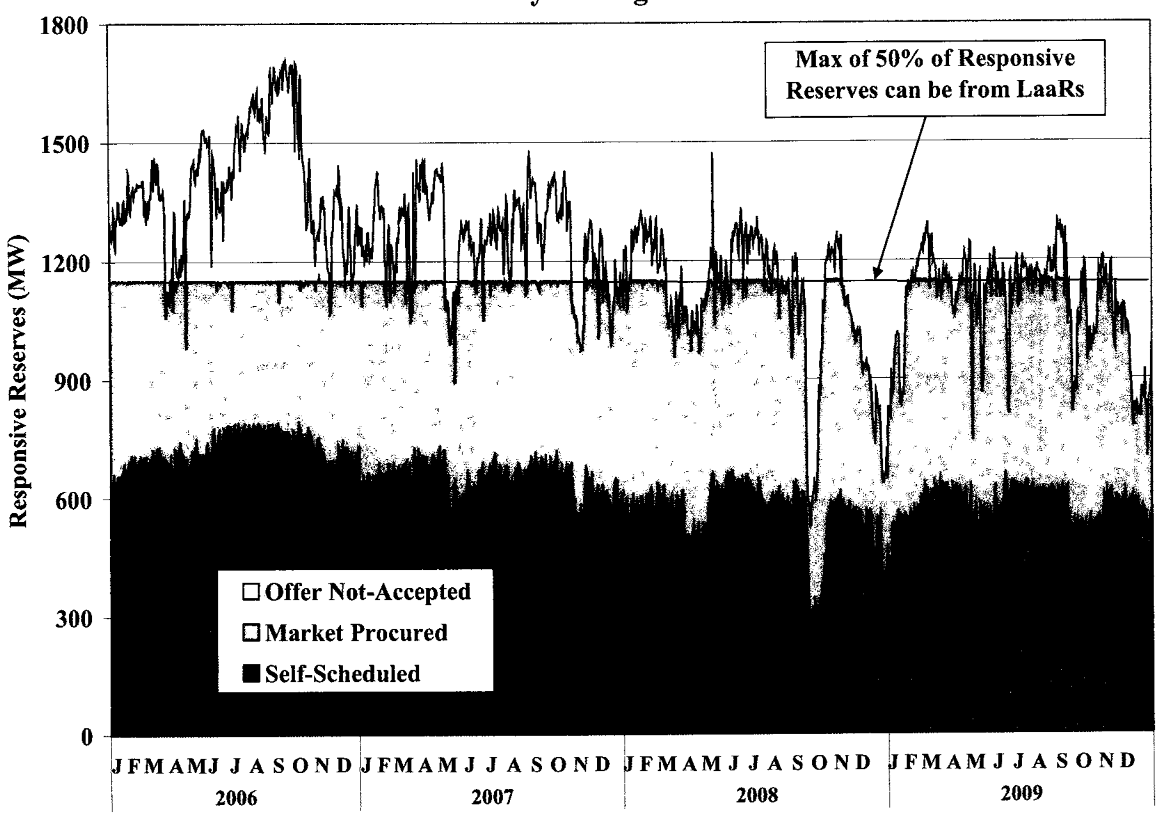
relay ("UFR") equipment. A load with UFR equipment is automatically tripped when the frequency falls below 59.7 Hz, which will typically occur only a few times in each year.

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

As of December 2009, over 2,200 MW of capability were qualified as LaaRs. These resources regularly provided reserves in the responsive reserves market, but never participated in the balancing energy market and only a very small portion participated in the non-spinning reserves market. Figure 41 shows the amount of responsive reserves provided from LaaRs on a daily basis in 2009.

Figure 41: Provision of Responsive Reserves by LaaRs
Daily Average



The high level of participation by demand response in the ancillary service markets sets ERCOT apart from other operating electricity markets. Figure 41 shows that the amount of responsive reserves provided by LaaRs has held fairly constant at 1,150 MW since the beginning of 2006. (For reliability reasons, 1,150 MW is the limit of participation in the responsive reserve market by LaaRs.) Exceptions include a decrease in September of 2008 corresponding to the Texas landfall of Hurricane Ike and a more prolonged reduction from November 2008 through January 2009 that was likely a product of the economic downturn and its effect on industrial operations.

Although LaaRs are active participants in the responsive reserves market, they did not offer into the balancing energy, regulation or non-spinning reserve services markets in 2009. This is not surprising because the value of curtailed load tends to be very high, and providing responsive reserves offers substantial revenue with very little probability of being deployed. In contrast, providing non-spinning reserves introduces a much higher probability of being curtailed. Participation in the regulation services market requires technical abilities that most LaaRs cannot meet at this point.

E. Net Revenue Analysis

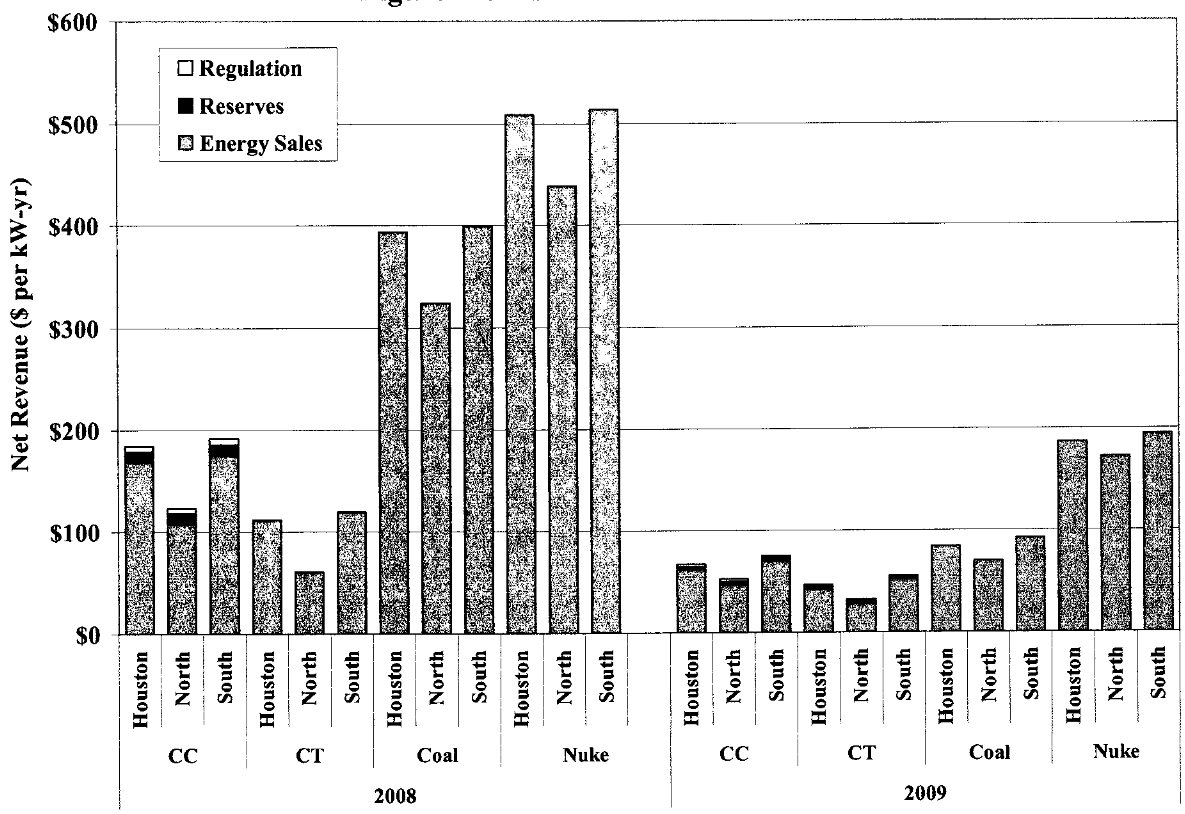
Net revenue is defined as the total revenue that can be earned by a generating unit less its variable production costs. Hence, it is the revenue in excess of short-run operating costs and is available to recover a unit's fixed and capital costs. Net revenues from the energy, operating reserves, and regulation markets together provide the economic signals that inform suppliers' decisions to invest in new generation or retire existing generation. In a long-run equilibrium, the markets should provide sufficient net revenue to allow an investor to break-even on an investment in a new generating unit. In the short-run, if the net short-run revenues produced by the market are not sufficient to justify entry, then one or more of three conditions exist:

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

Likewise, the opposite would be true if the markets provide excessive net revenues in the short-run. The persistence of excessive net revenues in the presence of a capacity surplus is an indication of competitive issues or market design flaws. In this section, we analyze the net revenues that would have been received by various types of generators in each zone.

Figure 42 shows the results of the net revenue analysis for four types of units in 2008 and 2009. These are: (a) a gas combined-cycle, (b) a combustion turbine, (c) a coal unit, and (d) a nuclear unit. In recent years, most new capacity investment has been in natural gas-fired technologies, although high prices for oil and natural gas have caused renewed interest in new investment in coal and nuclear generation. For the gas-fired technologies, net revenue is calculated by assuming the unit will produce energy in any hour for which it is profitable and by assuming it will be available to sell reserves and regulation in other hours that it is available (*i.e.*, when it is not experiencing a planned or forced outage). For coal and nuclear technologies, net revenue is calculated by assuming that the unit will produce at full output. The energy net revenues are computed based on the balancing energy price in each hour. Although most suppliers would receive the bulk of their revenues through bilateral contracts, the spot prices produced in the balancing energy market should drive the bilateral energy prices over time.

Figure 42: Estimated Net Revenue



For purposes of this analysis, we assume heat rates of 7 MMbtu per MWh for a combined cycle unit, 10.5 MMbtu per MWh for a combustion turbine, and 9.5 MMbtu per MWh for a new coal unit. We assume variable operating and maintenance costs of $4 per MWh for the gas units and

$5 per MWh for the coal unit. We assume fuel and variable operating and maintenance costs of $8 per MWh for the nuclear unit. For each technology, we assumed 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. The analysis also includes simplifying assumptions that can lead to over-estimates of the profitability of operating in the wholesale market. The following factors are not explicitly accounted for in the net revenue analysis: (i) start-up costs, which can be significant; and (ii) minimum running times and ramp restriction, which can prevent the natural gas generators from profiting during brief price spikes. Despite these limitations, the net revenue analysis provides a useful summary of signals for investment in the wholesale market.

Figure 42 shows that the net revenue decreased substantially in 2009 compared to each zone compared in 2008 and 2007. Based on our estimates of investment costs for new units, the net revenue required to satisfy the annual fixed costs (including capital carrying costs) of a new gas turbine unit ranges from $70 to $95 per kW-year. The estimated net revenue in 2009 for a new gas turbine was approximately $55, $47 and $32 per kW-year in the South, Houston and North Zones, respectively. For a new combined cycle unit, the estimated net revenue requirement is approximately $105 to $135 per kW-year. The estimated net revenue in 2009 for a new combined cycle unit was approximately $76, $67 and $52 per kW-year in the South, Houston and North Zones, respectively. These values indicate that the estimated net revenue in 2009 was well below the levels required to support new entry for a new gas turbine or a combined cycle unit in the ERCOT region.

Prior to 2005, net revenues were well below the levels necessary to justify new investment in coal and nuclear generation. However, high natural gas prices through 2008 allowed energy prices to remain at levels high enough to support new entry for these technologies. The production costs of coal and nuclear units did not change significantly over this period, leading to a dramatic rise in net revenues. With the significant decline in natural gas and energy prices in 2009, these results changed dramatically from recent years. For a new coal unit, the estimated net revenue requirement is approximately $190 to $245 per kW-year. The estimated net revenue in 2009 for a new coal unit was approximately $93, $84 and $70 per kW-year in the South, Houston and North Zones, respectively. For a new nuclear unit, the estimated net revenue requirement is approximately $280 to $390 per kW-year. The estimated net revenue in 2009 for a new nuclear unit was approximately $194, $187 and $172 per kW-year in the South, Houston and North Zones, respectively. These values indicate that the estimated net revenue for a new coal and nuclear unit in the South, Houston and North Zones was well below the levels required to support new entry in 2009.

Although estimated net revenue declined considerably in 2009 compared to the prior four years, there are other factors that determine incentives for new investment. First, market participants must anticipate how prices will be affected by the new capacity investment, future load growth, and increasing participation in demand response. Second, net revenues can be inflated when prices clear above competitive levels as a result of market power being exercised. Thus, a market participant may be deterred from investing in new capacity if it believes that prevailing net revenues are largely due to an exercise of market power that would not be sustainable after the entry of the new generation. Third, the nodal market design will have an effect on the profitability of new resources. In a particular location, nodal prices could be higher or lower than the prices in the current market depending on the pattern of congestion.

To provide additional context for the net revenue results presented in this section, we also compared the net revenue for natural gas-fired technologies in the ERCOT market with net revenue in other centralized wholesale markets. Figure 43 compares estimates of net revenue for each of the auction-based wholesale electricity markets in the U.S.: the ERCOT North Zone, the California ISO, the New York ISO, and PJM. The figure includes estimates of net revenue from energy, reserves and regulation, and capacity. ERCOT does not have a capacity market, and thus, does not have any net revenue from capacity sales.[20]

20 The California ISO does not report capacity and ancillary services net revenue separately, so it is shown as a combined block in Figure 43. 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.

Figure 43: Comparison of Net Revenue of Gas-Fired Generation between Markets

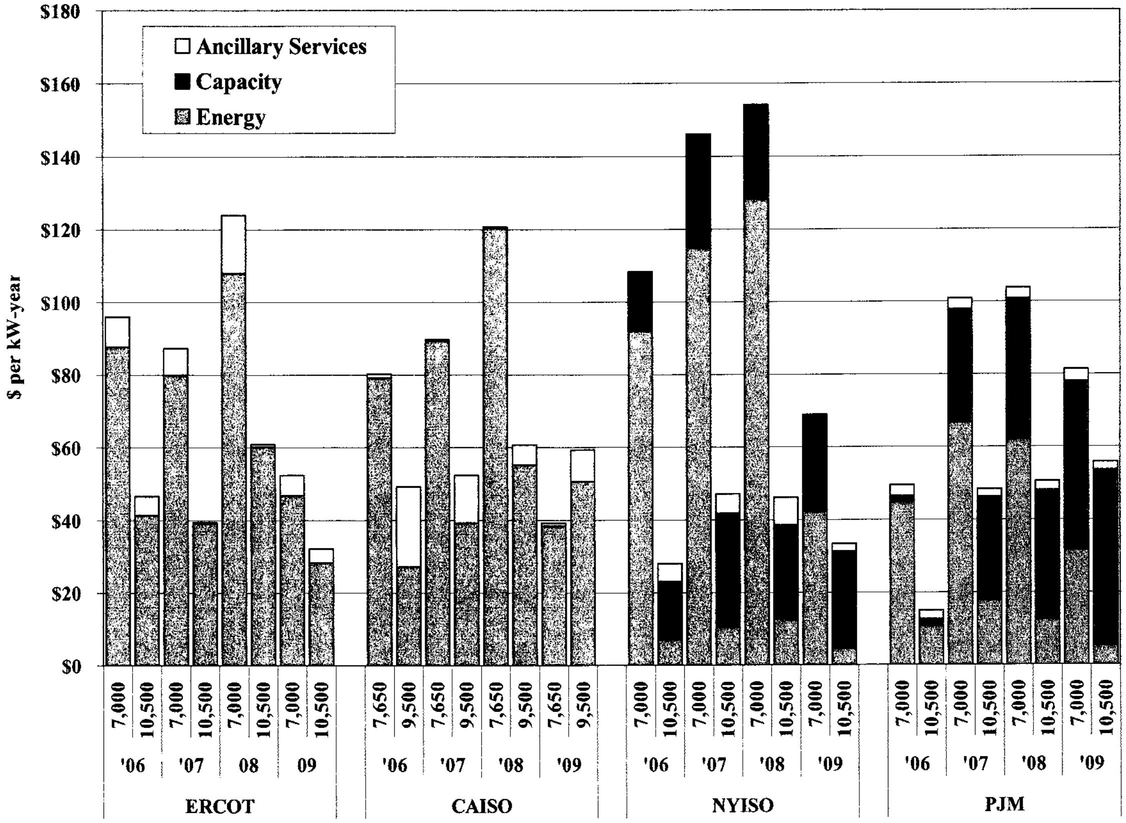


Figure 43 shows that net revenues decreased in all markets from 2008 to 2009, with the exception of gas peaking units in California ISO and PJM that remained flat. In the figure above, net revenues are calculated for central locations in each of the five markets. However, there are load pockets within each market where net revenue and the cost of new investment may be higher. Thus, even if new investment is not generally profitable in a market, it may be economic in certain areas. Finally, resource investments are driven primarily by forward price expectations, so historical net revenue analyses do not provide a complete picture of the future pricing expectations that will spur new investment.

F. Effectiveness of the Scarcity Pricing Mechanism

The PUCT adopted rules in 2006 that define the parameters of an energy-only market. These rules include a Scarcity Pricing Mechanism ("SPM") that relaxed the existing system-wide offer cap by gradually increasing it to $1,500 per MWh on March 1, 2007, $2,250 per MWh on March 1, 2008, and to $3,000 per MWh shortly after the implementation of the nodal market.

Additionally, market participants controlling less than five percent of the capacity in ERCOT by definition do not possess market power under the PUCT rules. Hence, these participants can submit very high-priced offers that, per the PUCT rule, will not be deemed to be an exercise of market power. However, because of the competition faced by the small market participants, the quantity offered at such high prices – if any – is very small.

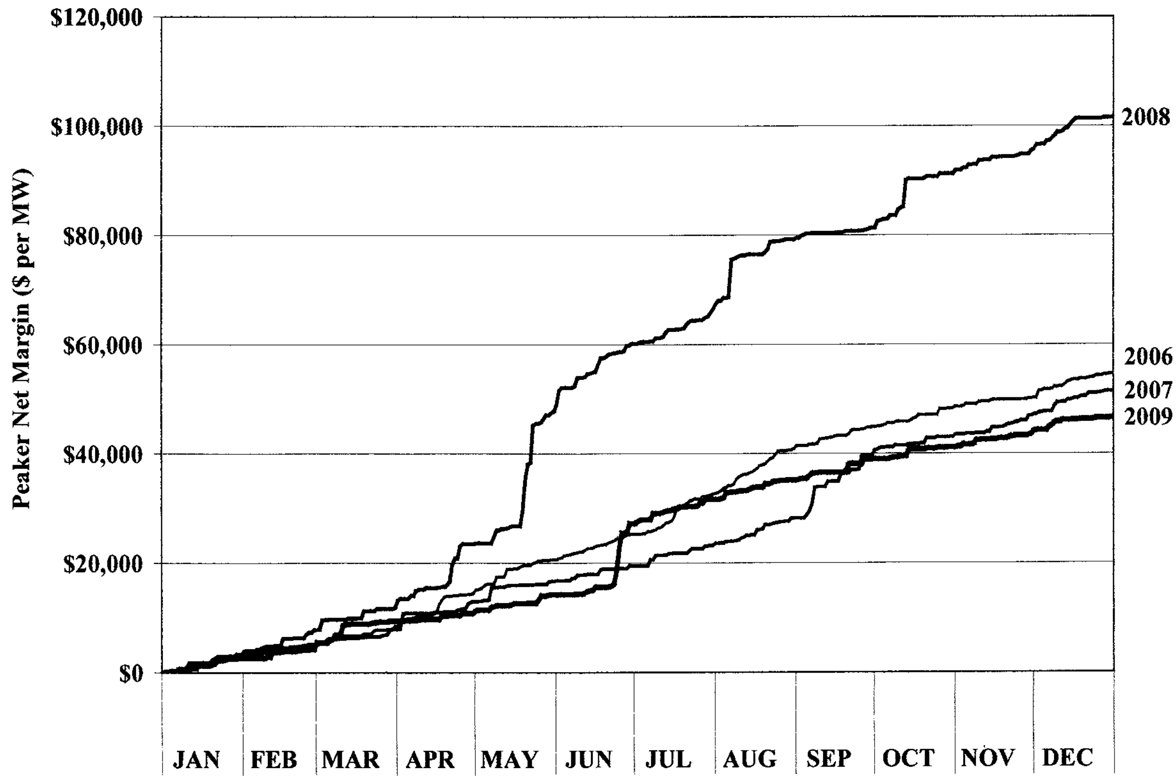
PUCT Subst. Rule 25.505 provides that the IMM may conduct an annual review of the effectiveness of the SPM. This subsection provides an assessment of the effectiveness of the SPM in 2009 under ERCOT's energy-only market structure.

Unlike markets with a long-term capacity market where fixed capacity payments are made to resources across the entire year regardless of the relationship of supply and demand, the objective of the energy-only market design is to allow energy prices to rise significantly higher during legitimate shortage conditions (*i.e.*, when the available supply is insufficient to simultaneously meet both energy and operating reserve requirements) such that the appropriate price signal is provided for demand response and new investment when required. During non-shortage conditions (*i.e.*, most of the time), the expectation of competitive energy market outcomes is no different in energy-only than in capacity markets.

Hence, in an energy-only market, it is the expectation of both the magnitude of the energy price during shortage conditions and the frequency of shortage conditions that will attract new investment when required. In other words, the higher the price during shortage conditions, the fewer shortage conditions that are required to provide the investment signal, and vice versa. While the magnitude of price expectations is determined by the PUCT energy-only market rules, it remains an empirical question whether the frequency of shortage conditions over time will be optimal such that the market equilibrium produces results that satisfy the reliability planning requirements (*i.e.*, the maintenance of a minimum 12.5 percent planning reserve margin).

The SPM includes a provision termed the Peaker Net Margin ("PNM") that is designed to measure the annual net revenue of a hypothetical peaking unit. Under the rule, if the PNM for a year reaches a cumulative total of $175,000 per MW, the system-wide offer cap is then reduced to the higher of $500 per MWh or 50 times the daily gas price index. Figure 44 shows the cumulative PNM results for each year from 2006 through 2009.[21]

Figure 44: Peaker Net Margin



As previously noted, the net revenue required to satisfy the annual fixed costs (including capital carrying costs) of a new gas turbine unit is approximately $70 to $95 per kW-year (i.e., $70,000 to $95,000 per MW-year). Thus, as shown in Figure 44 and consistent with the previous findings in this section relating to net revenue, the PNM reached the level sufficient for new entry in only one of the last four years (2008). In 2008, the peaker net margin and net revenue values rose substantially, surpassing the level required to support new peaker entry. However, a significant portion of the net revenue increase in 2008 was associated with extremely inefficient transmission congestion management and inefficient pricing mechanisms associated with the deployment of non-spinning reserves.[22] Both of these issues were corrected in the zonal market and will be further improved with the implementation of the nodal market in late 2010. With

21 The proxy combustion turbine in the Peaker Net Margin calculation assumes a heat rate of 10 MMBtu per MWh and includes no other variable operating costs or startup costs.

22 *See* 2008 ERCOT SOM Report at 81-87.

these issues addressed, the peaker net margin dropped substantially in 2009. Net revenues also dropped substantially for other technologies largely due to significant decreases in natural gas prices in 2009, but decreased natural gas price are not the driver for the reduction in net revenues for peaking resources. Beyond the correction of the market design inefficiencies that existed in 2008, there were three other factors that influenced the effectiveness of the SPM in 2009:

- A continued strong positive bias in ERCOT's day-ahead load forecast – particularly during summer on-peak hours – that creates the tendency to regularly commit online resources in excess of the quantity required to meet expected demand and operating reserve requirements;
- The implementation of PRR 776, which allows for quick-start gas turbines providing non-spinning reserves to offer the capacity into the balancing energy market; and
- The dependence on market participants to submit offers at or near the offer cap to produce scarcity level prices during legitimate operating reserve shortage conditions.

1. ERCOT Day-Ahead Load Forecast Error

ERCOT procedures include the operation of a day-ahead Replacement Reserve Service ("RPRS") market that is designed to ensure that adequate capacity is available on the system to meet reliability criteria for each hour of the following operating day. This includes an assessment of the capacity necessary to meet forecast demand and operating reserve requirements, as well as capacity required resolve transmission constraints.

An integral piece of the RPRS market is the day-ahead load forecast. If the day-ahead load forecast is significantly below actual load and no subsequent actions are taken, ERCOT may run the risk of there not being enough generating capacity online to meet reliability criteria in real-time. In contrast, if the day-ahead load forecast is significantly high, the outcome may be an inefficient commitment of excess online capacity in real-time.

Figure 45 shows the day-ahead load forecast error data for 2007 through 2009 with the average megawatt error displayed for each month in four hour blocks (hours ending). This figure shows a continuing bias toward over-forecasting summer peak loads by an average of 2,000MW.

Figure 45: Average Day Ahead Load Forecast Error by Month and Hour Blocks

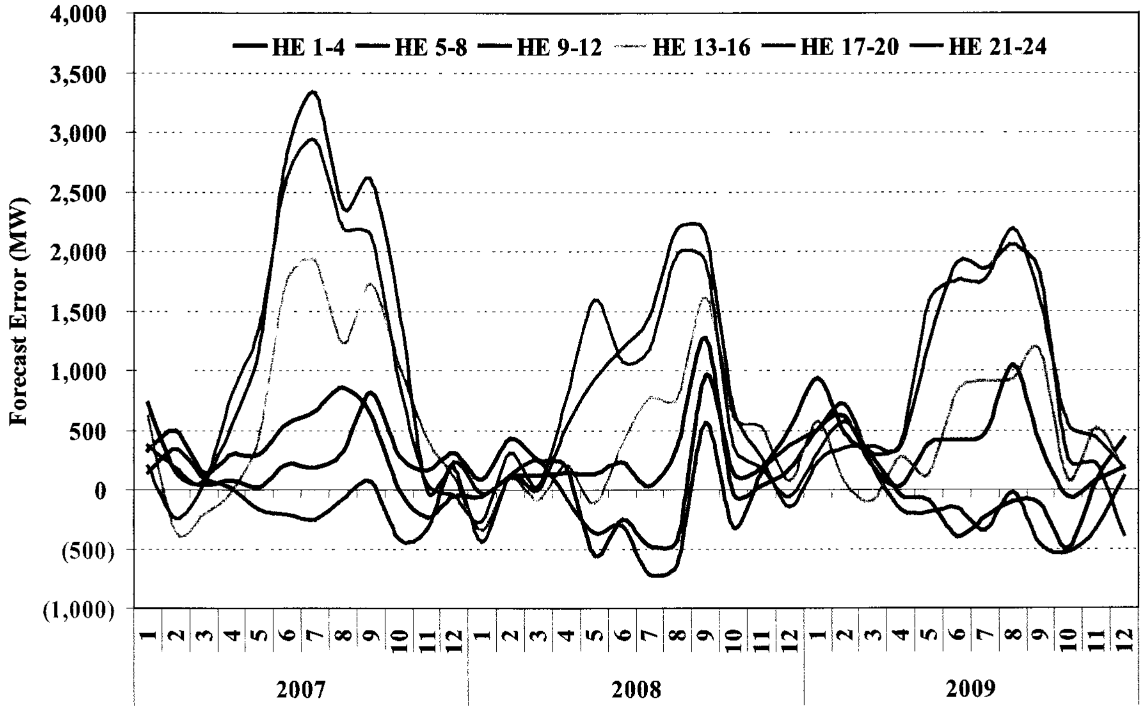
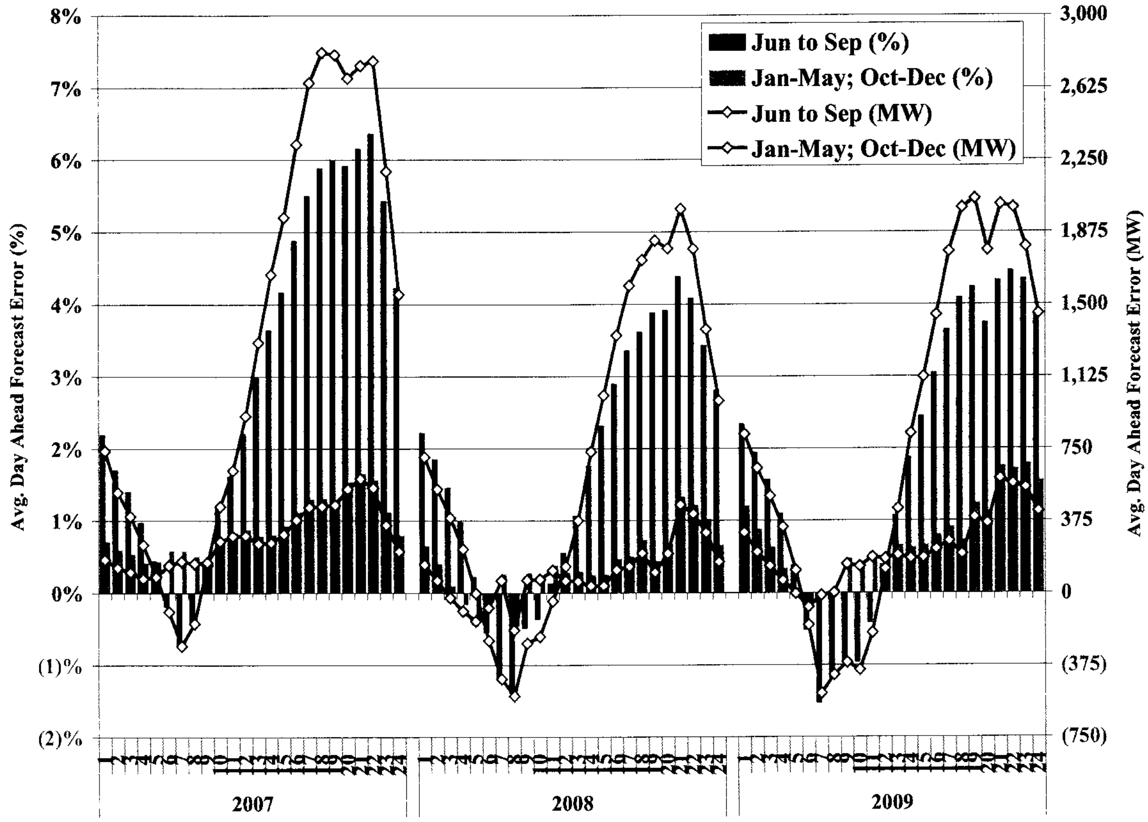


Figure 46 shows the average hourly day-ahead load forecast error for the summer months of June through September, and also for the months of January through May and October through December for 2007 through 2009. In this figure, positive values indicate a day-ahead load forecast that was greater than the actual real-time load. These data indicate a positive bias (*i.e.*, over-forecast) in the day-ahead load forecast over almost all hours in 2007 through 2009, with a particularly strong positive bias during the peak demand hours in the summer months. In terms of quantity, hour 17, for example, exhibited an average over-forecast of 300 MW for the non-summer months, and an average over-forecast of 2,000 MW for the four summer months in 2009. Figure 46 clearly shows that the positive day-ahead load forecast bias observed in 2007 and 2008 persists in 2009.

Figure 46: Average Day Ahead Hourly Load Forecast Error by Season



The existence of such a strong and persistent positive bias in the day-ahead load forecast will tend to lead to an inefficient over-commitment of resources and to the depression of real-time prices relative to a more optimal unit commitment. To the extent load uncertainty is driving the bias in the day-ahead load forecast, such uncertainty is more efficiently managed through the procurement of ancillary services such as non-spinning reserve, or through supplemental commitments of short-lead time resources at a time sufficiently prior to, but closer to real-time as uncertainty regarding real-time conditions diminishes.

In response to these observations in 2009 and prior years, the 2010 ERCOT ancillary service procurement methodologies was modified to adjust the ERCOT day-ahead load forecast to account for the historically measured net load forecast bias, and to compensate for this adjustment by increasing the quantity of non-spinning reserves procured. Although this solution is not ideal because it does not directly address the source of the forecast error bias, it is expected to have a positive effect toward reducing the average forecast error bias in 2010.

2. Implementation of PRR 776

Protocol Revision No. 776 related to the deployment and pricing of non-spinning reserve deployments was implemented in May 2009. Among other changes, the implementation of PRR 776 was expected to provide the following improvements related to non-spinning reserve deployments:

- Eliminate the previous *ex post* re-pricing provisions to provide for *ex ante* pricing during non-spinning reserve deployments, thereby providing more pricing certainty for resources and loads and significantly reducing the probability of *ex post* scarcity level prices during non-scarcity conditions;
- Allow quick start units providing non-spinning reserves to offer in the balancing energy market at a market-based price reflecting the cost and risks of starting and deploying these resources; and
- Reduce the probability of transitional shortages by providing more timely access to these reserves through the balancing energy market instead of manual operator deployments.

Generally, the implementation of PRR 776 performed as expected in 2009, providing increased efficiencies in market operations and pricing during the deployment of non-spinning reserves. As expected, the implementation of PRR 776 also significantly reduced the number of shortage intervals in 2009, as further discussed in the next subsection.

3. Dependence on High-Priced Offers by Market Participants

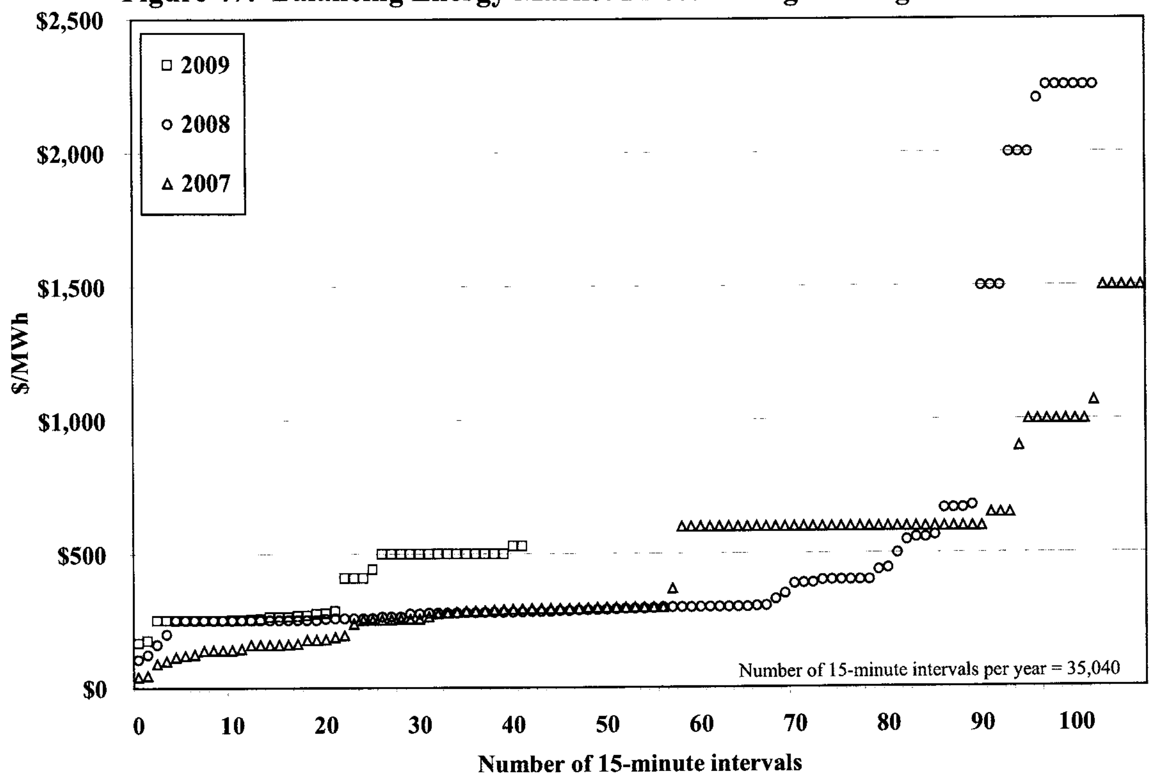
As a general principle, competitive and efficient market prices should be consistent with the marginal cost of the marginal action taken to satisfy the market's demand. In the vast majority of hours, the marginal cost of the marginal action is that associated with the dispatch of the last generator required to meet demand. It is appropriate and efficient in these hours for this generator to "set the price." However, this is not true under shortage conditions. When the system is in shortage, the demand for energy and operating reserves cannot be satisfied with the available resources, which will cause the system operator to take one or more of the following actions:

- Sacrifice a portion of the operating reserves by dispatching them for energy;
- Voluntarily curtail load through emergency demand response programs;
- Curtail exports or make emergency imports; or
- Involuntarily curtail load.

A market design that adheres to the pricing principles stated above will set prices that reflect each of these actions. When the market is in shortage, the marginal action taken by the system operator is generally to not satisfy operating reserves requirements (*i.e.*, dispatching reserves for energy). Diminished operating reserves results in diminished reliability, which has a real cost to electricity consumers. In this case, the value of the foregone reserves – which is much higher than the marginal cost of the most expensive online generator – should be reflected in energy prices to achieve efficient economic signals governing investment in generation, demand response and transmission.

Under the PUCT rules governing the energy-only market, the mechanism that allows for such pricing during shortage conditions relies upon the submission of high-priced offers by small market participants. Figure 47 shows the balancing market clearing prices during the 15-minute shortage intervals in 2007-2009.

Figure 47: Balancing Energy Market Prices during Shortage Intervals

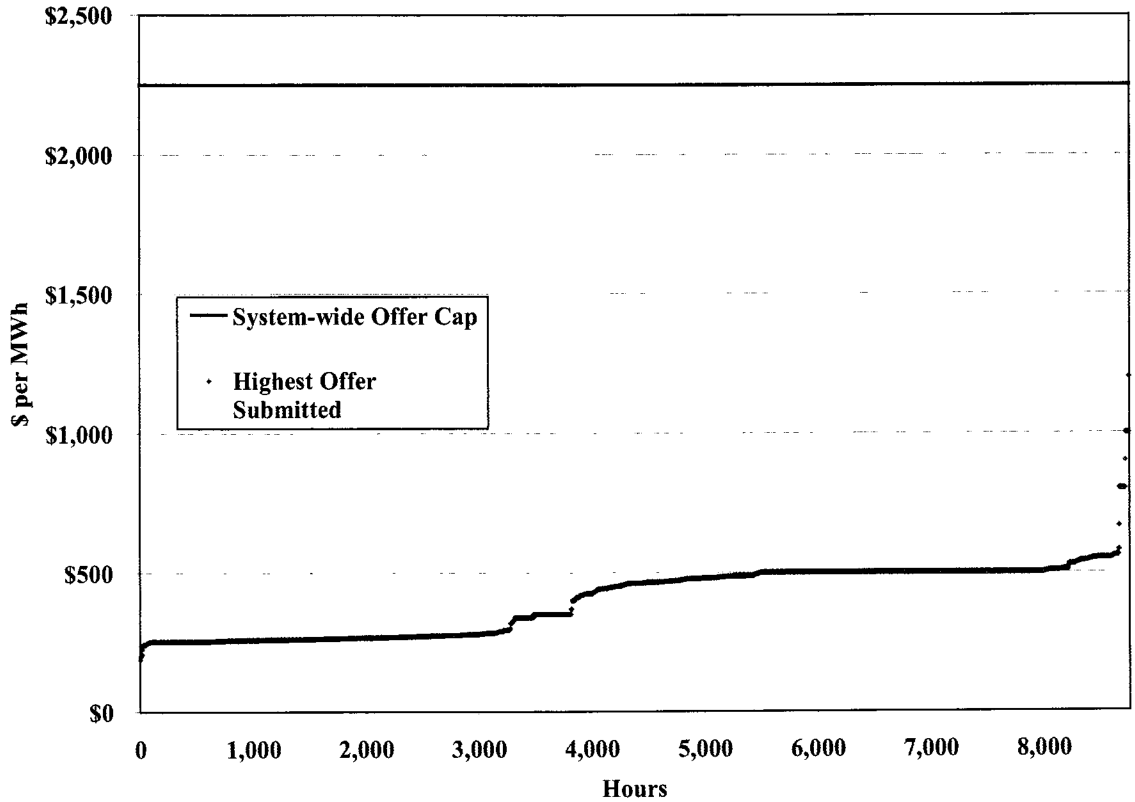


The 42 shortage intervals in 2009 are significantly fewer than the 108 and 103 shortage intervals that occurred in 2007 and 2008, respectively. This reduction can be primarily attributed to the implementation of PRR 776, which allowed more timely access to non-spinning reserves through the balancing energy market, thereby reducing the probability of transitional shortages of the core operating reserves. As shown in Figure 47, the prices during these 42 shortage intervals in 2009 ranged from $168 per MWh to $529 per MWh, with an average price of $364 per MWh and a median price of $283 per MWh.

Although each of the data points in Figure 47 represents system conditions in which the market was in shortage, the pricing outcomes are widely varied, with the majority of prices reflecting the marginal offer of the most expensive generation resource dispatched as opposed to the value of foregone operating reserves. Had an offer been submitted that established the MCPE at the system-wide offer cap in each of the 42 shortage intervals, the 2009 annual peaker net margin would have increased from $46,650 to $66,450 per MW-year, an increase of over 42 percent. The associated increase in the annual load-weighted average balancing energy price would have been less significant, increasing from $34.03 to $36.68 per MWh, an increase of 7.8 percent.

These results indicate that relying exclusively upon the submission of high-priced offers by market participants was generally not a reliable means of producing efficient scarcity prices during shortage conditions in 2007 through 2009. In fact, although the current system-wide offer cap is $2,250 per MWh (as represented by the maximum value of the y-axis in Figure 47), there were no hours in 2009 where an offer was submitted by a market participant that approached the offer cap. Figure 48 shows the highest balancing energy offer price submitted by all market participants in each hour of 2009, ranked from lowest to highest. This figure shows that there were only 33 hours (0.38 percent) with an offer that exceeded $1,000 per MWh, and the average of the highest offers submitted by any market participant in all hours in 2009 was approximately $400 per MWh.

Figure 48: Highest Hourly Balancing Energy Offer Prices



Despite the mixed and widely varied results of the SPM, private investment in generation capacity in ERCOT has continued, although such investment has been dominated by baseload (non-natural gas fueled) and wind generation. As indicated in the net revenue analyses, these investments have been largely driven by significant increases in natural gas prices in the four years prior to 2009. In contrast, private investment in peaking resources in ERCOT has been relatively thin. In an energy-only market, net revenue expectations for peaking resources are much more sensitive to the effectiveness of the shortage pricing mechanism than to the magnitude of natural gas prices.

More reliable and efficient shortage pricing could be achieved by establishing pricing rules that automatically produce scarcity level prices when operating reserve shortages exist. Such an approach would be more reliable because it would not be dependent upon the submission of high-priced offers by small market participants to be effective. It would also be more efficient during the greater than 99 percent of time in which shortage conditions do not exist because it would not be necessary for market participants to effectively withhold lower cost resources by offering relatively small quantities at prices dramatically higher than their marginal cost.

At least for the pendency of the zonal market, shortage pricing will remain dependent upon the existence of high-priced offers by market participants, and results such as those experienced in 2007 through 2009 will continue to frustrate the objectives of the energy-only market design. Further, although presenting some improvements, the nodal market design does not have a complete set of mechanisms to ensure the production of efficient prices during operating reserve shortage conditions. While important even in markets with a capacity market, efficient operating reserve shortage pricing is a particularly critical element in the ERCOT energy-only market to ensure that the long-term resource adequacy requirements are achieved.

III. TRANSMISSION AND CONGESTION

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

A. Electricity Flows between Zones

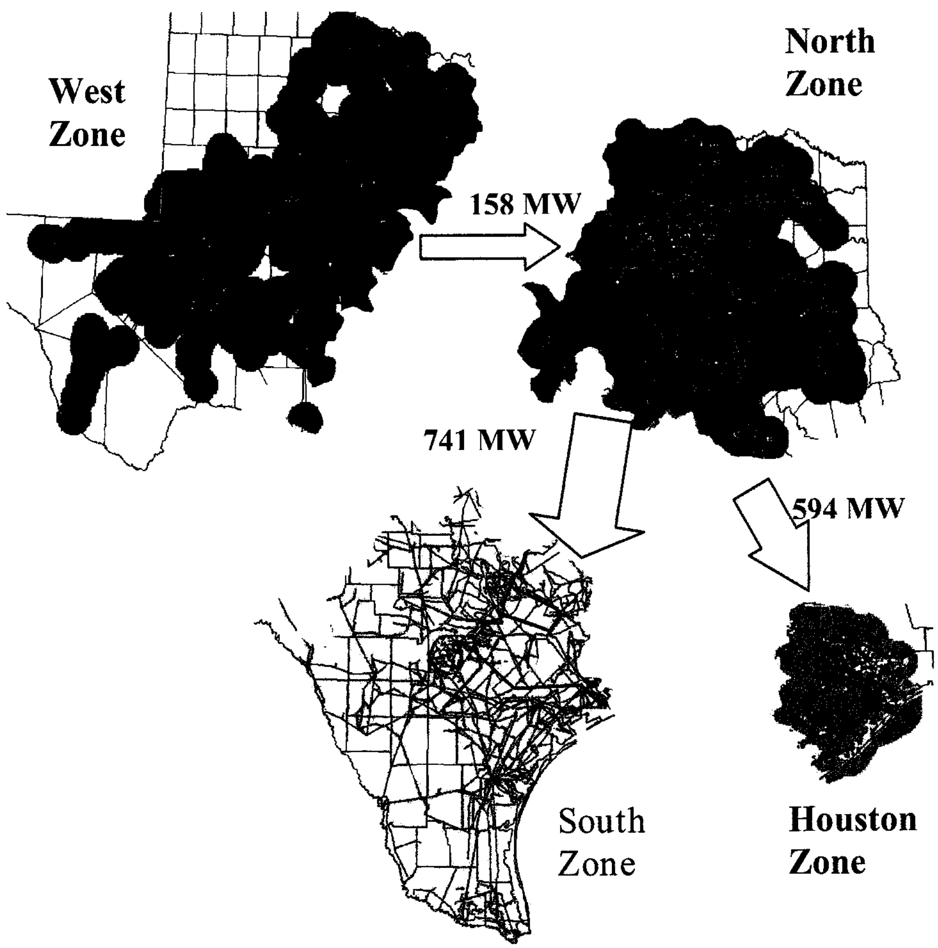
In 2009 there were four commercial pricing zones in ERCOT: (a) the North Zone, (b) the West Zone, (c) the South Zone, and (d) the Houston Zone. ERCOT operators use the Scheduling, Pricing and Dispatch ("SPD") software to economically dispatch balancing energy in each zone to serve load and manage congestion between zones. The SPD model embodies the market rules and requirements documented in the ERCOT protocols.

To manage interzonal congestion, SPD uses a simplified network model with four zone-based locations and five transmission interfaces. These five transmission interfaces, referred to as Commercially Significant Constraints ("CSCs"), are simplified representations of groups of transmission elements. ERCOT operators use planning studies and real-time information to set limits for each CSC that are intended to utilize the total transfer capability of the CSC. In this subsection of the report, we describe the SPD model's simplified representations of flows between zones and analyze actual flows in 2009.

The SPD model uses zonal approximations to represent complex interactions between generators, loads, and transmission elements. Because the model flows are based on zonal approximations, the estimated flows can depart significantly from real-time physical flows.

Estimated flows that diverge significantly from actual flows are an indication of inaccurate congestion modeling leading to inefficient energy prices and other market costs. This subsection analyzes the impact of SPD transmission flows and constraints on market outcomes.

Figure 49: Average SPD-Modeled Flows on Commercially Significant Constraints During All Intervals in 2009



Note: In the figure above, CSC flows are averaged taking the direction into account. So one arrow shows the average flow for the West-to-North CSC was 158 MW, which is equivalent to saying that the average for the North-to-West CSC was *negative* 158 MW.

Figure 49 shows the four ERCOT geographic zones as well as the five CSCs that interconnect the zones: (a) the West to North interface, (b) the South to North interface, (c) the North to South interface, (d) the North to Houston interface, and (e) the North to West interface. A single arrow is shown for the modeled flows of both the North to West and West to North CSCs and the South to North and North to South CSCs. Based on average SPD modeled flows, the North Zone exports a significant amount of power.

The most important simplifying assumption underlying the zonal model is that all generators and loads in a zone have the same effect on the flows over the CSC, or the same shift factor in relation to the CSC.[23] In reality, the generators and loads within each zone can have widely differing effects on the flows over a CSC. To illustrate this, we compared the flows calculated by using actual generation and zonal average shift factors to the average actual flows that occurred over each CSC. The flows over the North to West and South to North CSCs are not shown separately in the table below since they are equal and opposite the flows for the West to North CSC and North to South CSCs, respectively.

Table 2: Average Calculated Flows on Commercially Significant Constraints Zonal-Average vs. Nodal Shift Factors

CSC 2009	Flows Modeled by SPD (1)	Flows Calculated Using Actual Generation (2)	*Difference* *= (2) - (1)*	Actual Flows Using Nodal Shift Factors (3)	*Difference* *= (3) - (2)*
West-North	158	156	*-2*	207	*51*
North - South	741	688	*-53*	406	*-282*
North-Houston	594	541	*-53*	747	*206*

The first column in Table 2 shows the average flows over each CSC calculated by SPD. The second column shows the average flows over each CSC calculated by using zonal-average shift factors 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 shift factors, 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 calculated flows on each CSC that vary only slightly from the flows modeled by SPD.

23 For a generator, a shift factor indicates the portion of the incremental output of a unit that will flow over a particular transmission facility. For example, a shift factor of 0.5 would indicate that half of any incremental increase in output from a generator would flow over the interface. A negative shift factor would indicate a decrease in flow on an interface resulting from an increase in generation.

The fourth column in Table 2 reports the actual average flows over each CSC by using nodal shift factors applied to actual real-time generation and load. The difference in flows between columns (3) and (2) is attributable to using zonal average shift factors versus nodal shift factors for generation and load in each zone. These differences in flows are shown in the fifth column (in italics).

These results show that the heterogeneous effects of generators and load in a zone on the CSC flows can cause the actual flows to differ substantially from the SPD-calculated flows. Table 2 shows that by using nodal (actual) shift factors reduced the calculated flows on the North to South interface by 282 MW and increased the calculated flows on the North to Houston CSC by 206 MW.

The use of simplified generation-weighted shift factors prevents the SPD model from efficiently resolving and assigning the costs of interzonal congestion. In the long run, the use of generation-weighted shift factors for loads systematically biases prices, so that buyers in some zones pay too much, and others pay too little. Further, the use of average zonal shift factors creates significant operational challenges for ERCOT in the real-time management of zonal congestion because the response to zonal dispatch instructions can often affect the actual flow on a CSC in a manner that is significantly different than that calculated by the simplified assumptions in the SPD model. In turn, ERCOT will tend to operate the system more conservatively to account for the operational uncertainties introduced by the simplified assumptions in the SPD model, the effect of which is discussed in more detail later in this section.

To provide additional understanding of the electricity flows between zones prior to discussing the details of interzonal congestion in the next subsection, Figure 50 shows the actual average imports of power for each zone in 2009. In this figure, positive values represent imports, and negative values indicate exports.[24]

24 The Northeast Zone existed in 2005 and 2006, but was merged into a single North Zone in 2007 and 2008. The Northeast zone is included in the North zone for 2005 and 2006 in Figure 50.

Figure 50: Actual Zonal Net Imports

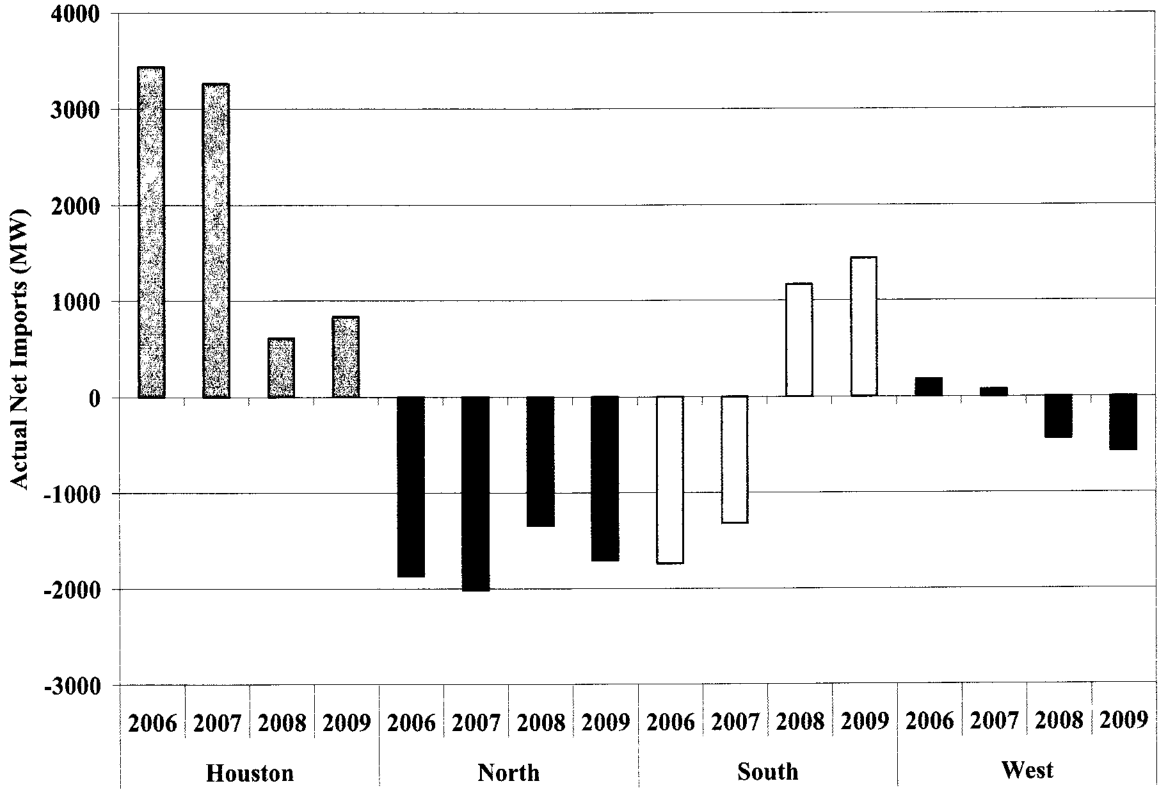


Figure 50 shows that the Houston Zone is a net importer of power, while the North Zone is a net exporter. The reduction in the Houston Zone imports in 2008 and 2009 and corresponding change in the South Zone from a net exporter to a net importer can be attributed to the movement of the 2,700 MW South Texas Nuclear Project from the South Zone to the Houston Zone in 2008. The West Zone transitioned from a net importer in 2006 and 2006 to a net exporter in 2008 and 2009. This reflects the significant increases in the installed capacity of wind resources in the West Zone that occurred over this time period.

B. Interzonal Congestion

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

Figure 51 shows the average SPD-calculated flows between the four ERCOT zones during constrained periods for the 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 51: Average SPD-Modeled Flows on Commercially Significant Constraints During Transmission Constrained Intervals in 2009

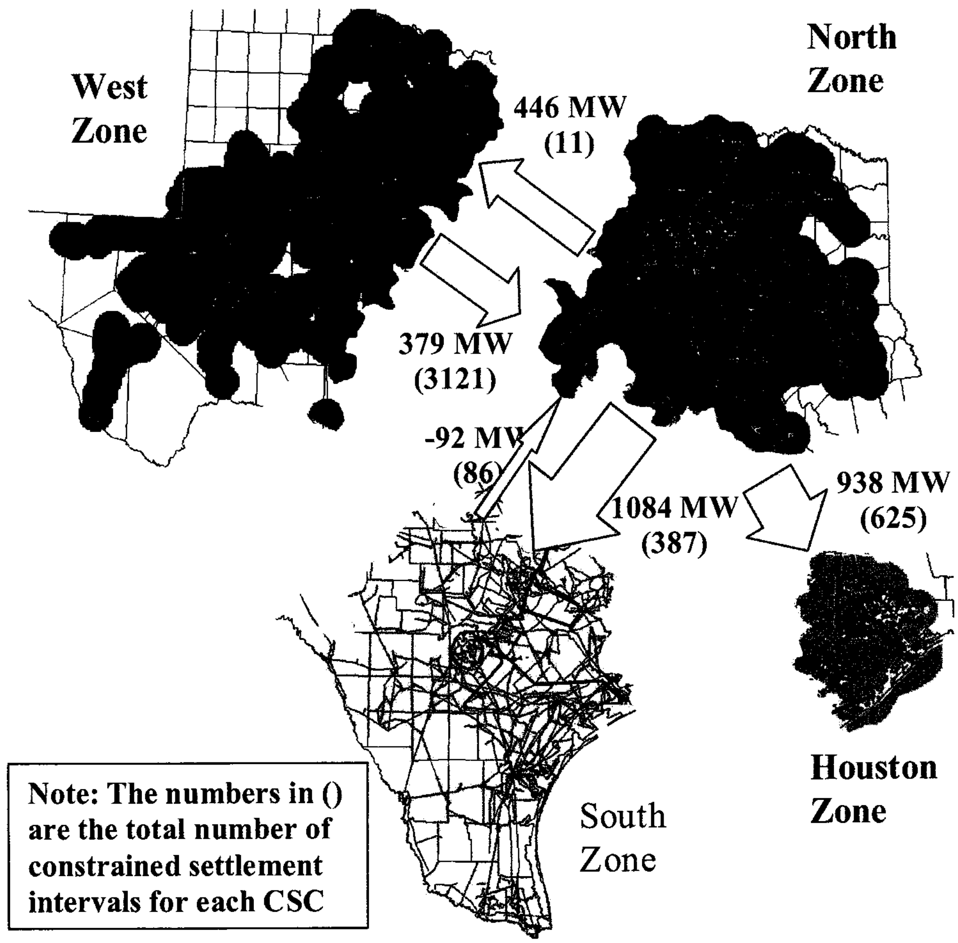


Figure 51 shows that inter-zonal congestion was most frequent in 2009 on the West to North and the North to Houston CSCs, followed by the North to South CSC. The West to North CSC exhibited SPD-calculated flows averaging 379 MW during 3,121 constrained 15-minute intervals (9 percent of the totals intervals in the year). The North to South CSC exhibited SPD-calculated flows averaging 1,084 MW during 387 constrained intervals (1 percent of the total intervals), and the SPD-calculated average flow for the North to Houston CSC was 938 MW during 625 constrained intervals (2 percent of the total intervals).

Table 3: Average Calculated Flows on Commercially Significant Constraints during Transmission Constrained Intervals Zonal-Average vs. Nodal Shift Factors

CSC 2009	Flows Modeled by SPD (1)	Flows Calculated Using Actual Generation (2)	*Difference* = (2) - (1)	Actual Flows Using Nodal Shift Factors (3)	*Difference* = (3) - (2)
North - South	1084	1058	*-26*	906	*-152*
North - Houston	938	889	*-49*	1171	*282*
South - North	-92	-198	*-106*	209	*406*
West - North	379	435	*56*	383	*-52*
North - West	446	632	*186*	623	*-9*

Table 3 shows data similar to that presented in Table 2, except that the data in Table 3 is limited for each CSC to only those intervals in which the transmission constraint was binding. Table 3 shows that the average SPD-modeled flows for the West to North and North to West CSCs were relatively close to actual flows, whereas the average actual flows for the North to South, South to North and North to Houston CSCs varied significantly from the average flows modeled by SPD.

The following subsections provide a more detailed assessment of the actual occurrences of congestion for each CSC in 2009, with the exception of the North to West CSC that was binding in only eleven 15-minute intervals in 2009.

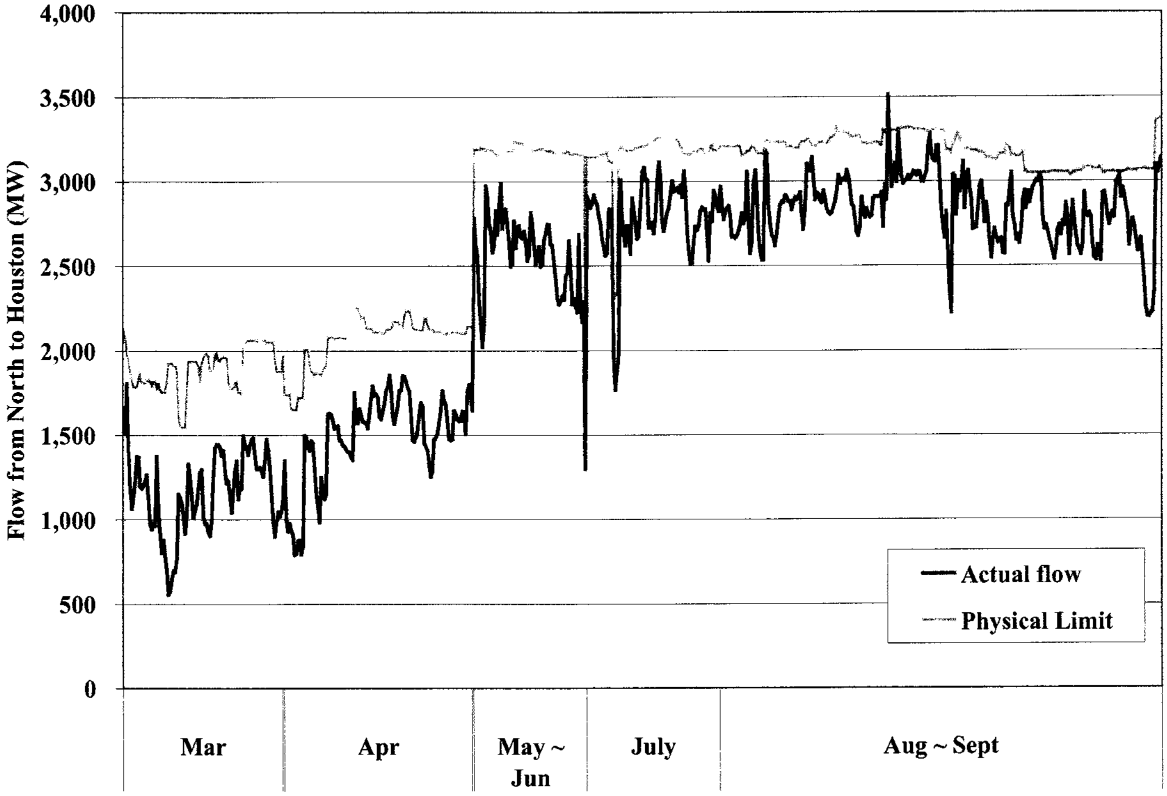
1. Congestion on the North to Houston CSC

The North to Houston CSC was binding in 625 15-minute intervals with an annual average shadow price of $2.01 per MW. These values represent a significant reduction in both the frequency and magnitude of congestion compared to 2008 when the North to Houston CSC was binding in 1,447 intervals with an annual average shadow price of $20.

The decreased congestion on the North to Houston CSC in 2009 is primarily attributable to the implementation of PRR 764 in June 2008 that revised the definition of valid zonal transmission constraints and improved the efficiency of transmission congestion management within the context of the zonal market model.[25] Figure 52 shows the actual flows versus the physical limit for the North to Houston CSC in 2009 during intervals when the CSC was binding.

[25] *See* 2008 ERCOT SOM Report at 81-87.

Figure 52: Actual Flows versus Physical Limits during Congestion Intervals North to Houston



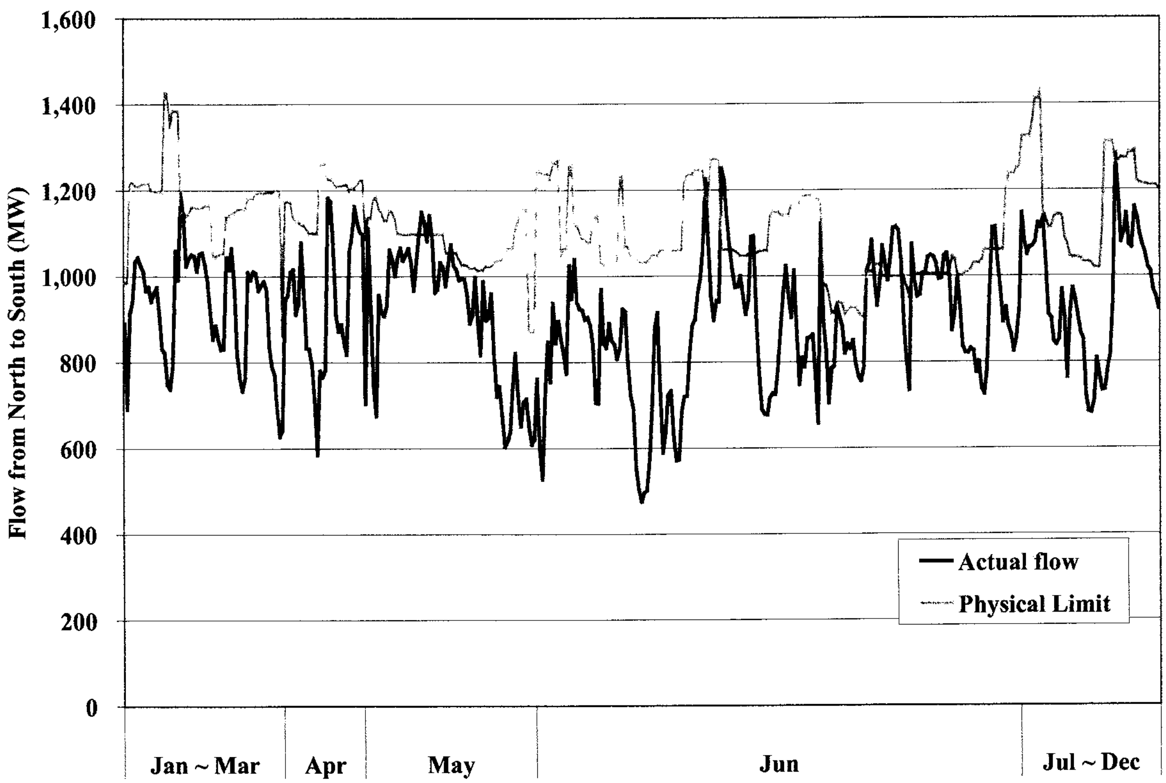
The average physical limit and actual flow for the North to Houston CSC during constrained intervals were 2,772 and 2,290 MW, respectively. Of the 625 intervals that the North to Houston CSC was binding, the actual flow was less than the physical limit in 623 intervals and greater than the physical limit in two intervals. In the 623 intervals where the actual flow was less than the physical limit and the constraint was binding, the average actual flow was 484 MW below the physical limit. In contrast, in the two intervals where the actual flow was greater than the physical limit, the average actual flow was 104 MW above the physical limit.

2. Congestion on the North to South CSC

In 2009 the North to South CSC was binding in 387 15-minute intervals with an annual average shadow price of $8.39 per MW. Like the North to Houston CSC, these values represent a significant reduction in both the frequency and magnitude of congestion compared to 2008 when the North to South CSC was binding in 2,531 intervals with an annual average shadow price of $22.

As was the case for the North to Houston CSC, the reduction in congestion on the North to South CSC in 2009 can be attributed to the implementation of PRR 764. Figure 53 shows the actual flows versus the physical limit for the North to South CSC in 2009 during intervals when the CSC was binding.

Figure 53: Actual Flows versus Physical Limits during Congestion Intervals North to South



The average physical limit and actual flow for the North to South CSC during constrained intervals were 1,117 and 906 MW, respectively, in 2009. Of the 387 intervals that the North to South CSC was binding, the actual flow was less than the physical limit in 353 intervals and greater than the physical limit in 34 intervals. In the 353 intervals where the actual flow was less than the physical limit and the constraint was binding, the average actual flow was 237 MW below the physical limit. In contrast, in the 34 intervals where the actual flow was greater than the physical limit, the average actual flow was 53 MW above the physical limit.

Figure 53 also shows that a significant percentage of the congestion on the North to South CSC occurred during June 2009. During this timeframe, the ERCOT market experienced very high temperatures and associated increases in load levels, as well as a number of outages at baseload generating facilities, particularly in the South Zone. This combination of events led to an increase in the frequency of congestion on the North to South CSC as well as local congestion related to import limitations into the San Antonio area from the north. In the zonal model, the most effective resolution to North to South congestion is to increase generation in the South Zone. However, effective zonal congestion management on the North to South CSC was affected by the local congestion in the San Antonio area, which is most effectively resolved by increasing generation in and South of San Antonio, and decreasing generation north of San Antonio. Because most of the generation resources located north of San Antonio required to decrease output to manage the local congestion in the San Antonio area are also in the South Zone that was broadly required to increase output to manage the zonal North to South congestion, competing reliability objectives were present that complicated the simultaneous resolution of both the North to South zonal congestion and the intrazonal San Antonio import-related congestion. Faced with these competing reliability objectives and the inability to resolve both reliability issues within the context of the zonal model and its bifurcated process of zonal and local transmission congestion management, ERCOT implemented a temporary transmission switching solution in late June that effectively increased the transfer capability on the North to South CSC, thereby resolving these competing reliability objectives under the atypical load and generator outage conditions experienced at that time.

3. Congestion on the West to North CSC

In 2009 the West to North CSC was binding in 3,121 15-minute intervals. This was more frequent than any other CSC in 2009 and, with the exception of the same CSC in 2008 that was binding for 5,320 intervals, more frequent than any other CSC since the inception of single control area operations in 2001. The primary reason for the high frequency of congestion on the West to North CSC in 2008 and 2009 is the significant increase in installed wind generation relative to the load in the West Zone and limited transmission export capability to the broader market.

Average load in the West Zone was 2,023 MW in 2009, with a minimum of 1,588 MW and a maximum of 2,744 MW. The average profile of West Zone wind production is negatively correlated with the load profile, with the highest wind production occurring primarily during the spring, fall and winter months, and predominately during off-peak hours. Figure 54 shows the average West Zone wind production for each month in 2009, with the average production in each month shown separately in four hour blocks.[26]

Figure 54: Average West Zone Wind Production

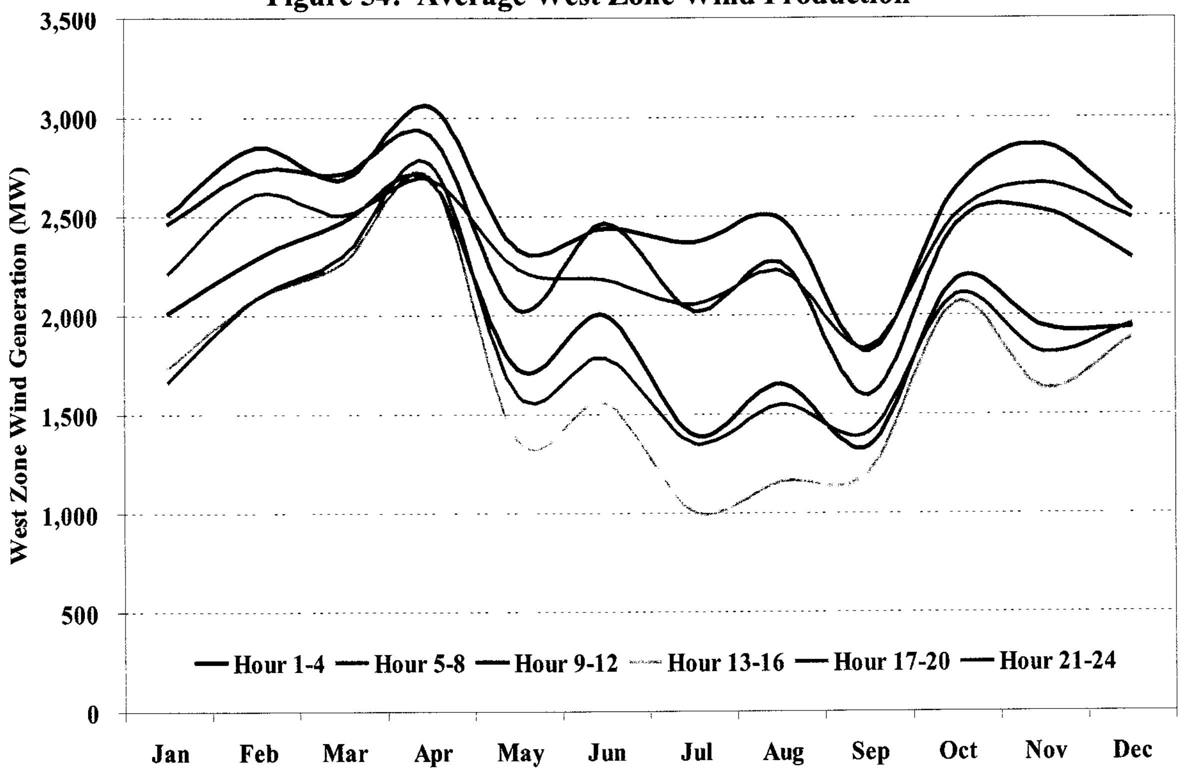
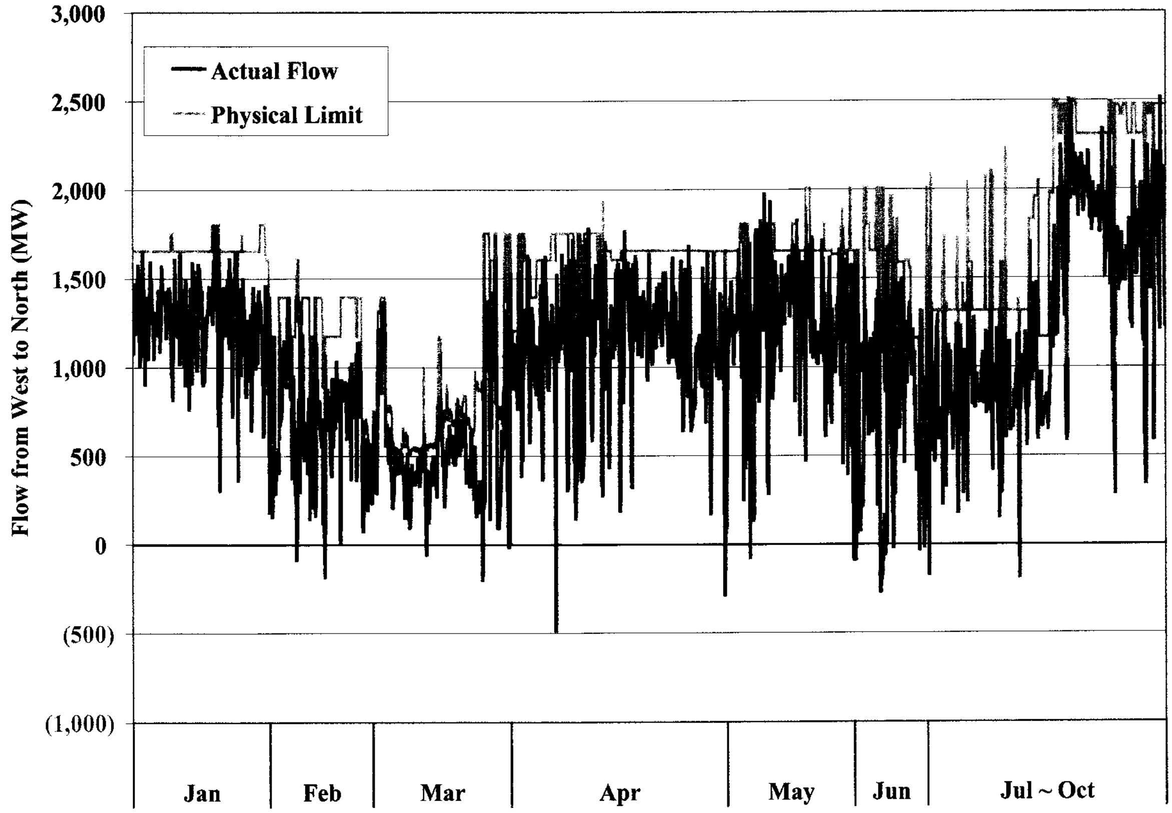


Figure 55 shows the actual flows and the physical limit for the West to North CSC in 2009 for intervals in which the CSC was binding. The average physical limit and actual flow for the West to North CSC during constrained intervals were 1,528 and 1,046 MW, respectively, in 2009. Of the 3,121 intervals that the West to North CSC was binding, the actual flow was less than the physical limit in 3,096 intervals and greater than the physical limit in 25 intervals. In the 3,096 intervals where the actual flow was less than the physical limit and the constraint was binding, the average actual flow was 487 MW below the physical limit. In contrast, in the 25 intervals where the actual flow was greater than the physical limit, the average actual flow was 42 MW above the physical limit.

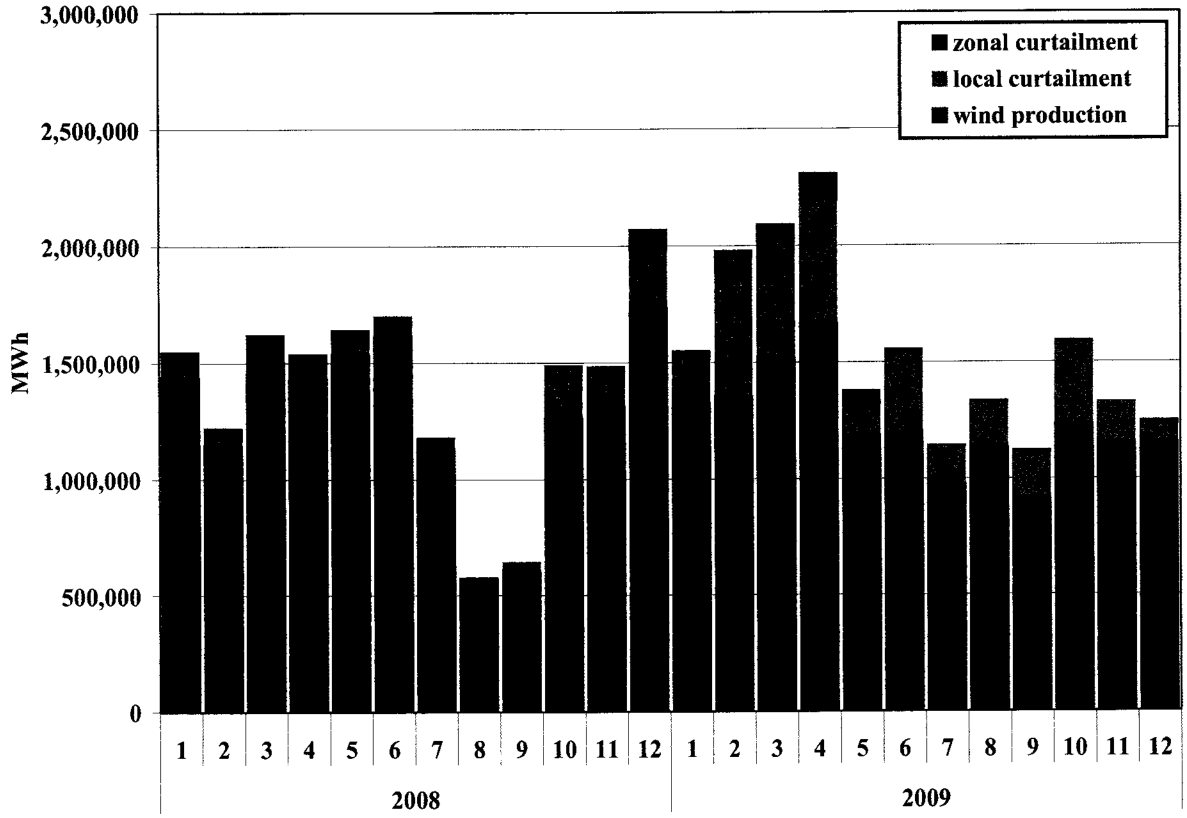
26 Figure 54 shows actual wind production, which was affected by curtailments at the higher production levels in 2008. Thus, the higher levels of actual wind production in Figure 54 are lower than the production levels that would have materialized absent transmission constraints.

Figure 55: Actual Flows versus Physical Limits during Congestion Intervals West to North



Although the frequency of zonal transmission congestion on the West to North CSC was very high in 2009, it was lower than in 2008. However, zonal congestion data do not provide a complete view of the congestion situation in the West Zone. Figure 56 shows the wind production and local and zonal curtailment quantities for the West Zone for each month of 2008 and 2009. This figure reveals that, while the quantity of zonal curtailments for wind resources in the West Zone was reduced from 604,000 MWh in 2008 to 442,000 MWh in 2009, the quantity of local curtailments increased significantly, rising from 812,000 MWh in 2008 to over 3,400,000 MWh in 2009.

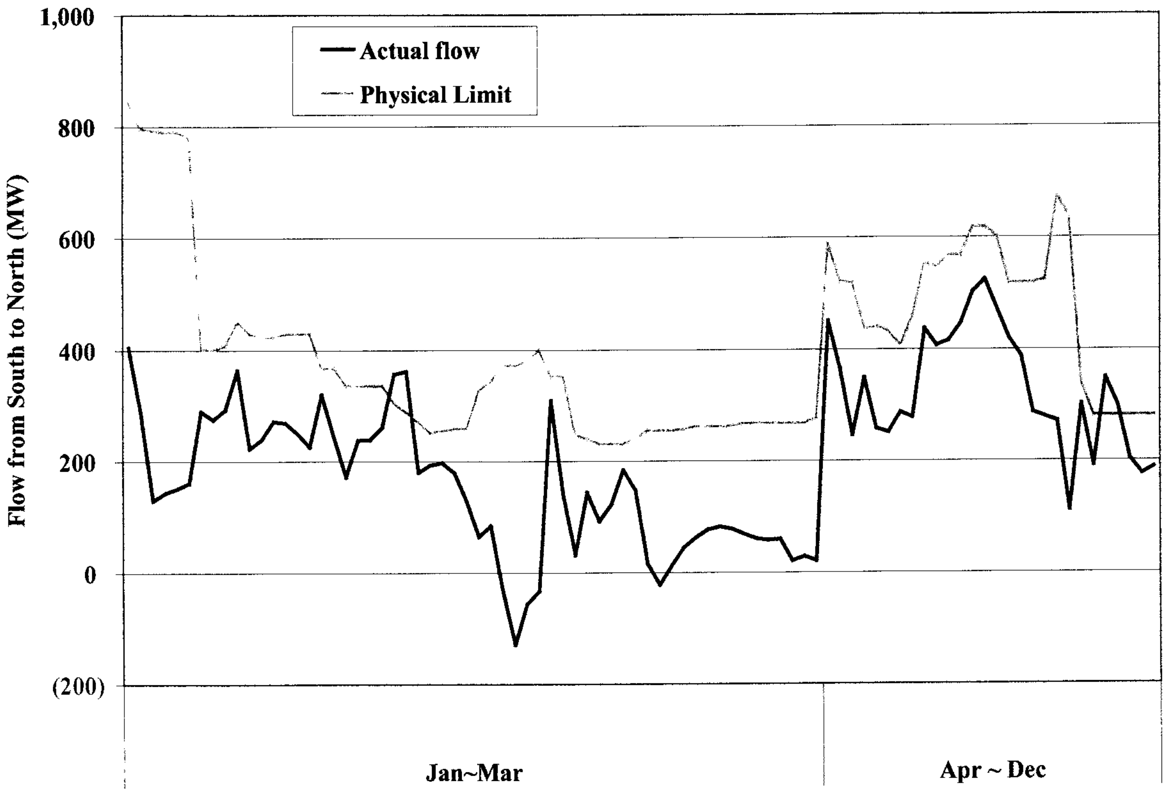
Figure 56: West Zone Wind Production and Curtailment



4. Congestion on the South to North CSC

The South to North CSC was binding in 86 15-minute intervals in 2009. Figure 57 shows the actual flows and the physical limit for the South to North CSC in 2009 for intervals in which the CSC was binding. The average physical limit and actual flow for the South to North CSC during constrained intervals were 402 and 209 MW, respectively, in 2009. Of the 86 intervals that the South to North CSC was binding, the actual flow was less than the physical limit in 82 intervals and greater than the physical limit in four intervals. In the 82 intervals where the actual flow was less than the physical limit and the constraint was binding, the average actual flow was 205 MW below the physical limit. In contrast, in the four intervals where the actual flow was greater than the physical limit, the average actual flow was 53 MW above the physical limit.

Figure 57: Actual Flows versus Physical Limits during Congestion Intervals South to North



5. Zonal Congestion Management Challenges

As discussed in the first part of this section, differences that exist between the commercial SPD model representation and the physical reality create operational challenges for ERCOT to efficiently manage zonal transmission congestion while also maintaining reliable operations. Table 4 shows the average physical limit, actual flow and the difference between the average physical limit and the actual flow for each CSC during binding intervals in 2009.

Table 4: CSC Average Physical Limits vs. Actual Flows during Constrained Intervals

CSC 2009	Average Physical Limit (MW)	Average Actual Flow (MW)	Avg. Physical Limit - Avg. Actual Flow (MW)
North to South	1117	906	211
North to Houston	2772	2290	483
South to North	401	208	193
West to North	1528	1046	483
North to West	780	623	157

Table 4 shows that, for all CSCs in 2009, the average actual flow was considerably less than the average physical limit. For all CSCs combined, the average actual flow was 23 percent less than the average physical limit. To maximize the economic use of the scarce transmission capacity, the ideal outcome would be for the actual flows to reach the physical limits, but not to exceed such limits to maintain reliable operations. However, primarily for the reasons discussed in the first part of this section, achieving such ideal outcomes is practically impossible in the context of the zonal market model. Further, as discussed in relation to the North to South CSC, the bifurcated process of resolving zonal and local congestion can at times lead to reliability conflicts that are difficult to resolve within the relatively inflexible framework of the zonal market design.

The nodal market will provide many improvements, including unit-specific offers and shift factors, simultaneous resolution of all transmission congestion, actual output instead of schedule-based dispatch, and 5-minute instead of 15-minute dispatch, among others. These changes should help to increase the economic and reliable utilization of scarce transmission resources well beyond that experienced in the zonal market, and in so doing, also dispatch the most efficient resources available to reliably serve demand.

C. Congestion Rights Market

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 constraint occurs market participants must compete to use the available transfer capability between zones. To allocate this capability efficiently, ERCOT establishes clearing prices for energy in each zone that will vary in the presence of congestion and charges the transactions between the zones the interzonal congestion price.

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

To analyze congestion rights in ERCOT, we first review the TCRs and PCRs that were auctioned or allocated for each CSC in 2009. Figure 58 shows the average number of TCRs and PCRs awarded for each of the CSCs in 2009 compared to the average SPD-modeled flows during the constrained intervals.

Figure 58: Transmission Rights vs. Real-Time SPD-Calculated Flows Constrained Intervals - 2009

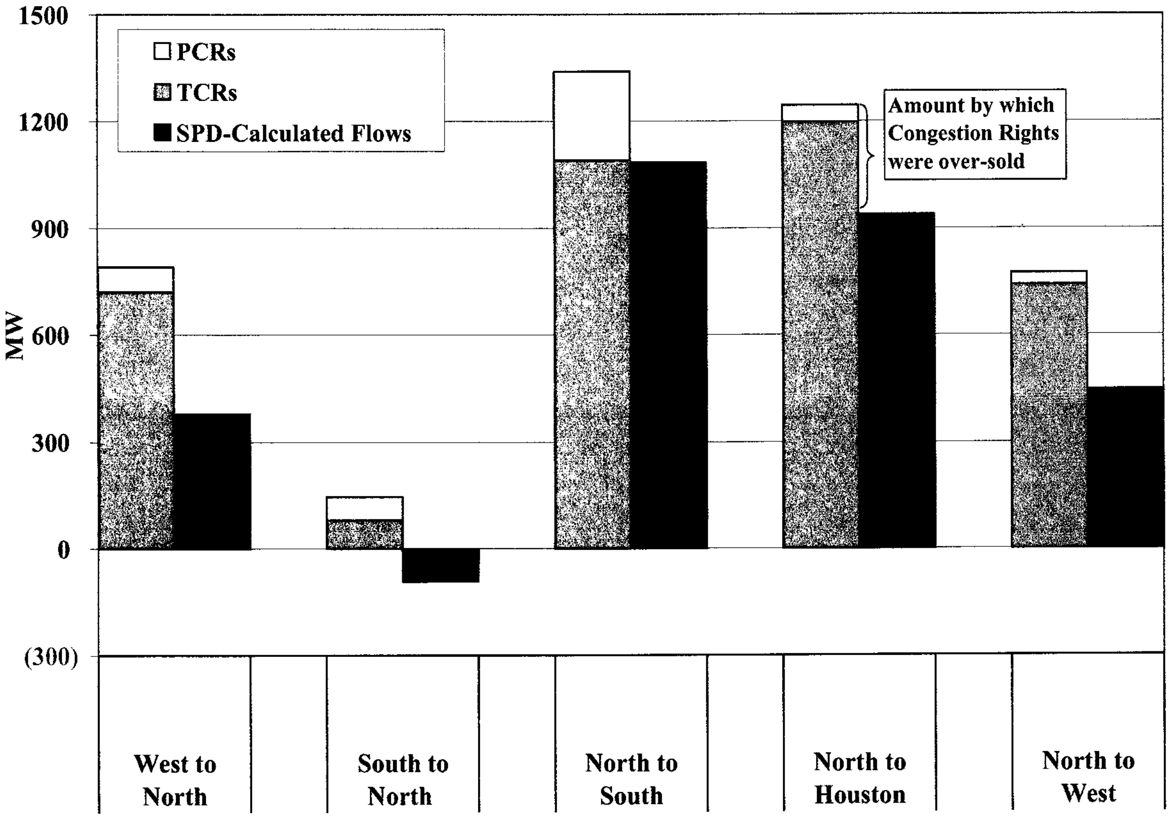


Figure 58 shows that total congestion rights (the sum of PCRs and TCRs) on all the interfaces exceeded the average real-time SPD-calculated flows during constrained intervals. These results indicate that the congestion rights were oversold in relation to the SPD-calculated limits. For example, congestion rights for the North to Houston CSC were oversold by an average of 328 MW. The average amount of TCRs awarded each month in 2009 is higher than in 2008.

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 flowing over the interface. When the SPD-calculated flows are consistent with the quantity of congestion rights sold over the interface, the congestion revenues will be sufficient to satisfy payments to the holders of the congestion rights. Alternatively, when the quantity of congestion rights exceeds the SPD-calculated flow over an interface, congestion revenues from the balancing energy market will not be sufficient to meet the financial obligations to congestion rights holders.

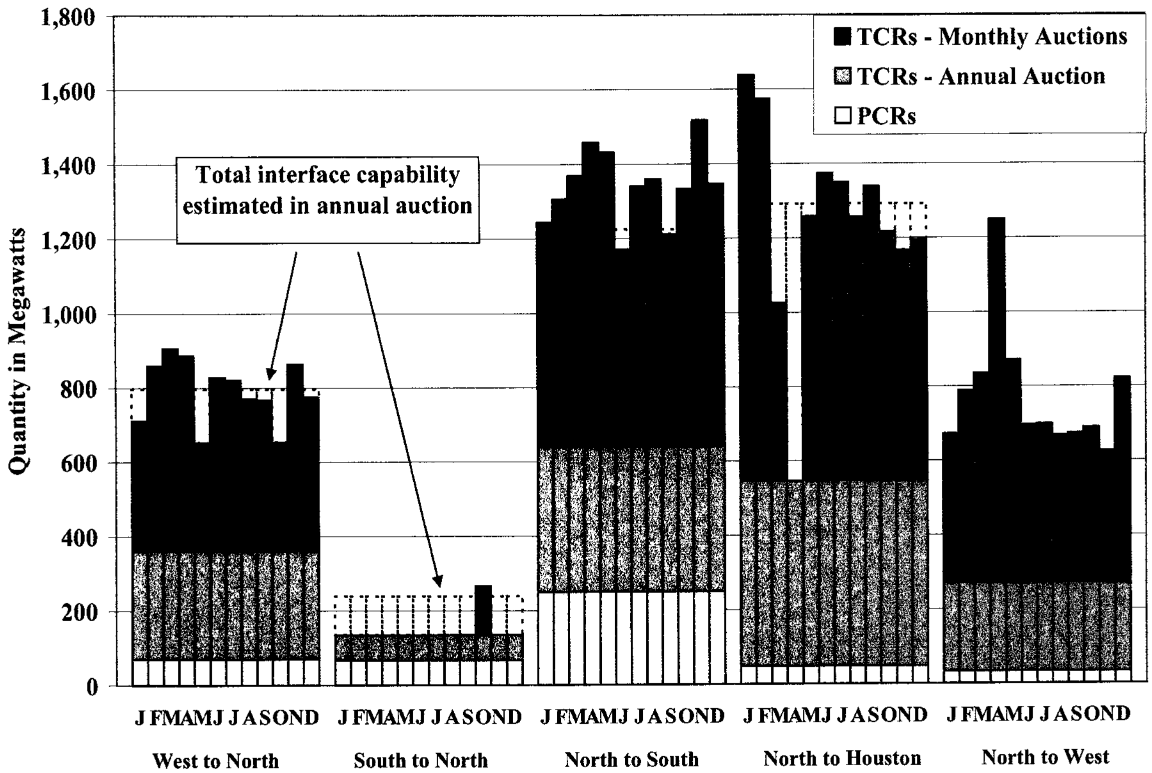
As an example, suppose the SPD-calculated flow limit is 300 MW for a particular CSC during a constrained interval and that holders of congestion rights own a total of 800 MW over the CSC. ERCOT will receive congestion rents from the balancing energy market to cover precisely 300 MW of the 800 MW worth 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 provide a better understanding of these relationships, we next review ERCOT's process to establish the quantity of congestion rights allocated or sold to participants. ERCOT performs studies to determine the capability of each interface under peak summer conditions. This summer planning study is the basis for offering 40 percent of the available TCRs for sale in the annual auction. These rights are auctioned during December for the coming year. Additional TCRs are offered for sale based on monthly updates of the summer study. Because 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 monthly studies used to designate the TCRs do not always accurately reflect real-time transmission conditions for two main reasons. First, transmission and generation outages can occur unexpectedly and can significantly reduce the transfer capability of a CSC. Even planned transmission outages may not be known to ERCOT when the summer studies are conducted. Second, conditions may arise causing the actual physical flow to be significantly different from the SPD modeled flow. As discussed above, ERCOT operators may need to respond by lowering the SPD-modeled flow limits to manage the actual physical flow. Accordingly, it is likely that the quantity of congestion rights awarded will be larger than available transmission capability in SPD.

To examine how these processes have together determined the total quantity of rights sold over each interface, Figure 59 shows the quantity of each category of congestion rights for each month during 2009. The quantities of PCRs and annual TCRs are constant across all months and were determined before the beginning of 2009, while monthly TCR quantities can be adjusted monthly.

Figure 59: Quantity of Congestion Rights Sold by Type

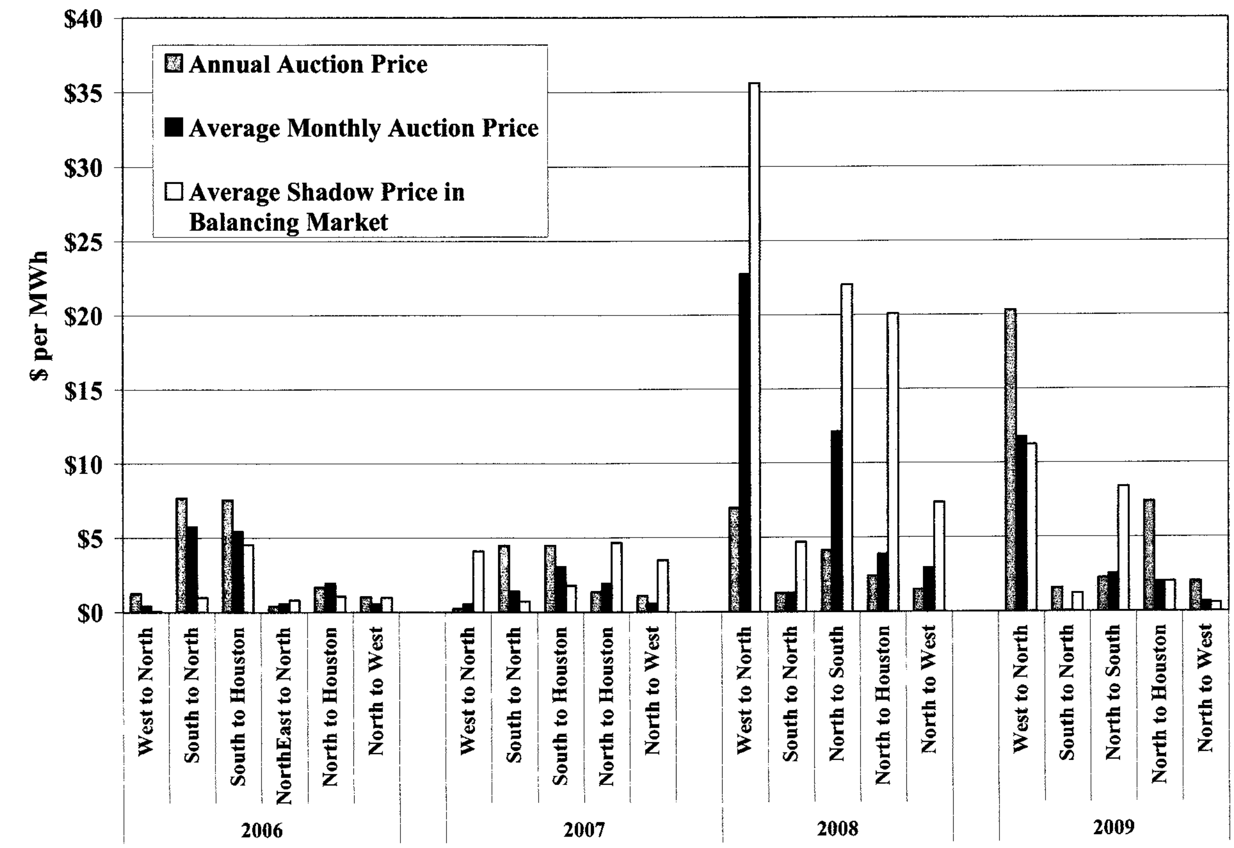


When the monthly planning studies indicate changes from the summer study, revisions are often made to the estimated transmission capability. Therefore, the auctioned congestion rights may increase or decrease relative to the amount estimated in the summer study. The shadow boxes in the figure represent the capability estimated in the summer study that is not ultimately sold in the monthly auction. When there is no shadow box in Figure 59, the total quantity of PCRs and TCRs sold in the annual and monthly auctions equaled or exceeded the summer estimate and therefore no excess capability is shown.

The South to North and North to Houston interfaces experienced the largest fluctuations in the estimates of transmission capacity between the annual auction and the monthly auctions. In fact, for several months South to North TCRs were not even offered for sale 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 and 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 60 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.

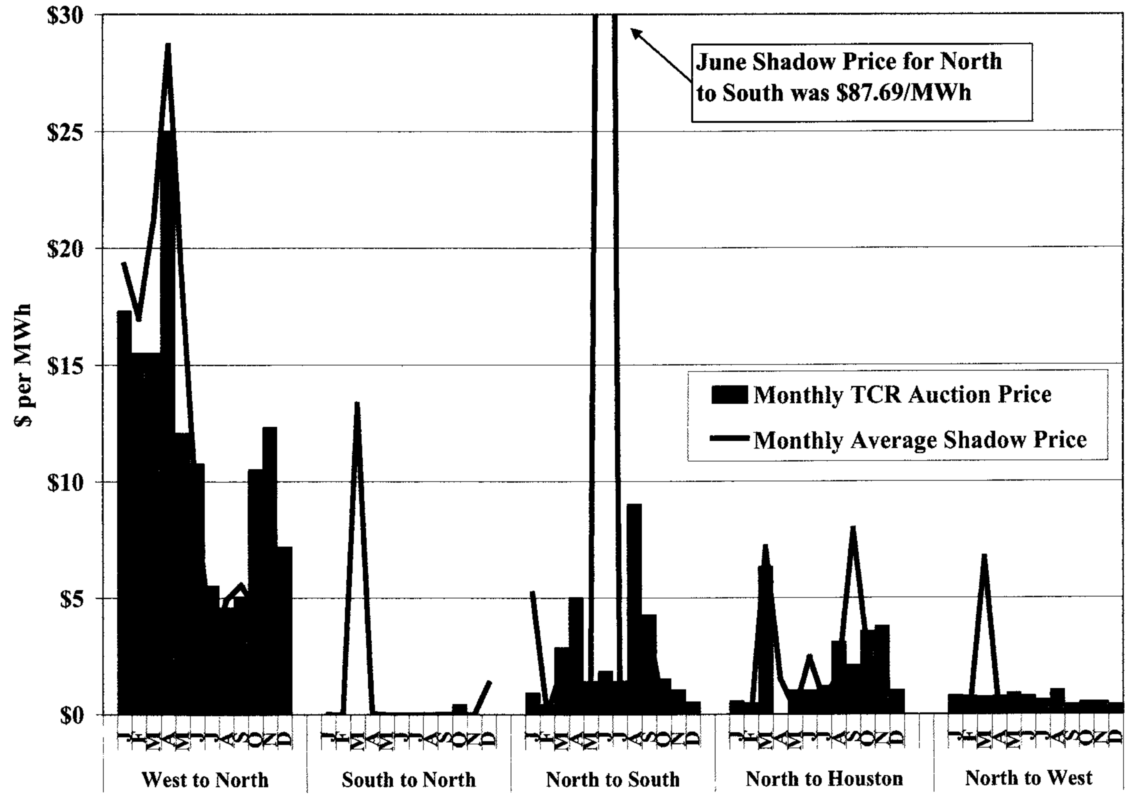
Figure 60: TCR Auction Prices versus Balancing Market Congestion Prices



This figure shows that the TCR annual auction prices were higher than the value of congestion in real-time for the West to North, North to Houston and North to West CSCs in 2009. In contrast, the annual auction price was significantly lower than the value of congestion in real-time for the North to South CSC in 2009. For the West to North, North to Houston and North West CSCs, the average monthly auction prices were more consistent with the value of congestion in real-time in 2009, indicating a more accurate forecast by the participants at the monthly auction than previous years for these CSCs. The North to South monthly auction price was significantly lower than the actual value of congestion in the real-time in 2009. This outcome is primarily due to the significant North to South congestion experienced in June 2009 that was influenced to a large degree by a number of baseload unit outages that were not foreseeable, as discussed previously in this Section.

Figure 61 compares monthly TCR auction prices with monthly average real-time CSC shadow prices from SPD for 2009. The TCR auction prices are expressed in dollars per MWh.

Figure 61: Monthly TCR Auction Price and Average Congestion Value



With the exception of the North to South CSC in June 2009 that diverged for the reasons previously discussed, the monthly TCR auction prices and the real-time shadow prices indicates that market participants improved their ability to predict and value the real-time cost of zonal congestion in 2009 compared to prior years.

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 per MWh and the price in the North Zone is $55 per MWh. The customers in the North Zone will pay $33,000 (600 MWh * $55 per MWh) while suppliers in the West Zone will receive $24,000 (600 MWh * $40 per MWh). The net result is that ERCOT collects $9,000 in congestion rent ($33,000 – $24,000) and uses it to fund payments to holders of TCRs.[27] 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 62 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.

27 This explanation is simplified for the purposes of illustration. Congestion rents are also affected by differences between calculated flows on CSCs from interzonal schedules using zonal average shift factors and actual flows on CSCs in real-time. As discussed in this Section, these differences can be significant.

Figure 62: TCR Auction Revenues, Credit Payments, and Congestion Rent

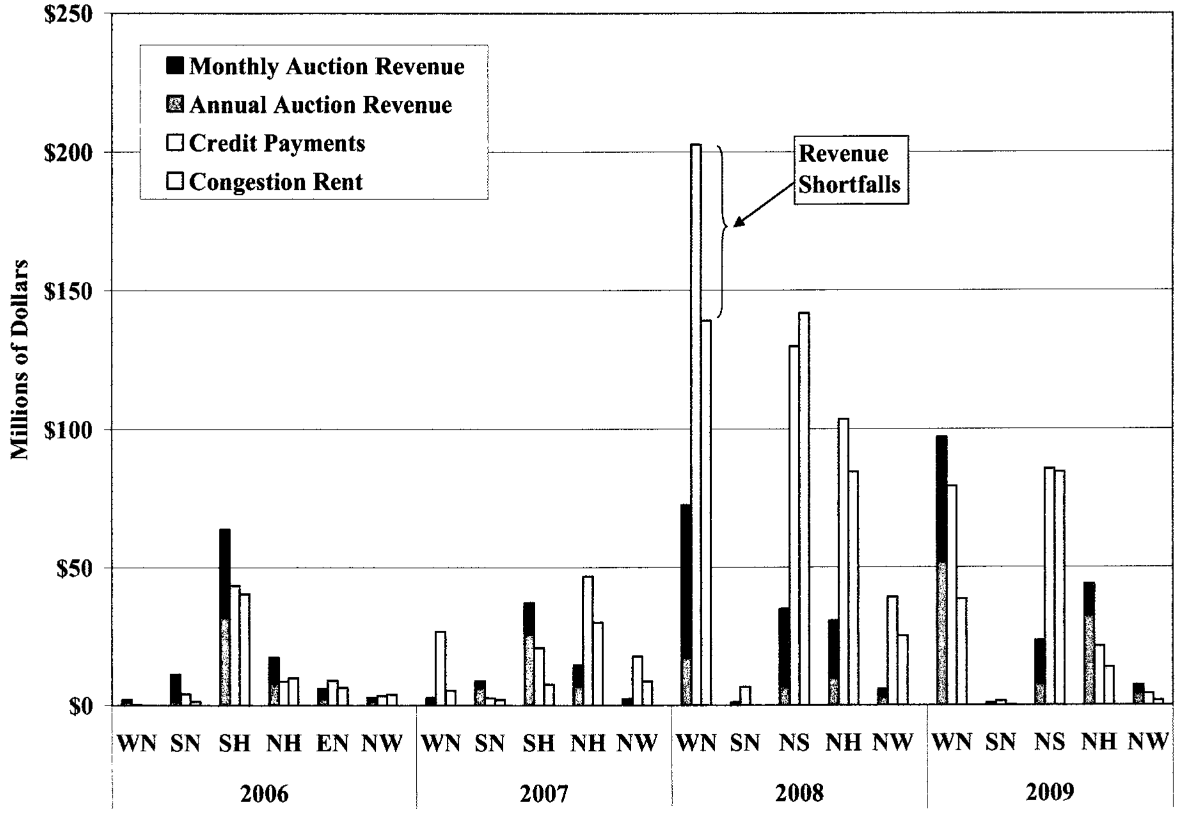


Figure 62 shows that the West to North and North to Houston had the most significant revenue shortfalls in 2009. When congestion rents fall significantly below payments to TCR holders, it implies that the SPD-calculated flows across constrained interfaces have been systematically lower than the amount of TCRs sold for the interfaces.

Figure 62 also shows that payments to TCR holders have consistently exceeded the congestion rents that have been collected from the balancing market in 2006 through 2009. Congestion rents covered 90, 47 and 79 percent of payments to TCR holders in 2006, 2007 and 2008, respectively. In 2009, Congestion rents covered 72 percent of the payments to TCR holders, with an annual net revenue shortfall of $53 million.

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

D. Local Congestion and Local/System Capacity Requirements

In this subsection, we address local congestion and local and system reliability requirements by evaluating how ERCOT manages the dispatch and commitment of generators when constraints and reliability requirements arise that are not recognized or satisfied by the current zonal markets. Local (or intrazonal) congestion occurs in ERCOT when a transmission constraint is binding that is not defined as part of a CSC or CRE. Hence, these constraints are not managed by the zonal market model. ERCOT manages local congestion by requesting that generating units adjust their output quantities (either up or down). When insufficient capacity is committed to meet local or system reliability requirements, ERCOT commits additional resources to provide the necessary capacity in either the day-ahead market or in the adjustment period, which includes the hours after the close of the day-ahead market up to one hour prior to real-time. Capacity required for local reliability constraints is procured through either the Replacement Reserve Service market ("Local RPRS") or as out-of-merit capacity ("OOMC"). Some of this capacity is also instructed to be online through Reliability Must Run ("RMR") contracts. Capacity required for system reliability requirements (*i.e.,* the requirement that the total system-wide online capacity be greater than or equal to the sum of the ERCOT load forecast plus operating reserves in each hour) is procured through either the RPRS market ("Zonal RPRS") or as OOMC.

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

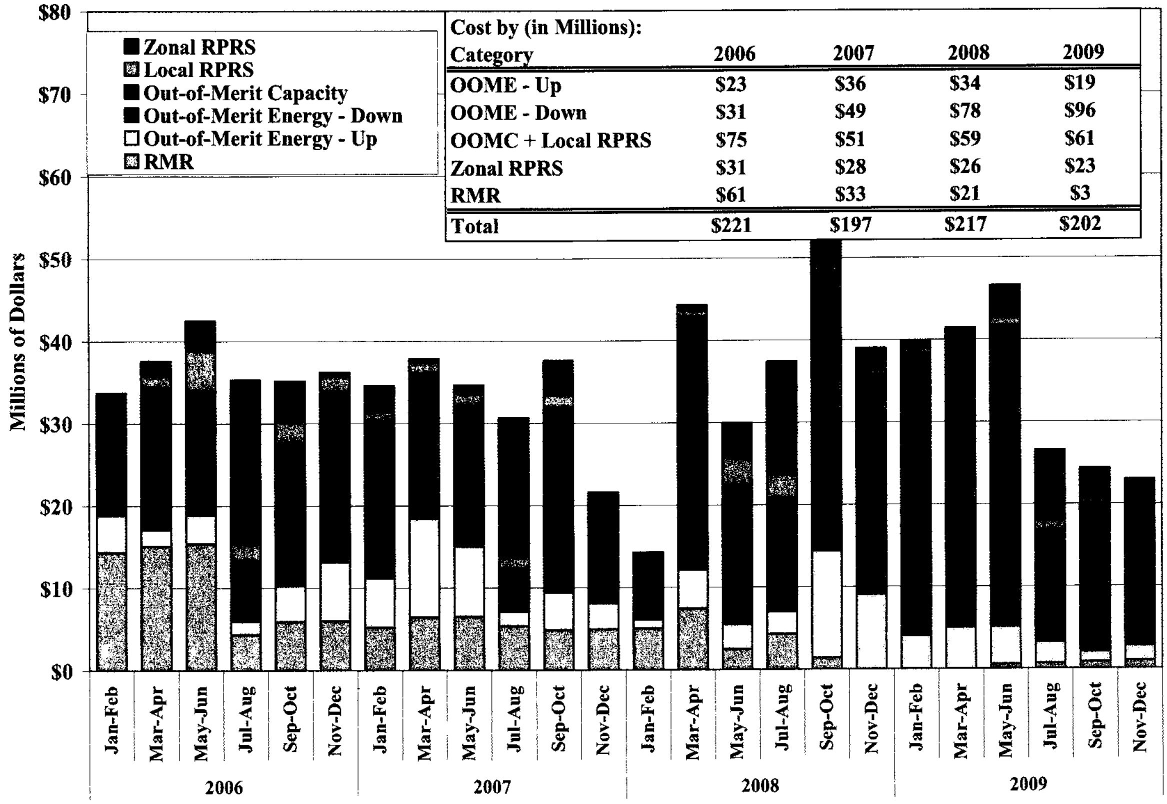
When a unit is dispatched out of merit (OOME up or OOME down), the unit is paid for a quantity equal to the difference between the scheduled output based on the unit's resource plan and the actual output resulting from the OOME instruction from ERCOT. The payment per MWh for OOME is a pre-determined 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 or Local RPRS 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 or Local RPRS instruction from ERCOT are obligated to offer any available energy from the resource into the balancing energy market. Zonal RPRS is selected based upon offer prices for startup and minimum energy and resources procured for Zonal RPRS are paid the market clearing price for this service.

Finally, RMR units committed or dispatched pursuant to their RMR agreements receive cost-based compensation. Since October 2002, ERCOT has entered into several RMR agreements with older, inefficient units that were planned to be retired. As a part of the RMR exit strategy process, all units were removed from RMR status by October 2008; however, two additional units entered into RMR agreements in May 2009. Units contracted to provide RMR service to ERCOT are compensated for start-up costs, energy costs, and are also paid a standby fee. Figure 63 shows each of the four categories of uplift costs from 2006 to 2009.

Figure 63: Expenses for Out-of-Merit Capacity and Energy



Cost by (in Millions): Category	2006	2007	2008	2009
OOME - Up	$23	$36	$34	$19
OOME - Down	$31	$49	$78	$96
OOMC + Local RPRS	$75	$51	$59	$61
Zonal RPRS	$31	$28	$26	$23
RMR	$61	$33	$21	$3
Total	$221	$197	$217	$202

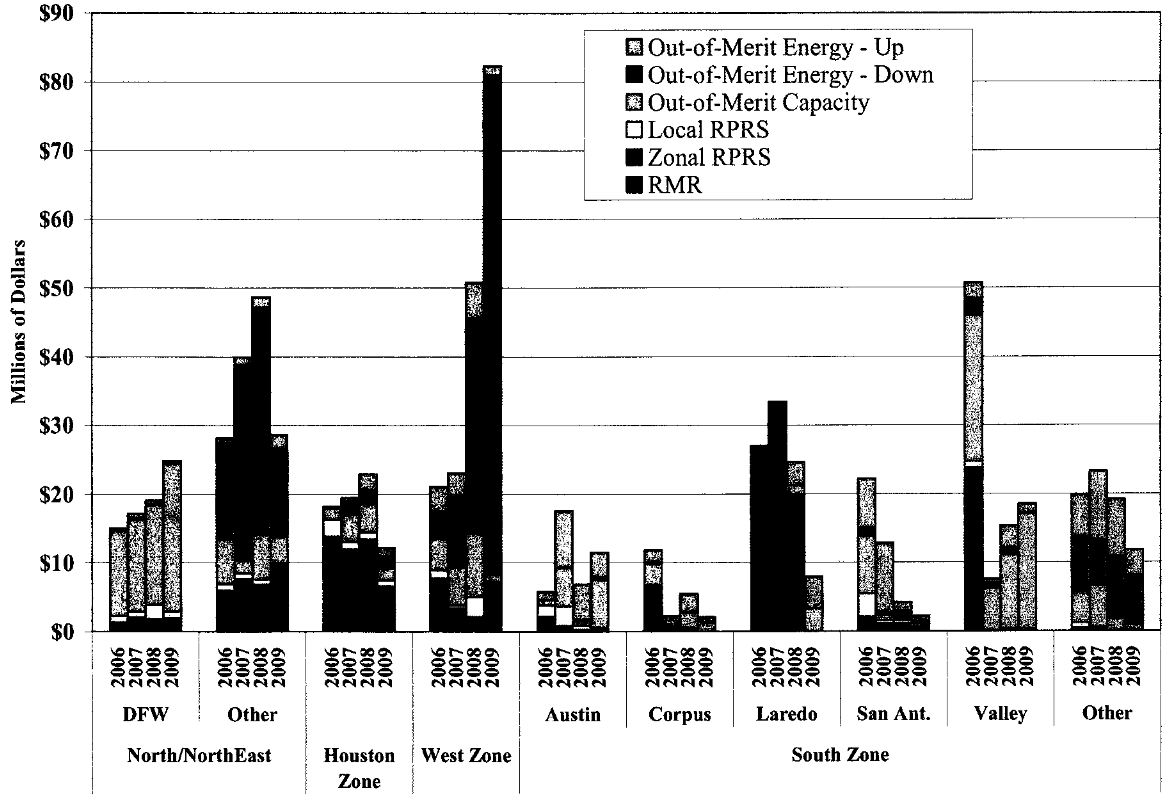
The results in Figure 63 show that overall uplift costs for RMR units, OOME units, OOMC/Local RPRS and Zonal RPRS units were $202 million in 2009, which is a $15 million decrease over the $217 million in 2008.[28] OOME Down and RMR costs accounted for the most significant portion of the change in 2009. OOME down increased from $78 million in 2008 to $96 million in 2009. These values represent significant increases in OOME Down costs from 2006 and 2007, and are primarily attributable to increases in OOME Down instructions for wind resources in the West Zone. RMR cost decreased from $21 million in 2008 to $3 million in 2009. Figure 63 also shows that the highest Zonal RPRS costs occur in July and August when electricity demand in the ERCOT region is at its highest levels.

Although the costs are borne by load throughout ERCOT, the costs are caused in specific locations because these actions, with the exception of zonal RPRS, are taken to maintain local

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

reliability. The rest of the analyses in this section evaluate in more detail where these costs were caused and how they have changed between 2006 and 2009. Figure 64 shows these payments by location.

Figure 64: Expenses for OOME, OOMC and RMR by Region



The most significant changes in local congestion costs in 2009 compared to 2008 shown in Figure 64 are as follows:

- OOME Down costs in the West Zone increased by $42 million in 2009. This increase was associated with the significant addition of wind capacity in the West Zone. OOMC cost in the West Zone decreased by $8 million in 2009.
- OOME Down costs in the North Zone decreased by $20 million in 2009. This decrease can be attributed to fewer transmission outages requiring the reduced output of coal/lignite units.
- RMR costs in the Laredo area of the South Zone decreased by $20 million to zero in 2009. This decrease was associated with the termination of the Laredo RMR contract in October 2008.
- OOMC costs in the Valley area of the South Zone increased by $6 million in 2009. This increase was associated with the more frequent need for local capacity to be online to maintain Rio Grande Valley import limits.

IV. 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 2009. We examine market structure by using a pivotal supplier analysis that indicates suppliers were pivotal in the balancing energy market in 2009 much less frequently than in 2007 and 2008 and significantly less frequently than in 2005 and 2006. These results highlight the trend of continued improvement in the structural competitiveness of the balancing energy market over the last five years. This analysis also shows that the frequency with which a supplier was pivotal increased at higher levels of demand, which is consistent with observations in prior years. To evaluate participant conduct we estimate measures of physical and economic withholding. We examine withholding patterns relative to the level of demand and the size of each supplier's portfolio. Based on these analyses, we find the overall performance of the ERCOT wholesale market to be competitive in 2009.

A. Structural Market Power Indicators

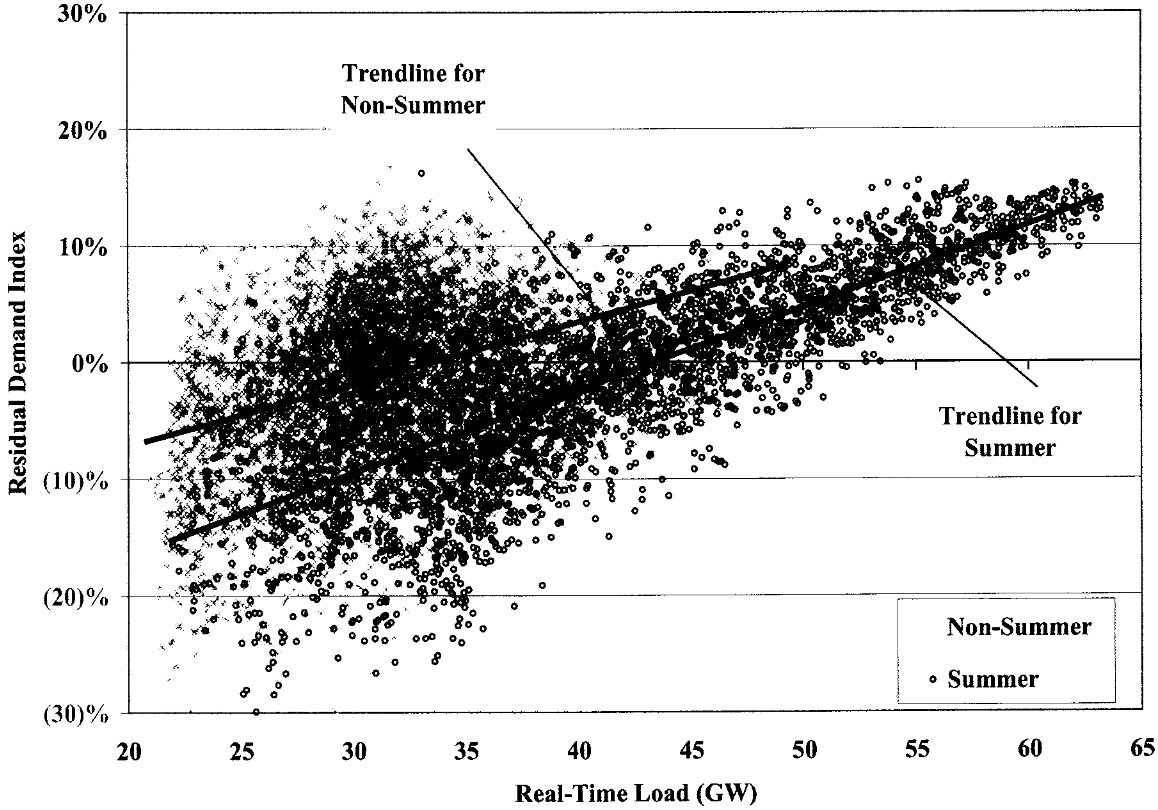
We analyze market structure by using the Residual Demand Index ("RDI"), a statistic that measures the percentage of load that could not be satisfied without the resources of the largest supplier. When the RDI is greater than zero, the largest supplier is pivotal (*i.e.*, its resources are needed to satisfy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

The RDI is a useful structural indicator of potential market power, although it is important to recognize its limitations. As a structural indicator, it does not illuminate actual supplier behavior to indicate whether a supplier may have exercised market power. The RDI also does not indicate whether it would have been 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 65 shows the RDI relative to load for all hours in 2009. The data are divided into two groups: (i) hours during the summer months (from May to September) are shown by darker points, while (ii) hours during other months are shown by lighter points. The trend lines for each data series are also shown and indicate a strong positive relationship between load and the RDI.

This analysis shown below is done at the QSE level because the largest suppliers that determine the RDI values own a large majority of the resources they are scheduling or offering. It is possible that they also control the remaining capacity through bilateral arrangements, although we do not know whether this is the case. To the extent that the resources scheduled by the largest QSEs are not controlled or providing revenue to the QSE, the RDIs will tend to be slightly overstated.

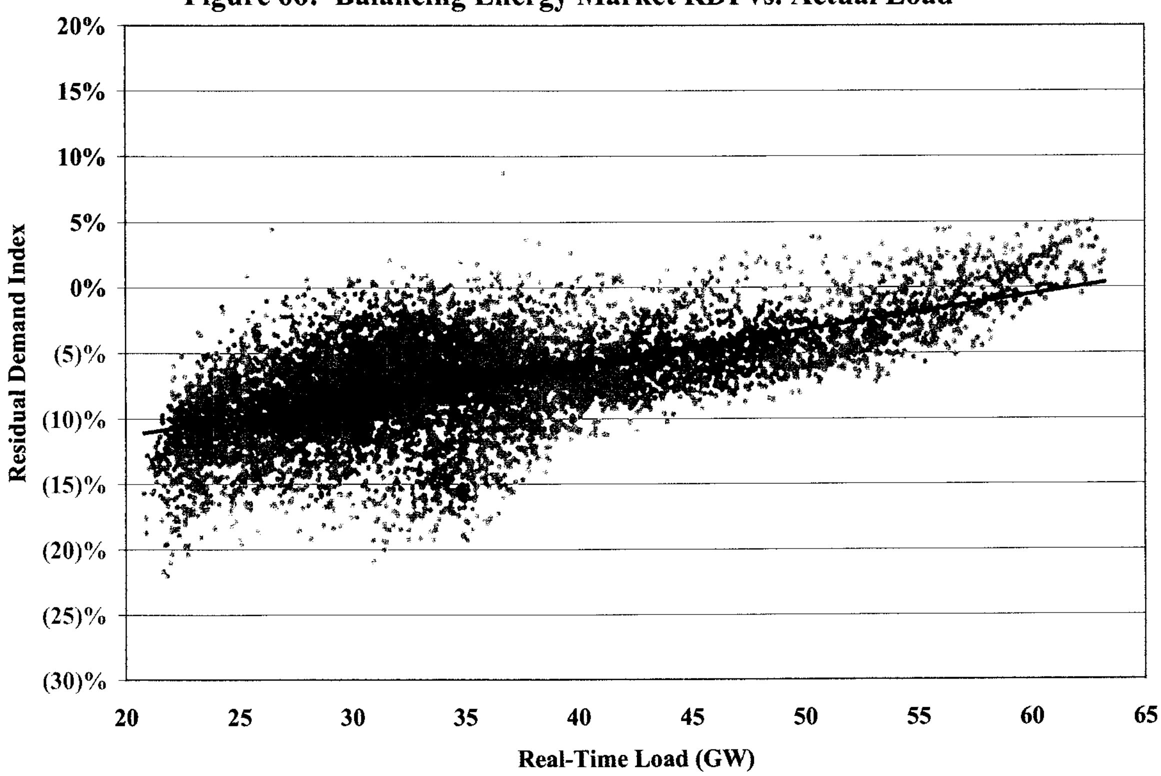
Figure 65: Residual Demand Index



The figure shows that the RDI for the summer (i.e. May to September) was usually positive in hours when load exceeded 40 GW. During the summer, the RDI was greater than zero in approximately 46 percent of all hours, reduced from 60 percent in 2008. The RDI was comparable at lower load levels during the spring and fall due to the large number of generation planned outages and less commitment. Hence, although the load was lower outside the summer, our analysis shows that a QSE was pivotal in approximately 46 percent of all hours during the non-summer period, reduced from 70 percent in 2008. It is important to recognize that inferences regarding market power cannot be made solely from this data. Retail load obligations can affect the extent of market power for large suppliers, since such obligations cause them to be much smaller net sellers into the wholesale market than the analysis above would indicate. Bilateral contract obligations can also affect a supplier's potential market power. For example, a smaller supplier selling energy in the balancing energy market and through short-term bilateral contracts may have a much greater incentive to exercise market power than a larger supplier with substantial long-term sales contracts. The RDI measure shown in the previous figure does not consider the contractual position of the supplier, which can increase a supplier's incentive to exercise market power compared to the load-adjusted capacity assumption made in this analysis.

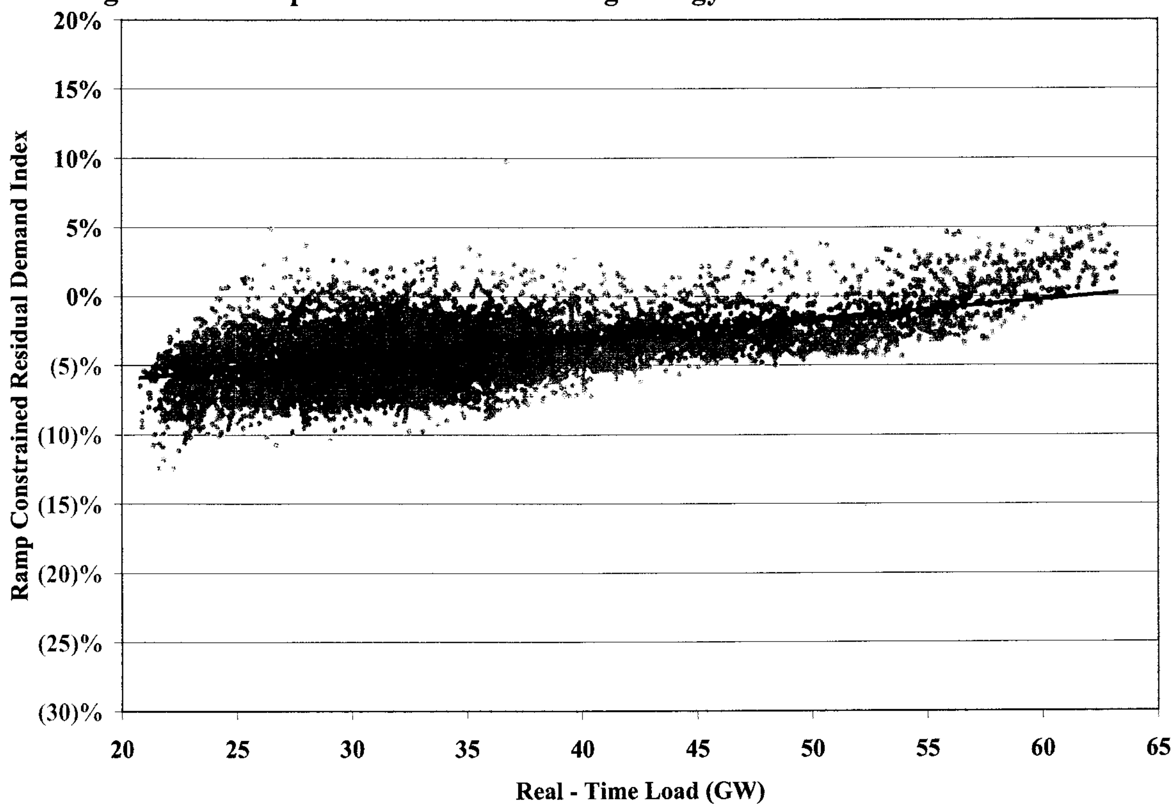
In addition, a supplier's ability to exercise market power in the current ERCOT balancing energy market may be higher than indicated by the standard RDI. Hence, a supplier may be pivotal in the balancing energy market when it would not have been pivotal according to the standard RDI shown above. To account for this, we developed RDI statistics for the balancing energy market. Figure 66 shows the RDI in the balancing energy market relative to the actual load level.

Figure 66: Balancing Energy Market RDI vs. Actual Load



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[29] owned by other suppliers. Figure 66 limits the other supplier's capacity to the capacity offered in the balancing energy market. When the RDI is greater than zero, the largest supplier's balancing energy offers are necessary to prevent a shortage of offers in the balancing energy market. Figure 67 shows the same data as in Figure 66 except that the balancing energy offers are further limited by portfolio ramp constraints in each interval.

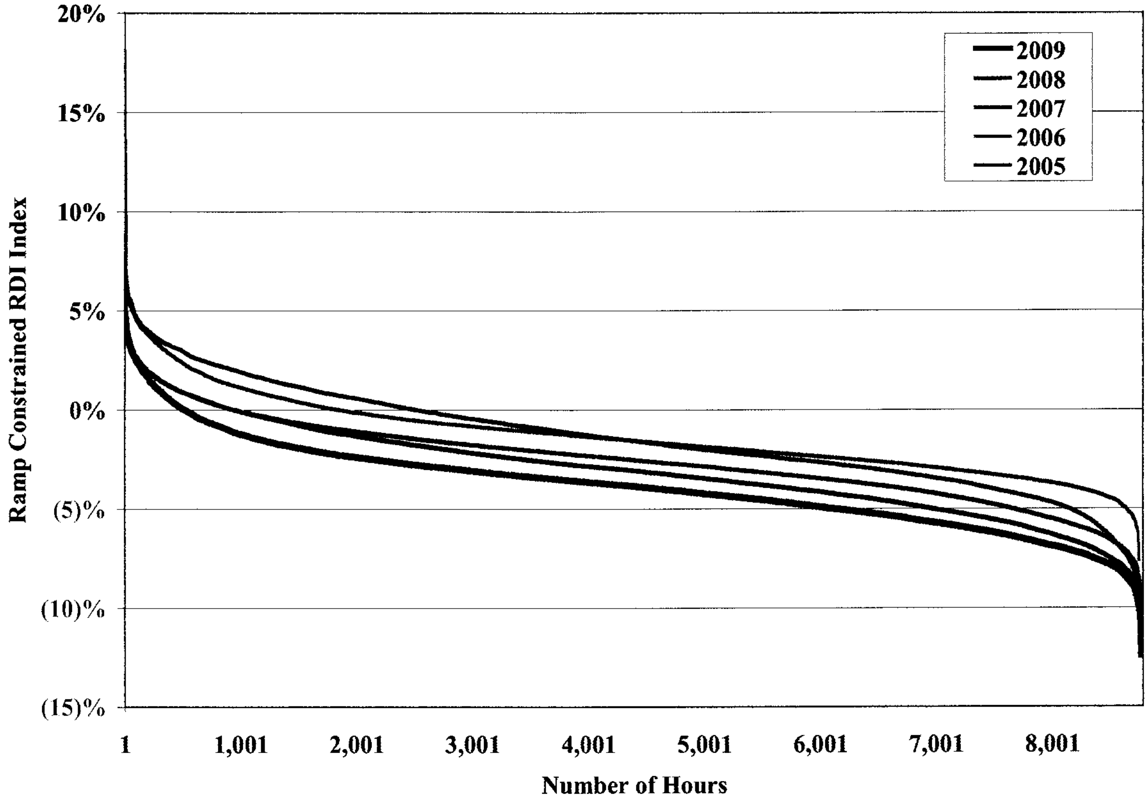
Figure 67: Ramp-Constrained Balancing Energy Market RDI vs. Actual Load



In 2009, the instances when the RDI was positive occurred over a wide range of load levels, from 25 GW to 63 GW. The balancing energy market RDI data and trend line for 2009 are similar in shape to prior years, with the frequency with which a supplier was pivotal generally increasing at higher levels of demand. However, the frequency of data points that are positive in 2009 is smaller than the frequency in prior years. This difference is highlighted in Figure 68, which compares the balancing energy market RDI duration curves for 2005 through 2009.

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

Figure 68: Ramp-Constrained Balancing Energy Market RDI Duration Curve



The frequency with which at least one supplier was pivotal in the balancing energy market *(i.e.*, an RDI greater than zero) has fallen consistently over the last five years from 29 and 21 percent of the hours in 2005 and 2006, respectively, to less than 11 percent of the hours in 2007 and 2008, to less than 6 percent of the hours in 2009. These results highlight the trend of continued improvement in the structural competitiveness of the balancing energy market over the last five years.

B. Evaluation of Supplier Conduct

The previous subsection presented a structural analysis that supports inferences about potential market power. In this section we evaluate actual participant conduct to assess whether market participants have attempted to exercise market power through physical or economic withholding. First, we review offer patterns in the balancing energy market. Then we examine unit deratings and forced outages to detect physical withholding and we evaluate the "output gap" to detect economic withholding.

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

1. Balancing Energy Market Offer Patterns

In this section, we evaluate balancing energy offer patterns by analyzing the rate at which capacity is offered.[30] Figure 69 shows the average amount of capacity offered to supply up balancing service relative to all available capacity.

Figure 69 shows a seasonal variation in 2009 over time in quantities of energy available and offered to the balancing energy market. Up balancing offers are divided into the portion that is capable of being deployed in one interval and the portion which would take longer due to portfolio ramp rate offered by the QSE (*i.e.*, "Ramp-Constrained Offers"). Capacity that is available but un-offered is represented by the white dashed portion of each column in the chart.

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

Figure 69: Balancing Energy Offers Compared to Total Available Capacity Daily Peak Load Hours

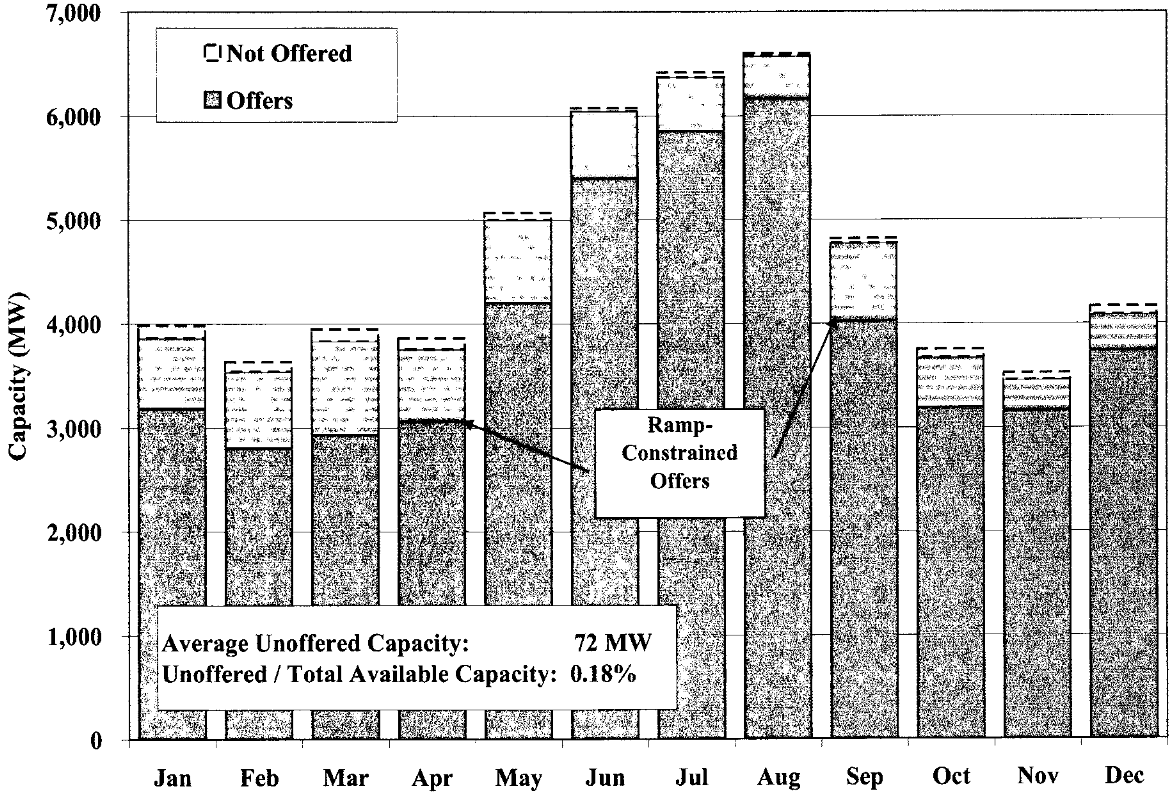


Figure 69 shows a seasonal variation in the quantity of energy available and offered in the balancing energy market, with higher quantities in the summer months than in the non-summer months. This figure also shows that the quantities of un-offered capacity were relatively small in all months in 2009.

Un-offered capacity can raise competitive concerns to the extent that it reflects withholding by a dominant supplier that is attempting to exercise market power. To investigate whether this has occurred, Figure 70 shows the same data as the previous figure, but arranged by load level for daily peak hours in 2009. Because prices are most sensitive to withholding under the tight conditions that occur when load is relatively high, increases in the un-offered capacity at high load levels would raise competitive concerns.