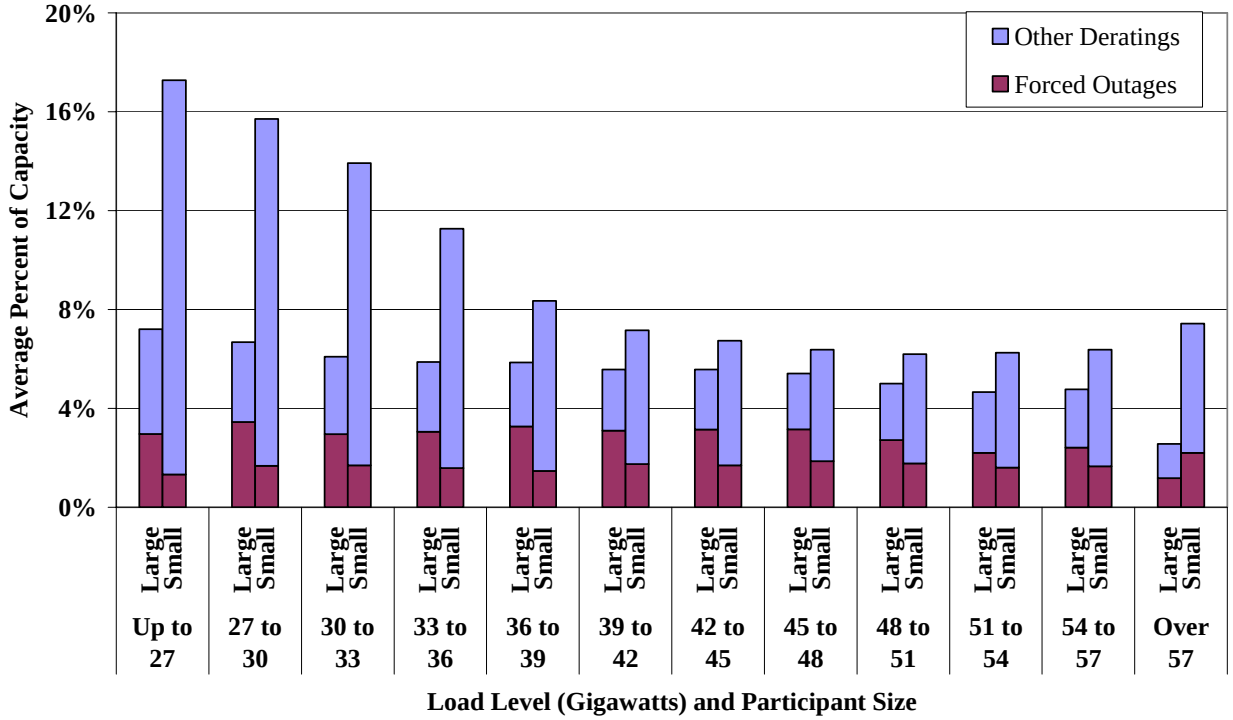
However, we further evaluate these trends by examining them by portfolio size. Portfolio size is important in determining whether individual suppliers have incentives to withhold available resources. Hence, the patterns of outages and deratings of large suppliers can be usefully evaluated by comparing them to the small suppliers' patterns.

Figure 82 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.[54] 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).[55]

[54] Like the prior analysis, long-term deratings and deratings by cogeneration and renewable energy resources are excluded.

[55] The four largest suppliers are Texas Utilities, Texas Genco, AEP, and Calpine

Figure 82: Short-Term Deratings by Load Level and Participant Size
June to August, 2004



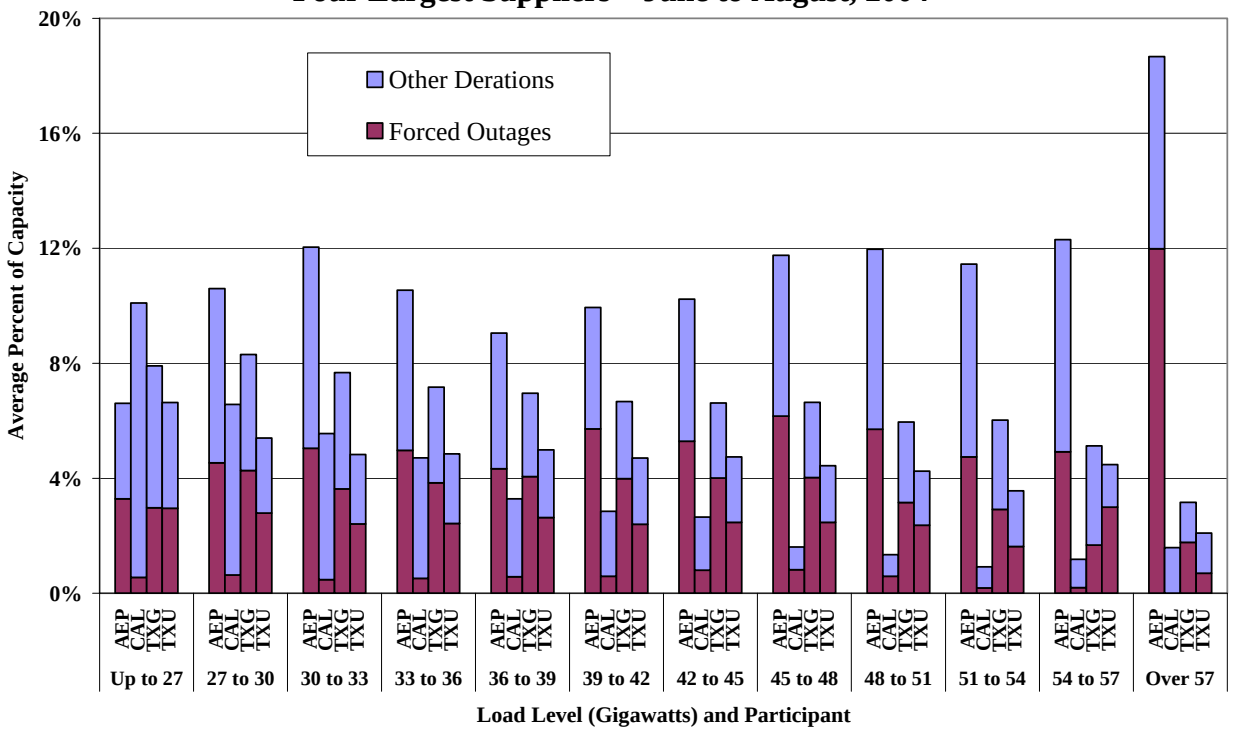
For large suppliers, the short-term derating or forced outage rates decreased from approximately 7 to 8 percent at low demand levels to about 2 to 4 percent at load levels above 54 GW. For small suppliers, the derating rates decreased from 10 to 17 percent at load levels below 36 GW to less than 7 percent at load levels above 54 GW. The deratings and outages for small suppliers rose to almost 8 percent in the small number of hours when demand exceeded 57 GW. As discussed above, the higher "other deratings" during lower load periods reflects a practice by some QSEs of designating some of their resources unavailable during weekends and nighttime periods.

Figure 82 also shows a distinction between forced outages and other deratings and indicates that a larger share of the large suppliers' deratings was comprised of forced outages. Given the extremely low forced-outage rates shown for small suppliers, it is likely that this difference is due, in part, to differences in forced outage reporting by smaller suppliers.

At all load levels, large suppliers have lower deratings rates than small suppliers. Furthermore, large suppliers' deratings and outages decline as load levels increase. Given that the market is most vulnerable to market power at the highest load levels, these derating patterns do not provide

evidence of physical withholding by the large suppliers. However, these results cannot exclude limited instances of withholding by either large or small suppliers. A more detailed analysis is needed before drawing conclusions about withholding behavior of either large or small suppliers. While investigating the specific conduct of large or small suppliers is beyond the scope of this report, the following figure summarizes the short-term deratings and outages of each of the largest four QSEs.

Figure 83: Short-Term Deratings by Load Level and Participant
Four Largest Suppliers – June to August, 2004



Whereas for TXU, Texas Genco, and Calpine, Figure 83 shows that the amounts of short-term deratings and outages generally decrease as load increases, AEP's deratings and outages are lowest when load is less than 30 GW, is relatively consistent when load is between 27 GW and 57 GW, and rises substantially when load is greater than 57 GW. While the level of deratings and outages for AEP raise some competitive concerns, the levels for Calpine, Texas Genco, and TXU are generally consistent with expectations of competitive conduct. The competitive concerns about AEP are diminished by several factors. First, AEP is the smallest of the four suppliers shown above making it less likely to adopt an aggressive withholding strategy. Second, a large share of AEP's fleet is under Reliability Must Run ("RMR") contracts that

greatly reduce their profits during high priced periods. This is because RMR units only retain 10 percent of the profits from selling into the balancing energy market. Third, AEP's fleet includes a large number of older units that are more likely to have unexpected deratings and forced outages. Based on the figures above, we cannot definitively conclude whether market participants have engaged in withholding, such instances can only be identified through a more detailed investigation.

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. Uncommitted capacity is considered to be in the output gap if the unit would have been substantially profitable given the prevailing balancing energy prices. The resource is counted in the output gap if its net revenue (market revenues less incremental production costs) exceeds the minimum commitment costs of the resource (including start-up and no-load costs) by a margin of at least $50 per MWh for its minimum output level over its minimum run-time.[56]

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

[56] The production costs are estimated using the Continuous Emissions Monitoring ("CEMS") data collected by the Environmental Protection Agency. This data is used to estimate incremental heat rates and heat input at minimum generation levels for ERCOT generating units. This analysis also assumes $4 per MWh variable operating and maintenance expenses. Whenever CEMS data is unavailable, minimum generation and incremental costs are estimated by looking at a sample of balancing energy prices that coincide with each resource's production over the previous 90 days.

patterns to those factors that increase the potential for market power, including load levels and portfolio size. Figure 84 shows the relationship between the output gap from committed resources and real-time load for all hours during 2004.

Figure 84: Output Gap from Committed Resources vs. Actual Load 2004

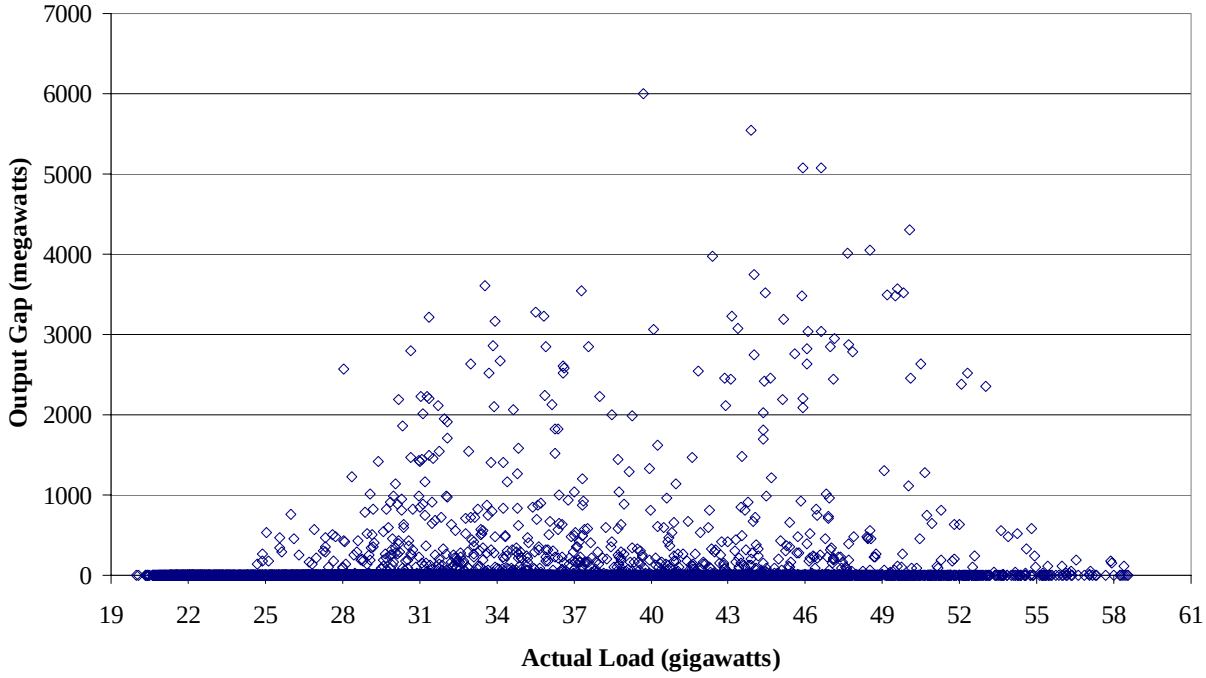


Figure 84 shows that the output gap from committed resources ranged from zero in most hours to a maximum of over 6000 MW during 2004. This figure also shows that there is no clear relationship between the output gap and real-time demand. The high output gap values generally occurred during transitory price spikes that occurred at a wide range of demand levels and tend to 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 dictate that resources cannot respond instantaneously to an unpredicted price spike. Even quick-start resources are sometimes unable to come on-line quickly enough to an unforeseen transitory price spike. The next analysis further examines the output gap results by size of supplier and load level.

Figure 85 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,[57] whereas the small supplier category includes the remaining suppliers that control 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 85: Output Gap by Load Level and Participant Size
2004

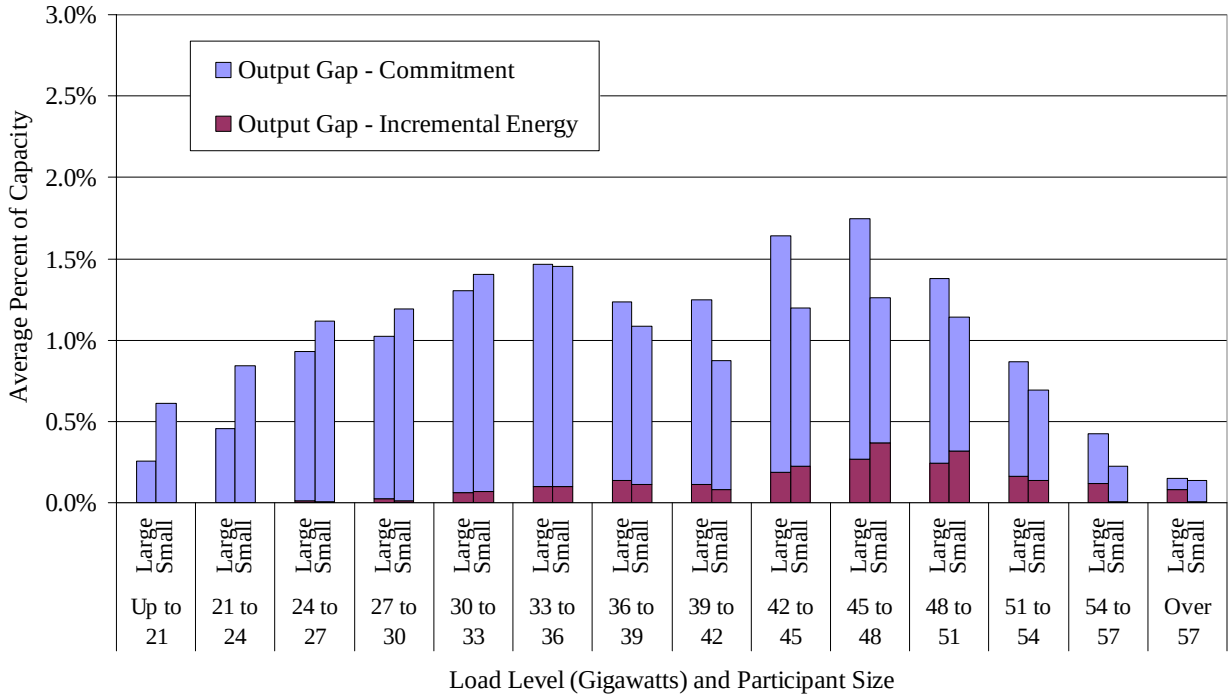


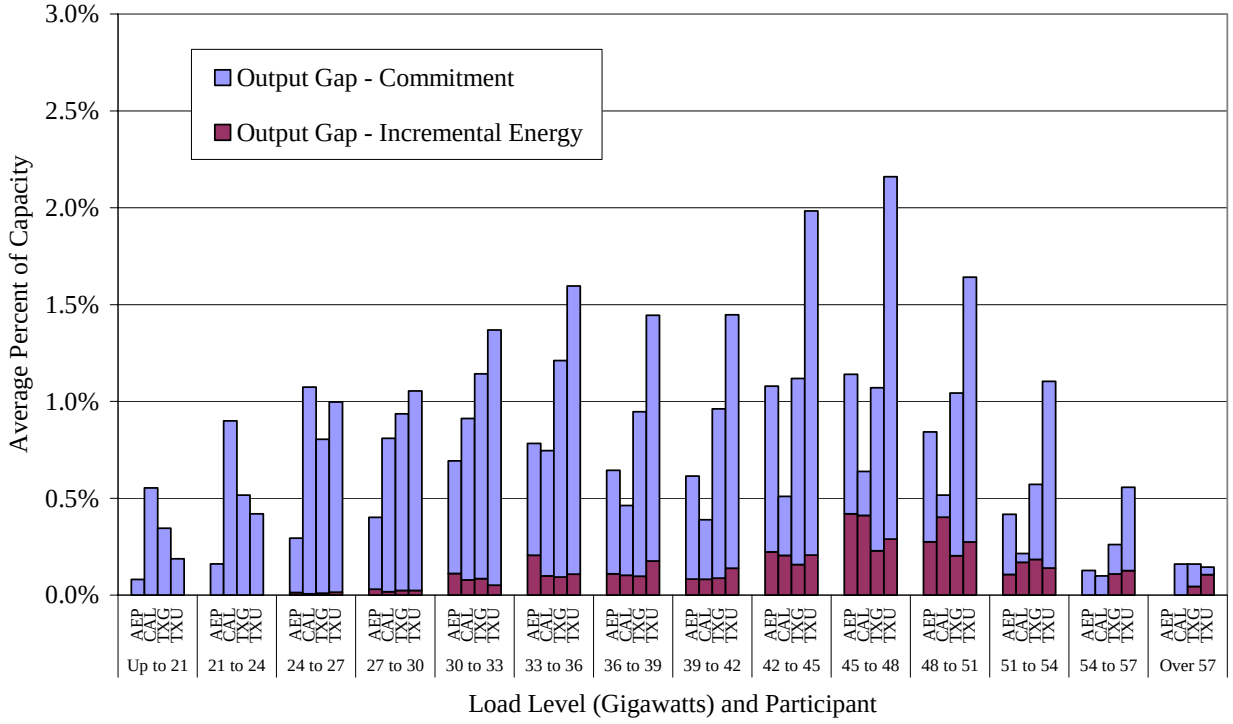
Figure 85 shows that compared to small suppliers, the large suppliers' total output gap was lower at load level below 33 GW but was higher at all other load levels. Additionally, the output gap associated with incremental energy on committed resources showed no clear pattern between large and small suppliers. For both large and small suppliers, the output gap declined as load rose above 48 GW. To the extent one expects the output gap to be a reflection of market power when load is the highest, under a hypothesis of market power, the output gap should increase as load increases. These results do not indicate that the large suppliers engaged in economic withholding during the highest load periods. However, as we have shown earlier in this report, the ERCOT balancing energy market frequently exhibits tight conditions or shortages when loads are not at peak levels for a variety of reasons. Therefore, the higher output gap amounts

[57] The four largest suppliers are Texas Utilities, Texas Genco, AEP, and Calpine.

from the large suppliers during mid-load periods may indicate economic withholding, although the quantities remain relatively low.

Large suppliers' output gap increased from close to zero at low demand levels to over 1.5 percent when demand levels were between 45 and 48 GW. For small suppliers, the output gap increased in a similar pattern from close to zero at low demand to almost 1.5 percent at load levels between 33 and 36 GW. At the very highest load levels, large and small suppliers' output gaps decreased to close to 0 percent. The following figure examines the output gap quantities for the large suppliers more closely by showing the information from Figure 85 separately for each of the four largest QSEs.

Figure 86: Output Gap by Load Level and Participant Size
2004



The output gap quantities shown in Figure 86 indicate that each of the four largest suppliers exhibit a similar pattern as load rises. In each case, the output gap rises from low load levels up to around 45 GW and then decreases as load rises above 45 GW. For each of the four QSEs, the incremental output gap (the darker bottom bar) is quite small, less than 0.5 percent of in-service capacity. The majority of the output gap is associated with off-line units (the lighter top bar) that would have been very profitable if committed. However, in many cases, the owner of the unit

would not know this ahead of time since price spikes frequently occur unexpectedly in ERCOT. Overall, TXU exhibits the largest amount of output gap from off-line units, although the average level, 1 to 2 percent of TXU's in-service capacity, is still relatively small.

Based on the analyses in this section of the report, there is no clear indication that suppliers have systematically exercised market power by economically or physically withholding capacity. However, this report is limited to evaluating overall patterns of conduct across the entire year. Isolated instances of significant physical or economic withholding would generally need to be identified on a case-specific basis.

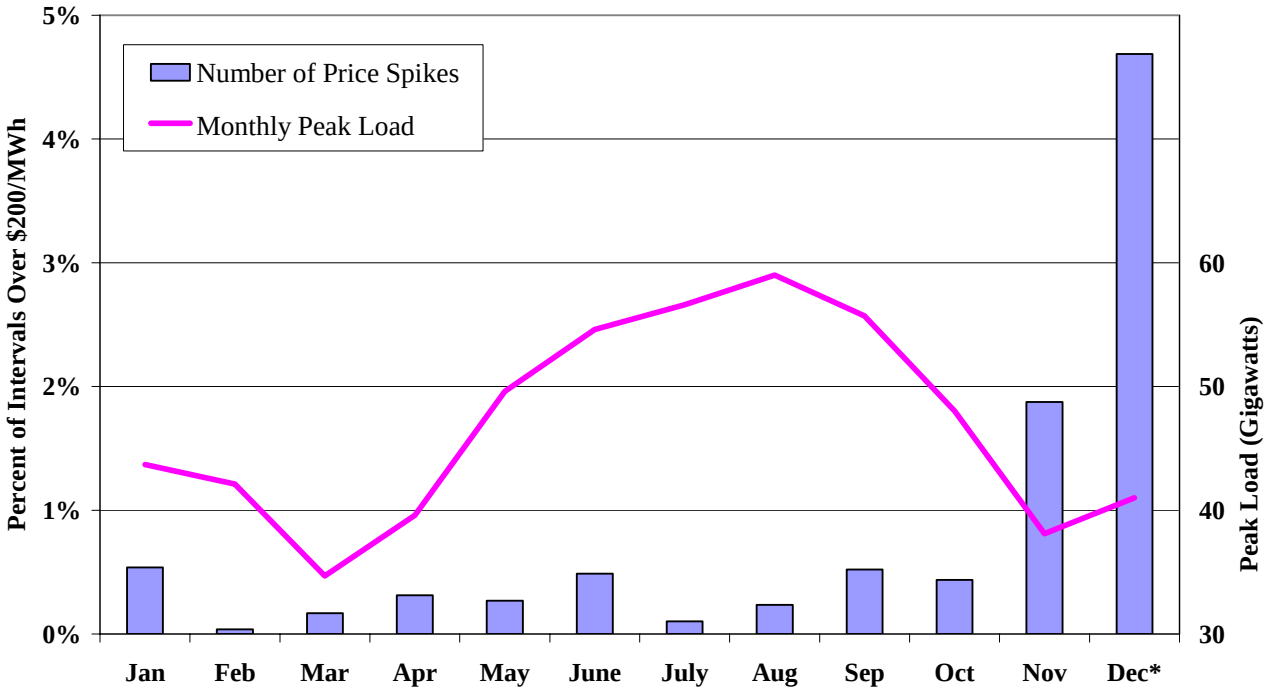
3. Investigation of Price Spikes from October 27 to December 8, 2004

ERCOT experienced a significant increase in the frequency of relatively high-priced intervals ("price spikes") in its balancing energy market during the fourth quarter of 2004.[58] However, these increases in prices were not precipitated by generation outages or changes in the transmission network. In fact, the price spikes occurred during off-peak months that are typically characterized by relatively low prices.

Figure 87 shows the frequency of price spikes and the peak demand level in each month during 2004. The figure shows that while monthly peak loads reached relatively low levels of 38 GW and 41 GW during November and December, the frequency of $200/MWh price spikes was four to nine times that of any other month. Compared with historic prices, the frequency of price spikes during November and December was unusual.

[58] For purposes of this investigation, we define "high-priced intervals" or "price spikes" as intervals with balancing energy prices exceeding $200 per MWh. We chose this level because prices higher than $200 per MWh exceed the competitive offer prices of most generating resources.

Figure 87: Frequency of $200 Price Spikes versus Peak Load
ERCOT Balancing Energy Auction – 2004



* Includes from December 1 to December 8, 2004.

The Public Utility Commission of Texas staff commissioned a report to assess the factors that led to this increase in price spikes for the period from October 27 to December 8, 2004 ("the study period"), the period during which the price spikes occurred.[59] In particular, we evaluate whether the high prices were the result of actions by TXU that constituted market power abuses.

The report focuses on the conduct of TXU because it changed its offer patterns substantially during the study period. TXU implemented its Rational Bidding Strategy ("RBS"), offering the energy from its quick-start gas turbines for more than $400/MWh, which is considerably higher than past offer prices associated with those units.[60] Moreover, these offers set the price in all of the high-priced intervals identified during this period and prices would have been substantially lower had TXU not employed its RBS.

59 *Investigation Into the Causes for the Shortages Of Energy in the ERCOT Balancing Energy Market and into the Wholesale Market Activities of TXU From October 27 To December 8, 2004*, Potomac Economics, April 2005.

60 TXU Response to Request for Information #1.

Based on our analysis of TXU's Rational Bidding Strategy, we found that the strategy was not consistent with competition and contributed to a significant increase in balancing energy prices during the study period. Prices during the high-priced intervals would generally have cleared at roughly 50 percent lower had TXU offered its gas turbines at competitive price levels.

However, we also identified a relatively large quantity of available energy that could have been produced from on-line and quick-start resources by rival suppliers that was not offered in the balancing energy market. If all of this energy had been offered, the price spikes would not have occurred. This reinforces the point that when significant quantities of available energy are not offered by smaller suppliers, it increases the ability of larger suppliers to increase the spot energy prices. For example, this un-offered energy is largely the reason why the results of the residual supply index in Figure 80 for the Balancing Energy Market are significantly worse (i.e., increased frequency of suppliers being pivotal) than the load-adjusted results in Figure 79.

As discussed earlier in this report, we believe that most of the un-offered energy is not offered due to barriers and economic risks inherent in the balancing energy market, rather than to physical or economic withholding by the smaller suppliers. Nonetheless, it does have a significant effect on the competitiveness of the balancing energy market, since it it increases the ability of larger suppliers to increase the spot energy prices

APPENDIX A

Frequent OOMC Resources

Resource	OOMC Uplift per MWh of Production	QSE	Zone
GEN_HLSES_UNIT2	$54.65	TXU ELECTRIC CO (RES)	NORTH
GEN_HLSES_UNIT5	$37.58	TXU ELECTRIC CO (RES)	NORTH
GEN_ATKINS_ATKINSG6	$35.83	BRYAN TEXAS UTILITIES (RES)	NORTH
GEN_SILASRAY_SILAS_9	$35.31	BROWNSVILLE PUBLIC UTILITY BOARD TENASKA (RES)	SOUTH
GEN_HLSES_UNIT4	$35.30	TXU ELECTRIC CO (RES)	NORTH
GEN_EMSES_UNIT1	$34.02	TXU ELECTRIC CO (RES)	NORTH
GEN_MCSES_UNIT7	$29.80	TXU ELECTRIC CO (RES)	NORTH
GEN_FTPP_FTPP_G1	$29.55	AMERICAN ELECTRIC POWER TEXAS NORTH COMPANY	WEST
GEN_MCSES_UNIT6	$24.15	TXU ELECTRIC CO (RES)	NORTH
GEN_LHSES_UNIT1	$19.36	TXU ELECTRIC CO (RES)	NORTH
GEN_NLSES_UNIT2	$19.12	TXU ELECTRIC CO (RES)	NORTH
GEN_HLSES_UNIT3	$17.16	TXU ELECTRIC CO (RES)	NORTH
GEN_MCSES_UNIT8	$17.15	TXU ELECTRIC CO (RES)	NORTH
GEN_NLSES_UNIT1	$16.84	TXU ELECTRIC CO (RES)	NORTH
GEN_NLSES_UNIT3	$10.53	TXU ELECTRIC CO (RES)	NORTH

Frequent OOME Up Resources

Resource	OOM-Up Uplift per MWh of Production	QSE	Zone
GEN_DECKER_DPGT_4	$21.37	CITY OF AUSTIN DBA AUSTIN ENERGY (RES)	SOUTH
GEN_DECKER_DPGT_3	$14.52	CITY OF AUSTIN DBA AUSTIN ENERGY (RES)	SOUTH
GEN_DECKER_DPGT_1	$13.94	CITY OF AUSTIN DBA AUSTIN ENERGY (RES)	SOUTH
GEN_DECKER_DPGT_2	$12.22	CITY OF AUSTIN DBA AUSTIN ENERGY (RES)	SOUTH
GEN_PBSES_CT1	$11.56	TXU ELECTRIC CO (RES)	WEST
GEN_SPNCER_SPNCE_4	$11.25	CITY OF GARLAND (RES)	NORTH
GEN_PBSES_CT2	$10.61	TXU ELECTRIC CO (RES)	WEST
GEN_DCSES_CT1	$8.08	TXU ELECTRIC CO (RES)	NORTH
GEN_DCSES_CT2	$7.89	TXU ELECTRIC CO (RES)	NORTH
GEN_ATKINS_ATKINSG6	$7.53	BRYAN TEXAS UTILITIES (RES)	NORTH
GEN_DCSES_CT3	$7.12	TXU ELECTRIC CO (RES)	NORTH
GEN_DCSES_CT4	$6.74	TXU ELECTRIC CO (RES)	NORTH
GEN_MCSES_UNIT6	$4.20	TXU ELECTRIC CO (RES)	NORTH

Frequent OOME Down Resources

Resource	OOM-Down Uplift per MWh of Production	QSE	Zone
GEN_DUKE_DUKE_GT2	$2.45	BROWNSVILLE PUBLIC UTILITY BOARD TENASKA (RES)	SOUTH
GEN_NEDIN_NEDIN_G2	$1.99	CALPINE CORP	SOUTH
GEN_DUKE_DUKE_GT1	$1.76	BROWNSVILLE PUBLIC UTILITY BOARD TENASKA (RES)	SOUTH
GEN_NEDIN_NEDIN_G1	$1.59	CALPINE CORP	SOUTH
GEN_AMOCOOIL_AMOCO_G1	$1.34	SOUTH HOUSTON GREEN POWER LP	HOUSTON
GEN_DUKE_DUKE_ST1	$1.18	BROWNSVILLE PUBLIC UTILITY BOARD TENASKA (RES)	SOUTH