the demand for energy, responsive reserve, up regulation and non-spinning reserve provided from online capacity or quick-start units. To evaluate the commitment of resources in ERCOT, Figure 37 plots the excess capacity in ERCOT during 2008. The figure shows the excess capacity in only the peak hour of each weekday because largest amount of additional generation commitment usually occurs at the peak hour. Hence, one would expect larger quantities of excess capacity in other hours.

Figure 37: Excess On-Line and Quick Start Capacity During Weekday Daily Peaks

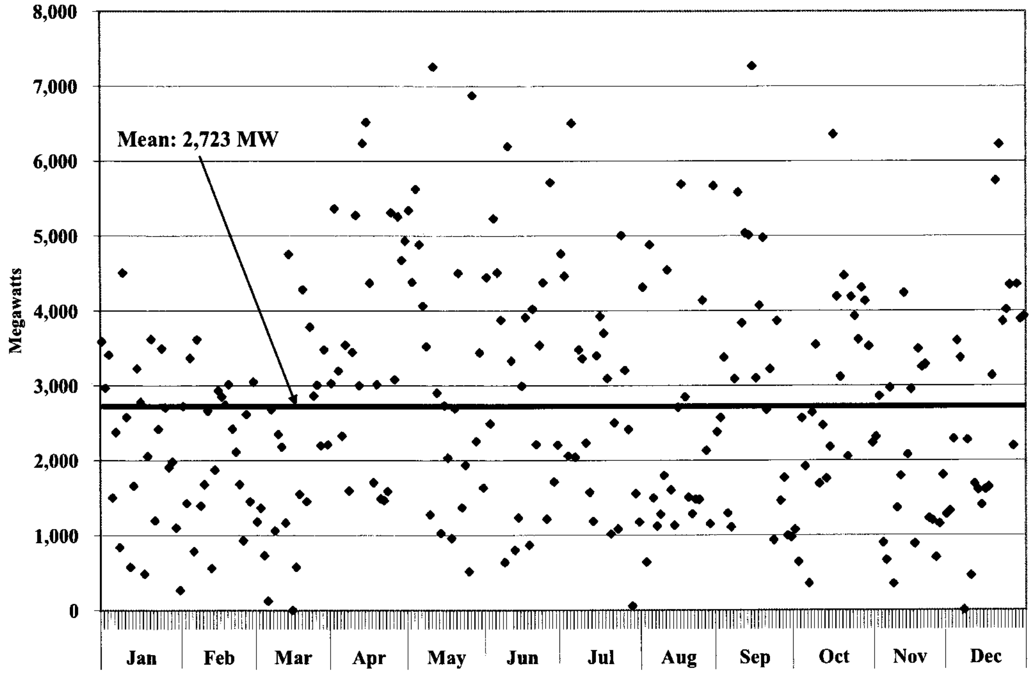


Figure 37 shows that the excess on-line capacity during daily peak hours on weekdays averaged 2,723 MW in 2008, which is approximately 7.6 percent of the average load in ERCOT. This is a reduction from prior years in which the same measure of on-online capacity average averaged 4,313, 2,927, and 3,020 MW in 2005, 2006 and 2007, respectively.

The overall trend in excess on-line capacity also indicates a movement toward more efficient unit commitment across the ERCOT market; however, the current market structure is still based primarily upon a decentralized unit commitment process whereby each participant makes independent generator commitment decisions that are not likely to be optimal. Further

provide balancing energy.

contributing to the suboptimal results of the current unit commitment process is that the decentralized unit commitment is comprised of non-binding resource plans that form the basis for ERCOT's day-ahead planning decisions. However, these non-binding plans can be modified by market participants after ERCOT's day ahead planning process has concluded causing ERCOT to take additional actions that may be more costly and less efficient. Hence, the introduction of a day-ahead energy market with centralized Security Constrained Unit Commitment ("SCUC") that is financially binding under the nodal market design promises substantial efficiency improvements in the commitment of generating resources.

D. Demand Response Capability

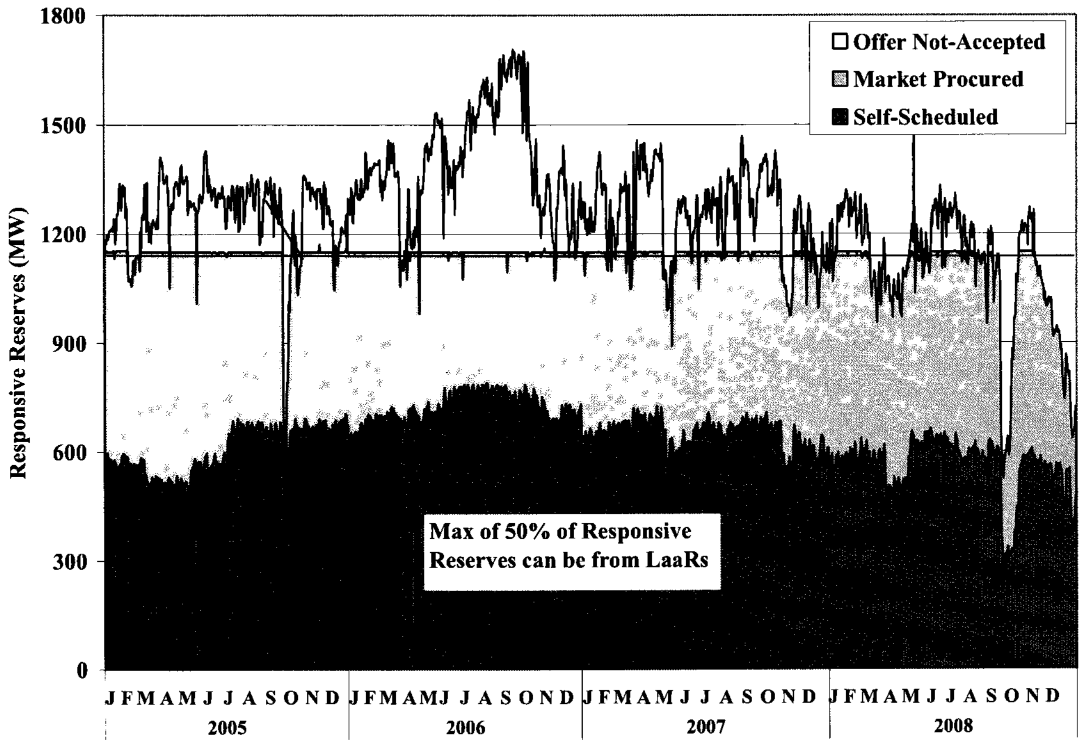
Demand response is a term that broadly refers to actions that can be taken by end users of electricity to reduce load in response to instructions from ERCOT or in response to certain market or system conditions. The ERCOT market allows participants with demand-response capability to provide energy and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

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

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

As of December 2008, 2,158 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 38 shows the amount of responsive reserves provided from LaaRs on a daily basis in 2008.

Figure 38: Provision of Responsive Reserves by LaaRs
Daily Average



The high level of participation by demand response participating in the ancillary service markets sets ERCOT apart from other operating electricity markets. Figure 38 shows that the amount of responsive reserves provided by LaaRs has held fairly constant at 1,150 MW since the beginning of 2005 (for reliability reasons, 1,150 MW is the limit of participation in the responsive reserve market by LaaRs). Notable exceptions were a period in September/October 2005 corresponding to Hurricane Rita, and a more prolonged decrease in September/October of 2008 corresponding to the Texas landfall of Hurricane Ike. Of interest in late 2008 is the post-hurricane recovery of the quantity of LaaRs providing Responsive Reserve followed by a steady reduction for the remainder of the year, which 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 or regulation services markets and their participation in the non-spinning reserves market was negligible in 2008. 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.

One change that may increase the participation in the non-spinning reserve market is the implementation of Protocol Revision Request No. 776 that was developed in late 2008 and was implemented in May 2009. This change will allow LaaRs that choose to provide this service to receive a daily capacity payment as in the past, but will also allow the LaaR to then offer the non-spinning reserve capacity into the balancing energy market that will determine the energy price at which the LaaR is willing to curtail its load. This change offers two benefits for LaaRs. First, assuming that the opportunity cost for LaaRs is typically much higher than the marginal cost of generating resources, the probability of deployment for LaaRs providing non-spinning reserves will be lower than in the past when non-spinning reserves were deployed independent of the marginal cost of the providers. Second, by allowing for the deployments to be based on energy price offers, LaaRs providing non-spinning reserves will be able to better predict and control the economics of providing the service in light of their particular business circumstances, as opposed to the prior practice in which non-spinning reserves were deployed as a price taker. These changes are expected to lead to the entry of some quantity of LaaRs into the non-spinning reserve market, which would obviously be an improvement compared to the history of no participation by LaaRs in the non-spinning reserve market.

E. Net Revenue Analysis

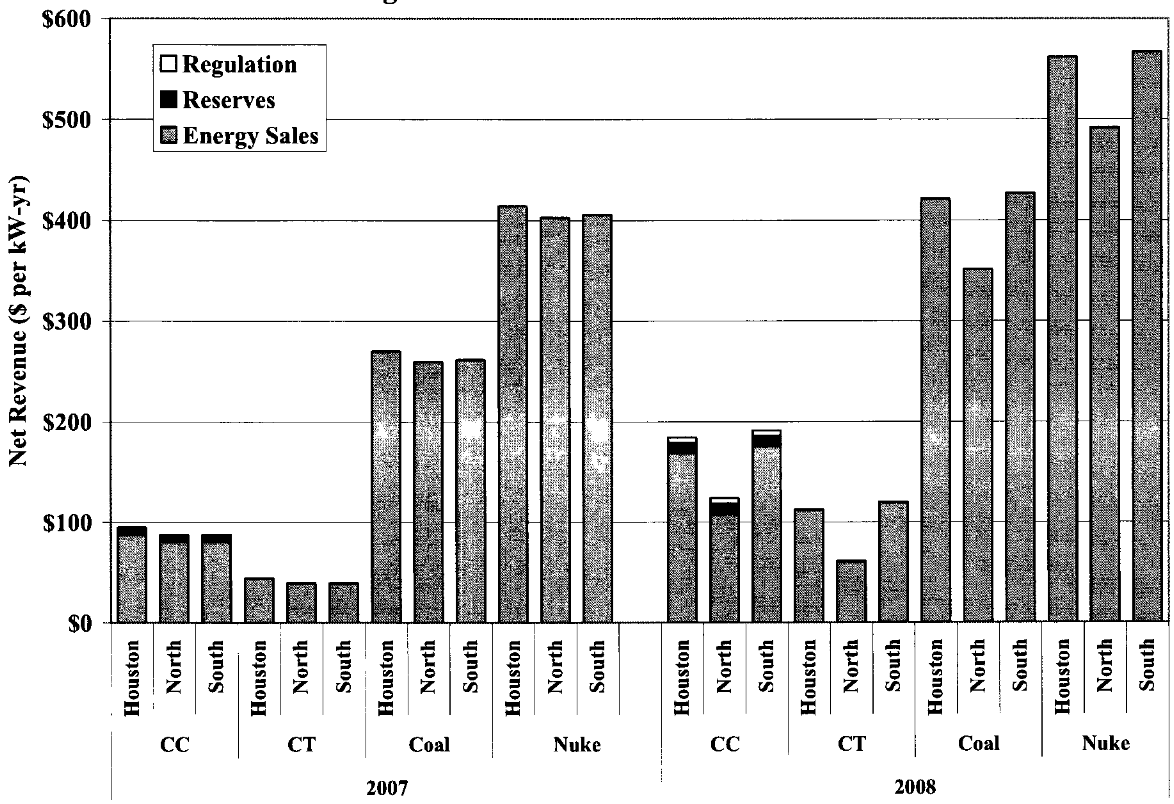
Net revenue is defined as the total revenue that can be earned by a generating unit less its variable production costs. Hence, it is the revenue in excess of short-run operating costs and is available to recover a unit's fixed and capital costs. Net revenues from the energy, operating reserves, and regulation 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 39 shows the results of the net revenue analysis for four types of units in 2007 and 2008. These are: (a) a gas combined-cycle, (b) a combustion turbine, (c) a new coal unit, and (d) a new nuclear unit. In recent years, most new capacity investment has been in natural gas-fired technologies, although high prices for oil and natural gas have caused renewed interest in new investment in coal and nuclear generation. For the gas-fired technologies, net revenue is calculated by assuming the unit will produce energy in any hour for which it is profitable and by assuming it will be available to sell reserves and regulation in other hours that it is available (*i.e.*, when it is not experiencing a planned or forced outage). For coal and nuclear technologies, net revenue is calculated by assuming that the unit will produce at full output. The energy net revenues are computed based on the balancing energy price in each hour. Although most suppliers would receive the bulk of their revenues through bilateral contracts, the spot prices produced in the balancing energy market should drive the bilateral energy prices over time.

Figure 39: 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 MMbtu per MWh for a new coal unit. We assume variable operating and maintenance costs of $4 per MWh for the gas units and $1 per MWh for the coal unit. We assume variable costs of $5 per MWh for the nuclear unit. For each technology, we assumed a total outage rate (planned and forced) of 10 percent.

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 39 shows that the net revenue increased substantially in 2008 in each zone compared to 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 2008 for a new gas turbine was approximately $120, $113 and $61 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 2008 for a new combined cycle unit was approximately $191, $185 and $124 per kW-year in the South, Houston and North Zones, respectively. These values indicate that the estimated net revenue in 2008 was sufficient to support new entry for a new gas turbine in the South and Houston zones and for a combined cycle unit in the South, Houston and North zones. However, as discussed later in this subsection, significant portions of the net revenue results for gas turbine and combined cycle units in 2008 can be attributed to anomalous market design related inefficiencies rather than fundamentals that would support an investment decision for new gas turbines and combined-cycle units.

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 have allowed energy prices to remain at levels high enough to support new entry for these technologies. The production costs of coal and nuclear units did not change significantly over this period, leading to a dramatic rise in net revenues. For a new coal unit, the estimated net revenue requirement is approximately $190 to $245 per kW-year. The estimated net revenue in 2008 for a new coal unit was approximately $427, $421 and $351 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 2008 for a new nuclear unit was approximately $567, $562 and $492 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 sufficient to support new entry in 2008, as was the case in 2005, 2006 and 2007. Thus, it is not surprising that some market participants are building new baseload facilities and that several others have initiated activities that may lead to the construction of additional baseload facilities in the ERCOT region.

Although estimated net revenue grew considerably in 2008 compared to prior years, there are other factors that determine incentives for new investment. First, market participants must anticipate how prices will be affected by the new capacity investment, future load growth, and increasing participation in demand response. Second, net revenues can be inflated when prices clear above competitive levels as a result of market power being exercised. Thus, a market participant may be deterred from investing in new capacity if it believes that prevailing net revenues are largely due to an exercise of market power that would not be sustainable after the entry of the new generation. Third, the nodal market design will have an effect on the profitability of new resources. In a particular location, nodal prices could be higher or lower than the prices in the current market depending on the pattern of congestion. Finally, and most importantly in 2008, net revenues can be inflated when prices clear at high levels due to inefficiencies in the market design. Similar to the case of market power, a market participant may be deterred from investing in new capacity if it believes that prevailing net revenues are largely due to market design inefficiencies that will be corrected.

Such market design inefficiencies were apparent in 2008. As discussed in Section III, the vast majority of price excursions in 2008 – particularly in the South and Houston Zones – were not a function of market fundamentals; rather, the price excursions were driven by inefficient congestion management techniques that have since been corrected and are not expected to materialize in the future, especially upon implementation of the nodal market in 2010. In addition to these transmission congestion issues, in 2008 the ERCOT Protocols provided for *ex post* re-pricing provisions in intervals in which non-spinning reserve prices were deployed that frequently resulted in scarcity level prices at times when ERCOT's operating reserve levels were not deficient. These rules were changed as a part of the aforementioned PRR 776, thereby reducing the probability of scarcity level prices during non-scarcity conditions going forward. Hence, a significant portion of the net revenue produced in 2008 is not reflective of fundamentals that would support an investment decision for new gas turbines and combined cycle units.

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 40 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.[14]

Figure 40: Comparison of Net Revenue of Gas-Fired Generation between Markets

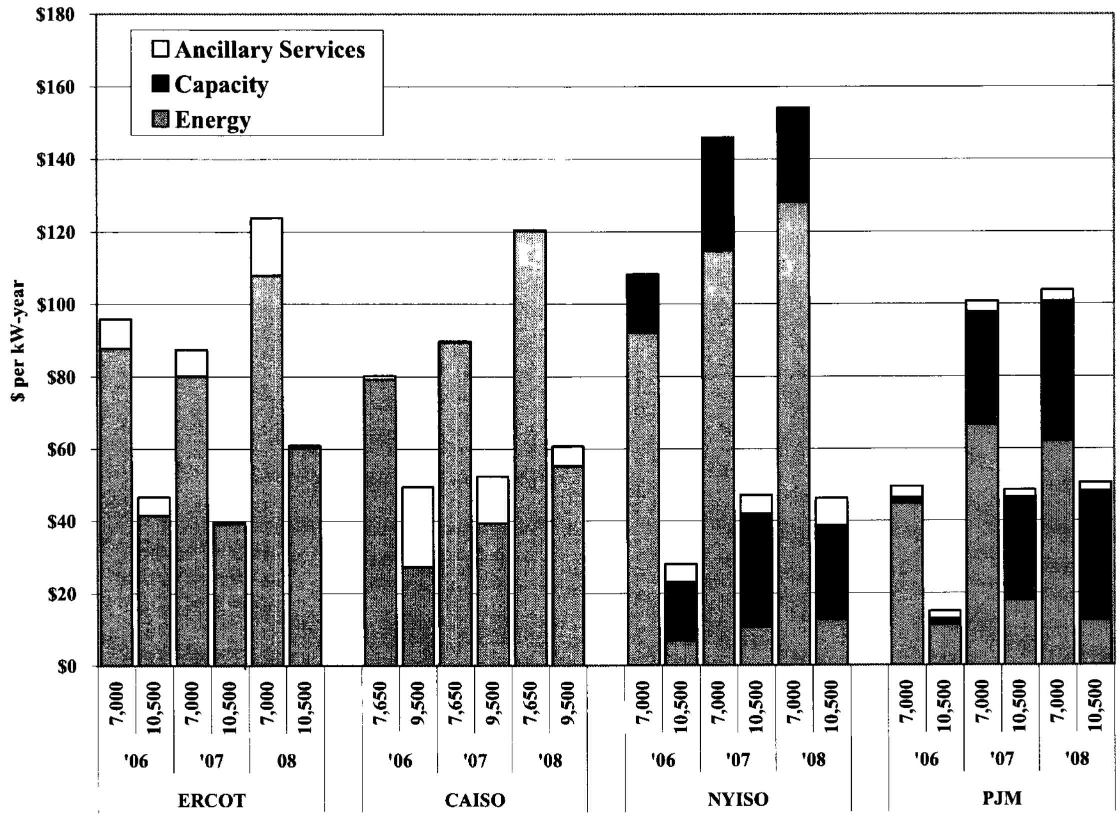


Figure 40 shows that net revenues increased in all markets from 2007 to 2008, with the exception of gas peaking units in New York that remained flat. In the figure above, net revenues are calculated for central locations in each of the five markets. However, there are load pockets within each market 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.

14 The California ISO does not report capacity and ancillary services net revenue separately, so it is shown as a combined block in Figure 40. 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.

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.

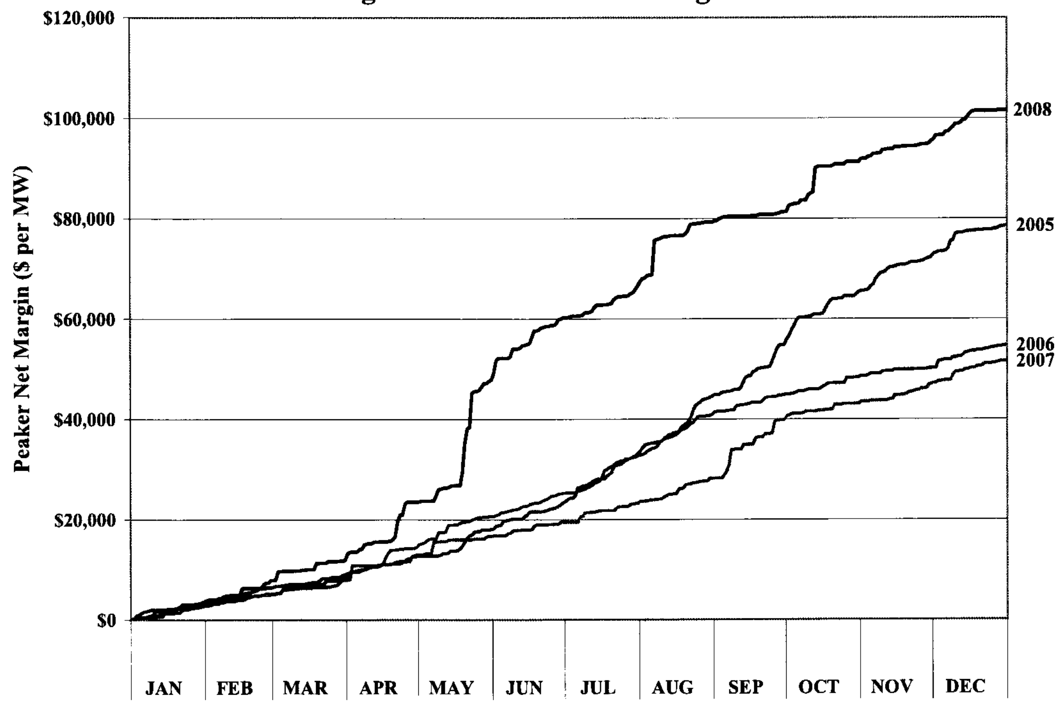
PUCT Subst. Rule 25.505 provides that the IMM may conduct an annual review of the effectiveness of the SPM. This subsection provides an assessment of the effectiveness of the SPM in 2008 under ERCOT's energy-only market structure.

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. Although the PNM was not in effect prior to 2007, Figure 41 shows the cumulative PNM that would have been produced for each year from 2002 through 2007.[15]

Figure 41: 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 41 and consistent with the previous findings in this section relating to net revenue, the PNM reached the level sufficient for new entry in only two of the last five years (2005 and 2008). In 2008, the peaker net margin and net revenue values rose substantially, surpassing the level required to support new peaker entry. However, as previously discussed, a significant portion of the net revenue increase in 2008 was associated with extremely inefficient transmission congestion management and inefficient

15 The proxy combustion turbine in the Peaker Net Margin calculation uses a heat rate of 10 MMbtu per MWh and includes no other variable operating costs.

pricing mechanisms associated with the deployment of non-spinning reserves. Both of these issues have been corrected in the zonal market and will be further improved with the implementation of the nodal market in 2010. Absent these inefficiencies, net revenues would not have been sufficient to support new peaker entry in 2008. Beyond these anomalies, there were three other factors that significantly influenced the effectiveness of the SPM in 2008:

- A substantial decrease in out-of-merit deployments by ERCOT during declared short-supply conditions;
- A continued strong positive bias in ERCOT's day-ahead load forecast that tended to regularly commit online resources in excess of the quantity required to meet expected demand and operating reserve requirements; and
- The dependence on market participants to submit offers at or near the offer cap to produce scarcity level prices during legitimate shortage conditions.

1. Out-of-Merit Deployments during Shortage Conditions

As discussed in the 2007 SOM Report, ERCOT implemented new operating procedures in 2007 whereby it deployed Non-Spinning Reserve Service ("NSRS") when Adjusted Responsive Reserves ("ARR") were reduced to 2,500 MW. If NSRS was not procured, had already been deployed, or could not be timely deployed, ERCOT issued out-of-merit ("OOM") instructions to offline, quick-start units. ARR is a measure that is based upon available responsive reserves, but incorporates a discount factor that is applied to the capacity of online generating units. This discount factor was developed by ERCOT based on prior experience during emergency operating conditions, and is intended to account for the uncertainty in the actual maximum capacity that is deliverable when called upon during emergency conditions.

Although well-intended from a reliability perspective, from a market efficiency perspective, the use of the discount factor in 2007 created an "overlap" between market and reliability operations that often led to inefficient pricing outcomes during shortage and near-shortage conditions.

Efforts in 2007 to address these inefficiencies led to an interim measure that was implemented in January 2008 that increased the procurement of responsive reserves to offset the effect of the application of the discount factor, thereby significantly reducing the "overlap" between market and reliability operations that was frequently experienced in 2007. The responsive reserve procurement increase was linked directly to the magnitude of the discount factor. Additionally, Protocol Revision Request No. 750 was adopted in 2007 that provided for unannounced generator capacity testing with the objective of providing ERCOT enough confidence to eliminate the discount factor, which would also eliminate the increased procurement of responsive reserves.

In the 2007 SOM Report, we noted that "implementation of PRR 750 in 2008 will not only lead to the elimination of the discount factor, but will also eliminate the interim measure of increased procurement of responsive reserves. Ultimately, the successful implementation of PRR No. 750 should lead to more reliable and efficient operations in the ERCOT wholesale market." In fact, by August 2008, ERCOT had performed sufficient unit testing under PRR 750 that provided it with the confidence to substantially reduce the discount factor to a level that eliminated the increased procurements of responsive reserves. Together, the interim increase in responsive reserve procurements and implementation of PRR 750 worked very successfully to virtually eliminate the out-of-merit deployments by ERCOT during shortage conditions in 2008, thereby improving both the efficiency and reliability of market operations.

2. ERCOT Day-Ahead Load Forecast Error

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

An integral piece of the RPRS market is the day-ahead load forecast. If the day-ahead load forecast is significantly below actual load and no subsequent actions are taken, ERCOT may run the risk of being unable to meet reliability criteria in real-time. In contrast, if the day-ahead load forecast is significantly high, the outcome may be an inefficient commitment of excess online capacity in real-time.

Figure 42: Average Day Ahead Hourly Load Forecast Error by Season

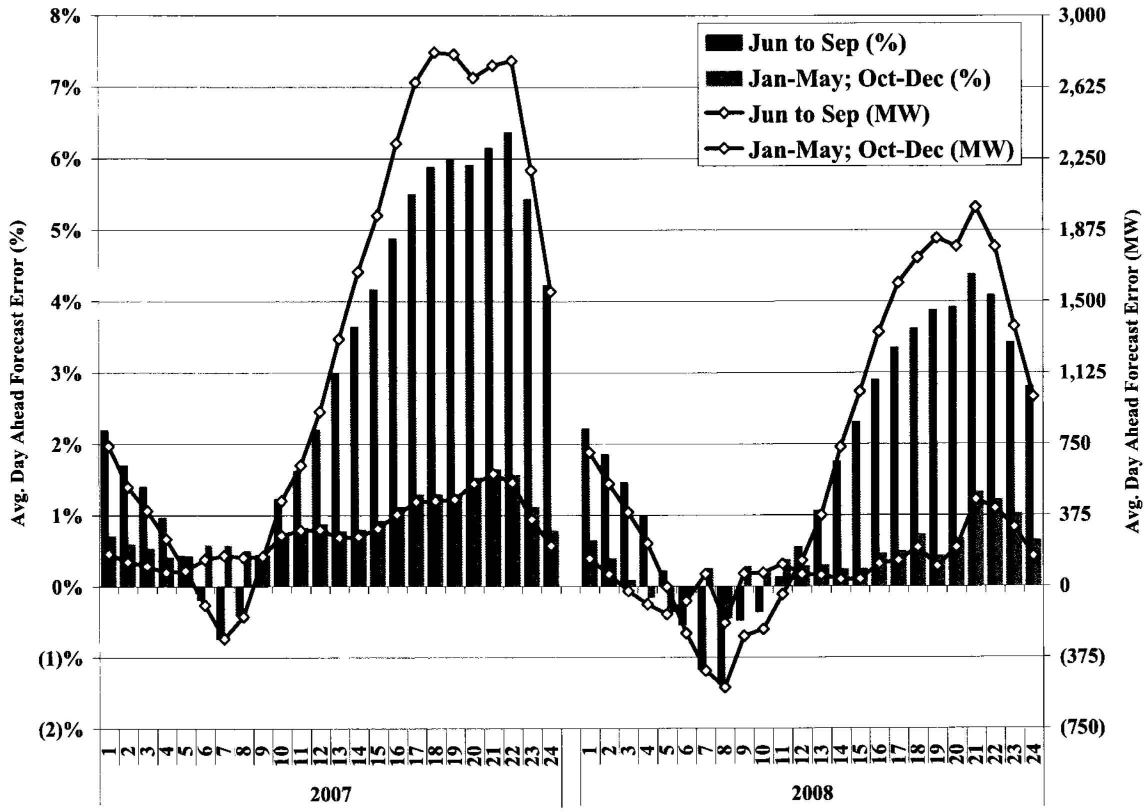
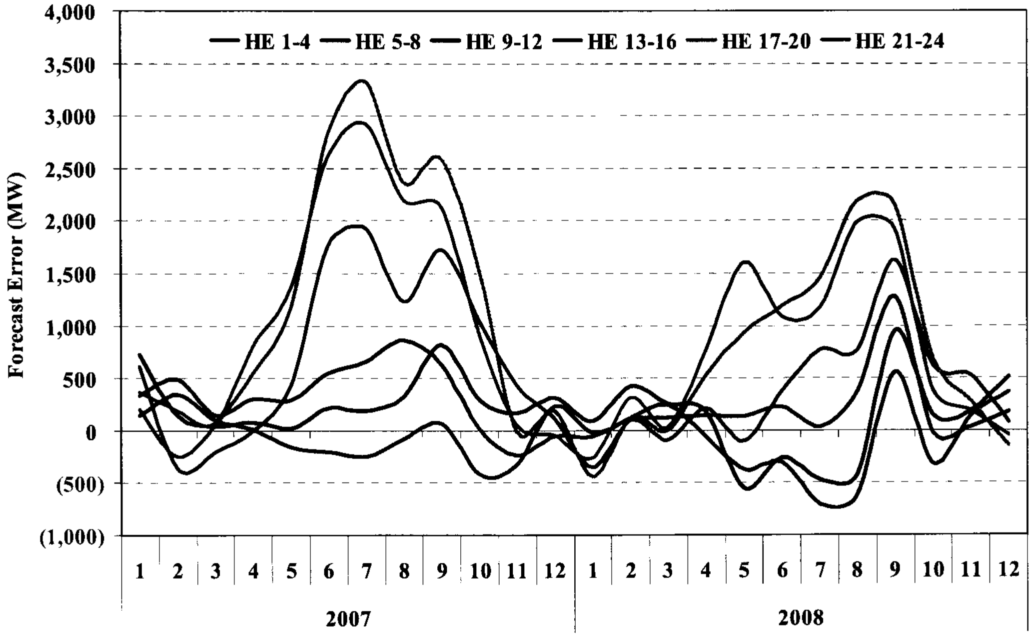


Figure 42 shows the average hourly day-ahead load forecast error for the summer months of June through September, and also for the months of January through May and October through December for 2007 and 2008. In this figure, positive values indicate a day-ahead load forecast that was greater than the 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 and 2008, 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 205 MW for the non-summer months, and an average over-forecast of 1,729 MW for the four summer months in 2008. Although the performance in 2008 was generally improved compared to 2007, Figure 42 clearly shows that the positive day-ahead load forecast bias observed in 2007 persists in 2008. Figure 43 shows another view of the same load forecast error data as in Figure 42 for 2007 and 2008 with the average megawatt error displayed for each month in four hour blocks (hours ending).

Figure 43: Average Day Ahead Load Forecast Error by Month and Hour Blocks



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.[16]

3. Dependence on High-Priced Offers by Market Participants

As a general principle, competitive and efficient market prices should be consistent with the cost of the marginal action taken to satisfy the market's demand. In the vast majority of hours, the marginal action is the dispatch of the most expensive online generator. 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:

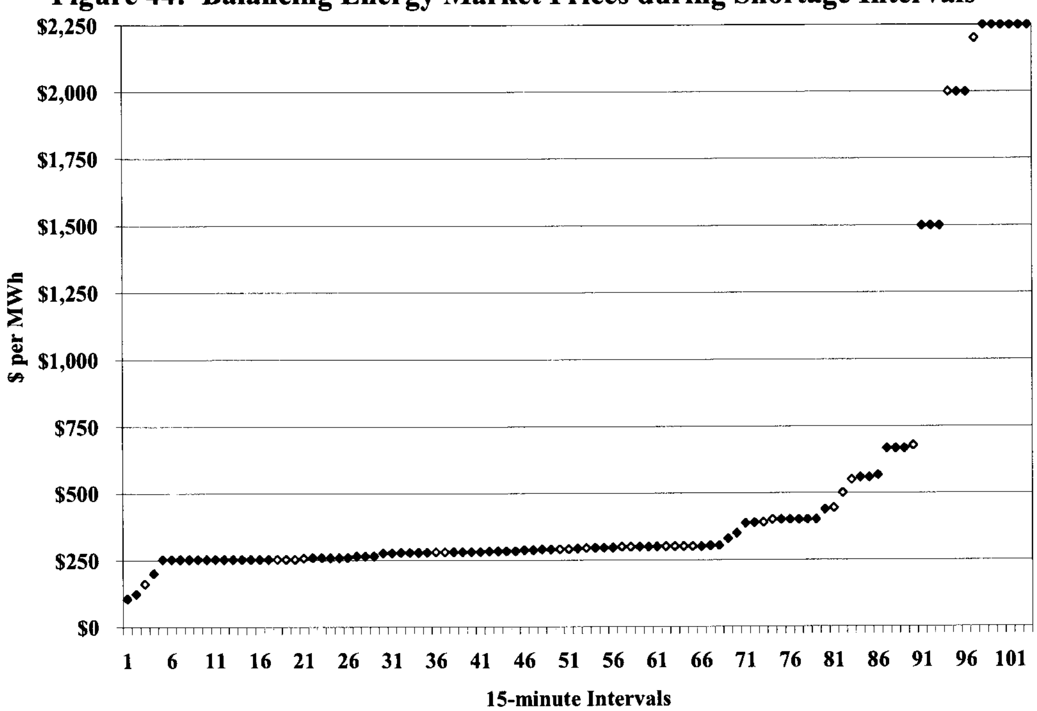
16 It is our understanding that ERCOT's current procedures allow to some extent for the deferral of the commitment of short-lead time resources.

- 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 44 shows the balancing market clearing prices during the 103 15-minute shortage intervals in 2008.

Figure 44: Balancing Energy Market Prices during Shortage Intervals



As shown in Figure 44, the prices during these 103 shortage intervals in 2008 ranged from $105 per MWh to the offer cap of $2,250 per MWh (prior to March 1, 2008, the offer cap was $1,500 per MWh), with an average price of $534 per MWh and a median price of $293 per MWh. The results in 2008 are similar to those in 2007 when there were 108 shortage intervals with an average price of $476 per MWh and a median price of $299 per MWh.

The data in Figure 44 are separated into solid blue and red outlined points. The blue points (79) represent true shortage conditions, whereas the red points (24) represent artificial shortage prices occurring as a result of large generation schedule reductions at the top of the hours from 10 PM to 1 AM. As discussed in more detail in Section I, the production of such artificial shortage prices under these conditions is the result of inefficiencies inherent to the current market design that will be significantly improved with the implementation of the nodal market.

Although each of the data points in Figure 44 represents system conditions in which the market was in shortage, the pricing outcomes are widely varied, with the majority of prices reflecting the marginal cost of the most expensive generation resource dispatched as opposed to the value of foregone operating reserves. 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 and 2008.

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 are largely driven by significant increases in natural gas prices in recent years. In contrast, private investment in mid-merit and peaking resources in ERCOT has been relatively thin. In an energy-only market, net revenue expectations for mid-merit and peaking resources are much more sensitive the effectiveness of the shortage pricing mechanism than to factors such as 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 small market participants to effectively withhold lower cost resources by offering at prices dramatically higher than their marginal cost.

At least for the pendency of the zonal market, shortage pricing will continue to remain dependent upon the existence of high-priced offers by market participants, and results such as those experienced in 2007 and 2008 will continue to frustrate the objectives of the energy-only market design.[17] 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 shortage conditions. While important even in markets with a capacity market, efficient shortage pricing is a particularly critical element in the ERCOT energy-only market to ensure that the long-term resource adequacy requirements are achieved.

17 Net revenue in 2008 was sufficient to support new entry for peaking resources in the Houston and South Zones. However, as discussed in Section III, the vast majority of price excursions in 2008 were not a function of market fundamentals; rather, the price excursions were driven by inefficient congestion management techniques and pricing mechanisms associated with the deployment of non-spinning reserves that have been corrected and are not expected to materialize in the future. Hence, the net revenue values produced in 2008 are generally not reflective of fundamentals that would support an investment decision for new gas turbine and combined cycle units.

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.

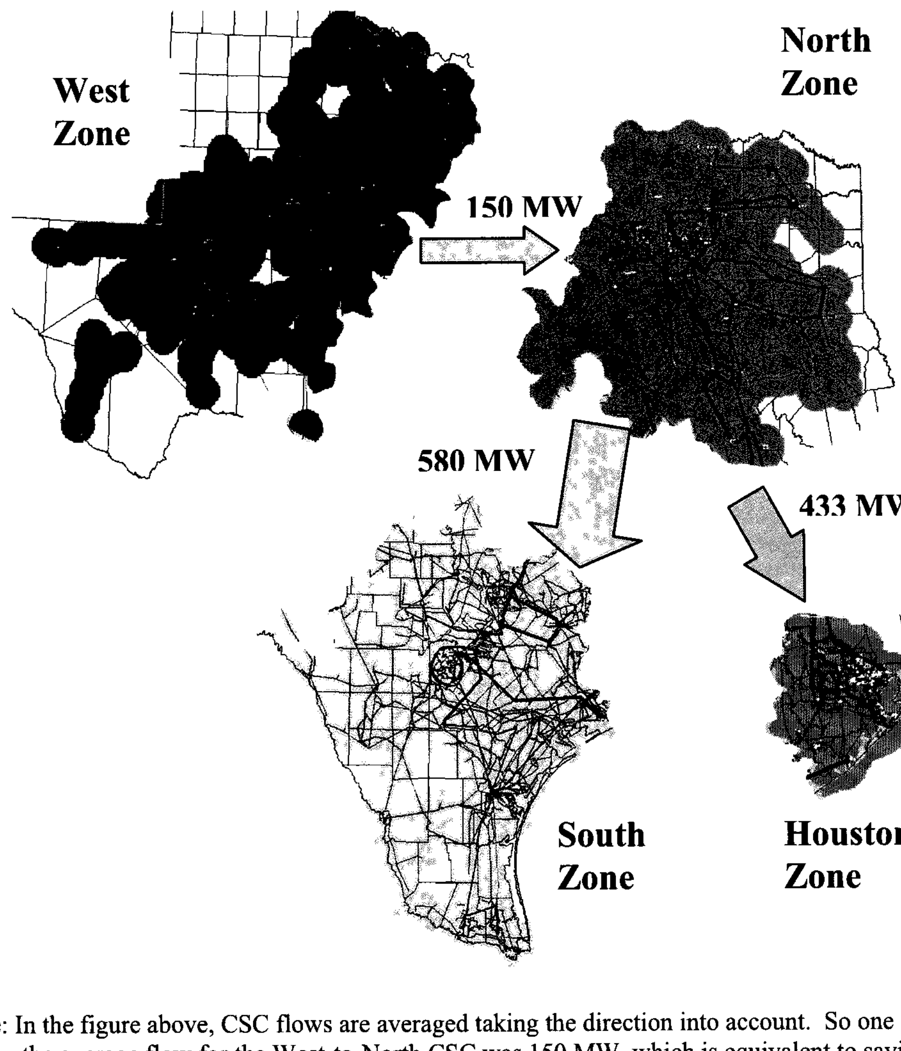
A. Electricity Flows between Zones

In 2008 there were four commercial pricing zones in ERCOT: (a) the North Zone, (b) the West Zone, (c) the South Zone, and (d) the Houston Zone. From year-to-year, slight adjustments are sometimes made to the boundaries of the commercial pricing zones. However, the vast majority of customers remained in the same zone from 2007 to 2008. 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 2008.

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 45: Average SPD-Modeled Flows on Commercially Significant Constraints During All Intervals in 2008



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 150 MW, which is equivalent to saying that the average for the North-to-West CSC was *negative* 150 MW.

Figure 45 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. [18] In reality, the generators and loads within each zone can have widely varying 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 2008	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	150	139	*-11*	128	*-11*
North - South	580	570	*-10*	-44	*-614*
North-Houston	433	436	*3*	753	*317*

The first column in Table 2 shows the average flows over each CSC calculated by SPD. The second column shows the average flows over each CSC calculated using zonal-average 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.

18 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 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 614 MW and increased the calculated flows on the North to Houston CSC by 317 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 46 shows the actual average imports of power for each zone in 2008. In this figure, positive values represent imports, and negative values indicate exports.[19]

[19] 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 46.

Figure 46: Actual Zonal Net Imports

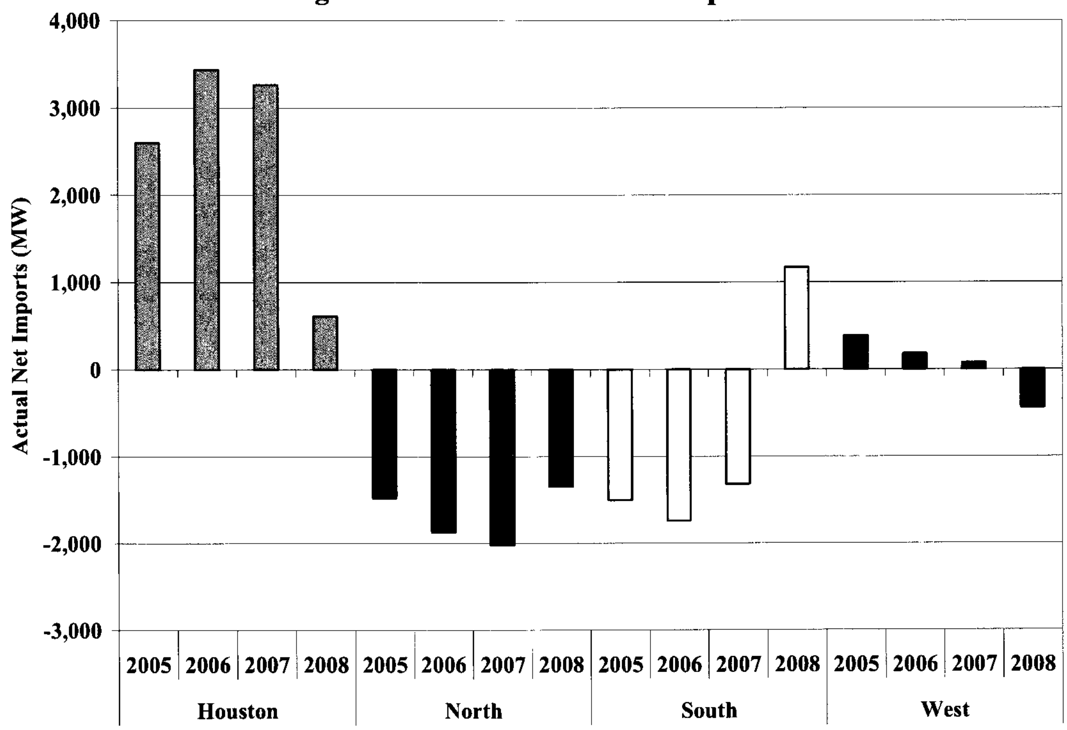


Figure 46 shows that the Houston Zone is a net importer of power, while the North Zone is a net exporter. The reduction in the Houston Zone imports in 2008 and corresponding change in the South Zone from a net exporter to a net importer can be attributed to the movement of the 2,700 MW South Texas Nuclear Project from the South Zone to the Houston Zone in 2008. The West Zone transitioned from a net importer in 2005 to a net exporter in 2008, which is reflective of 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 47 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 47: Average SPD-Modeled Flows on Commercially Significant Constraints During Transmission Constrained Intervals in 2008

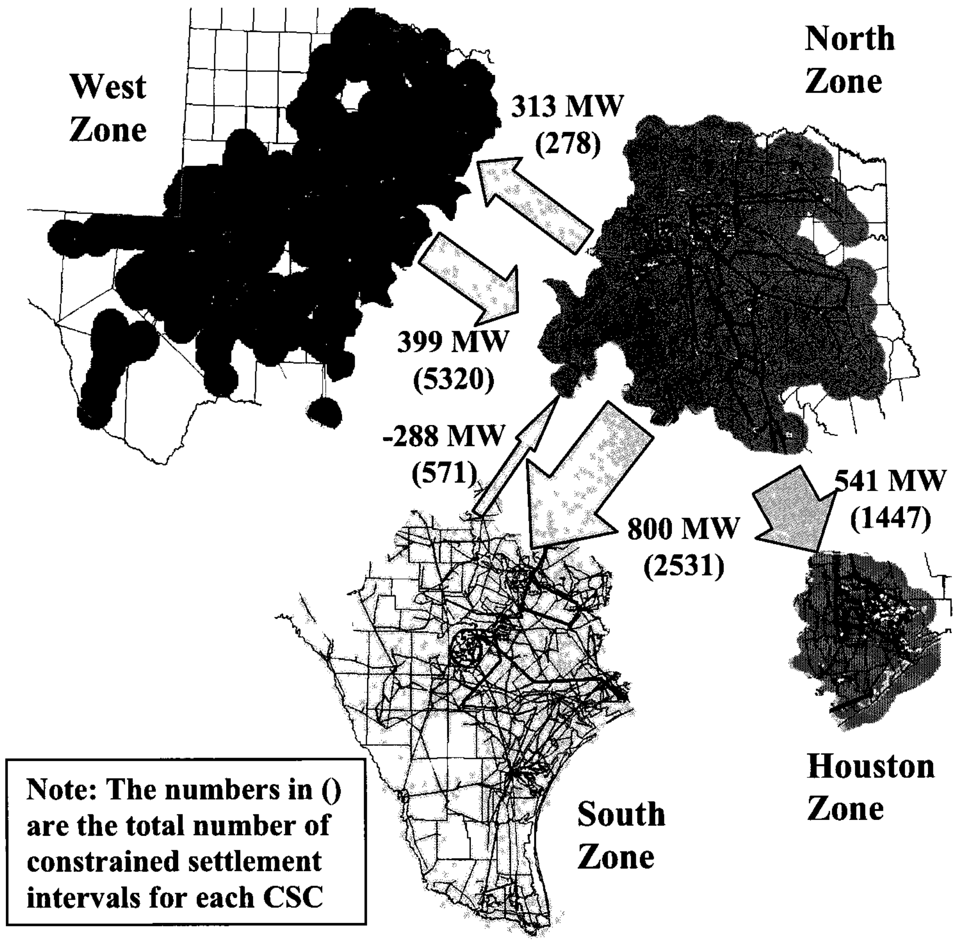


Figure 47 shows that inter-zonal congestion was most frequent in 2008 on the West to North and the North to South CSCs, followed by the North to Houston CSC. The West to North CSC exhibited SPD-calculated flows averaging 399 MW during 5,320 constrained intervals (15 percent of the totals intervals in the year). The North to South CSC exhibited SPD-calculated flows averaging 800 MW during 2,531 constrained intervals (7 percent of the total intervals), and the SPD-calculated average flow for the North to Houston CSC was 541 MW during 1,447 constrained intervals (4 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 2008	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	399	345	*-54*	335	*-10*
North - South	800	633	*-167*	139	*-494*
North-Houston	541	418	*-123*	827	*409*

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 CSC were relatively close to actual flows, whereas the average actual flows for the North to South 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 2008.

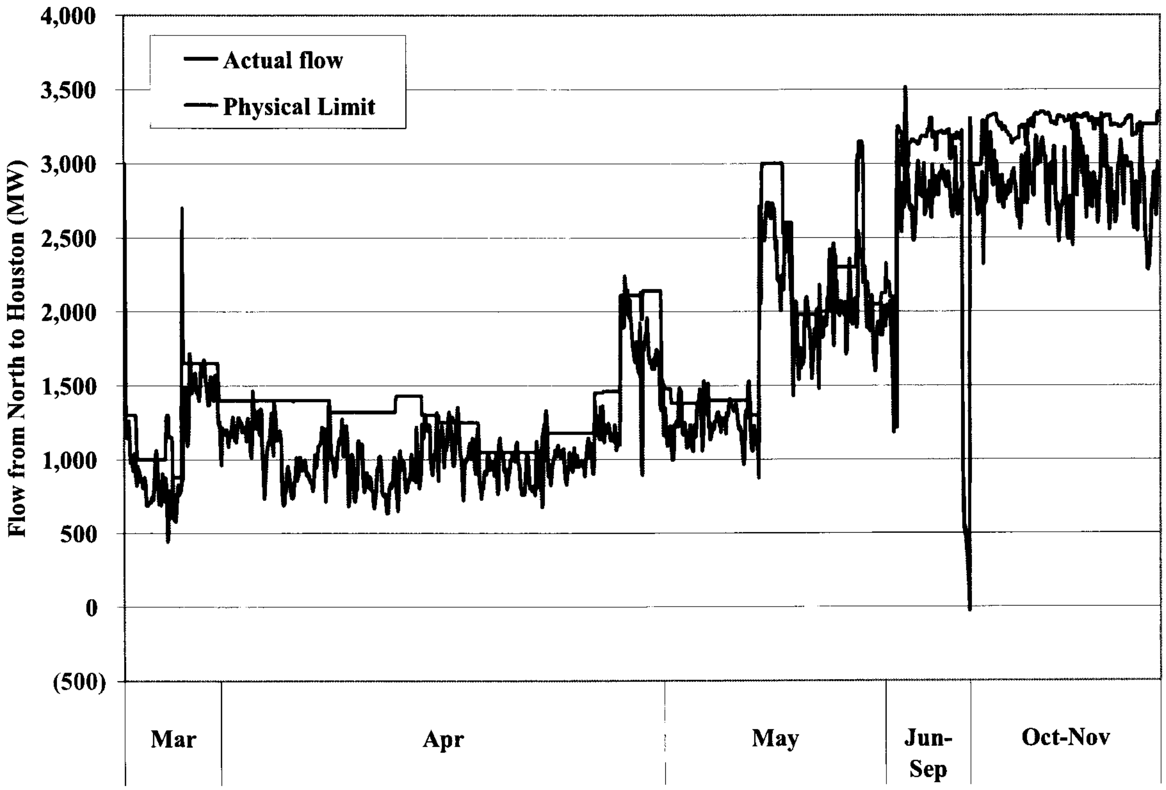
1. Congestion on North to South and North to Houston CSCs

The majority of congestion activity for the North to South and North to Houston CSCs was affected by common factors in 2008. Therefore, the analysis of both of these CSCs is provided in this single subsection.

The North to South CSC was new in 2008 and was added based on experience in 2007 of increased local congestion activity on this interface. In 2008 the North to South CSC was binding in 2,531 intervals (7 percent) with an annual average shadow price of $22 per MW. The North to Houston CSC was binding in 1,447 intervals (4 percent) with an annual average shadow price of $20.

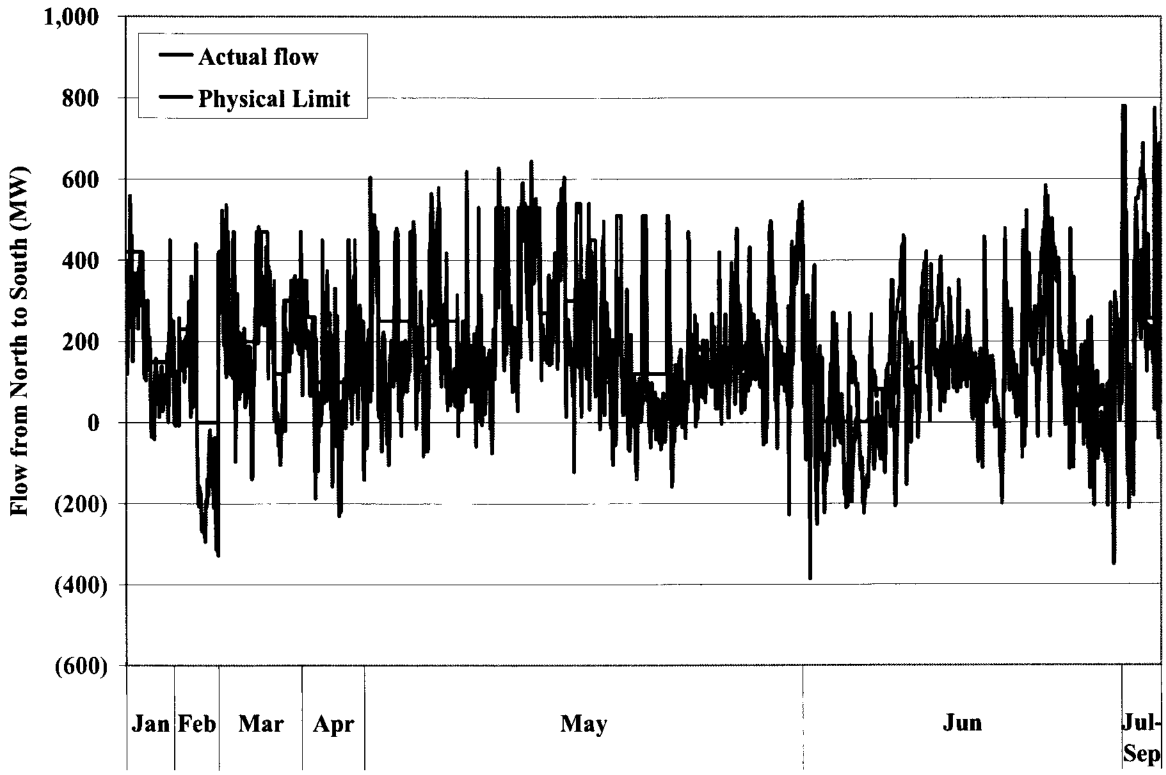
Beginning in April and continuing into May 2008, the frequency of congestion on the North to Houston CSC began to increase, at times becoming so significant that the constraint was unable to be resolved with available balancing energy in the zones where it was required. When congestion on a CSC cannot be resolved, maximum shadow prices are produced for the CSC that, in turn, produce balancing energy market prices in the deficient zones that can approach or even exceed the system-wide offer caps (the system-wide offer cap was $2,250 per MWh beginning March 1, 2008). Figure 48 shows the actual flows versus the physical limit for the North to Houston CSC in 2008 during intervals when the CSC was binding.

Figure 48: Actual Flows versus Physical Limits during Congestion Intervals North to Houston



Similar to the North to Houston CSC, the frequency of congestion on the North to South CSC began to increase beginning in May and continuing into early June 2008. Like the North to Houston CSC, the North to South CSC also experienced an increasing number of intervals in which the congestion on the interface could not be resolved, thereby producing maximum shadow prices on the constraint and associated high balancing energy market prices in the deficient zones (Houston and South). Figure 49 shows the actual flows versus the physical limit for the North to South CSC in 2008 during intervals when the CSC was binding.

Figure 49: Actual Flows versus Physical Limits during Congestion Intervals North to South



Historically, the inability to resolve a zonal constraint has been a relatively rare occurrence. In fact, excluding the North to Houston and North to South CSCs, the other CSCs together averaged only 15 intervals in 2008 with shadow prices that were greater than or equal to the current maximum CSC shadow price of $5,000 per MW. In contrast, the North to South and North to Houston CSCs experienced shadow prices greater than or equal to $5,000 per MW in 92 and 87 intervals, respectively.

The sharp increase in the frequency of occurrence of unresolved congestion on the North to Houston and North to South CSCs prompted the IMM, in consultation with ERCOT and the PUCT, to initiate in early May 2008 a detailed examination of ERCOT's congestion management procedures. This investigation quickly revealed that ERCOT rules permitted certain transmission elements to be managed with zonal balancing energy deployments when, in actuality, the congestion on these elements was neither effectively nor efficiently resolvable with zonal balancing energy deployments (the transmission elements that can be designated to be managed with zonal balancing energy deployments in the same manner as the CSC are referred to as "Closely Related Elements (CREs)" in the ERCOT Protocols).

Under the current zonal market model, transmission congestion is resolved through a bifurcated process that consists of either (1) zonal balancing energy deployments, or (2) local, unit-specific deployments. Because the pricing and incentives for both load and resources are better aligned with zonal congestion management techniques under the zonal market model, it is preferable to manage transmission congestion using zonal balancing energy to the extent it is effective and efficient.

However, for the CREs in question related to the North to Houston and North to South CSCs, zonal balancing energy deployments were neither effective nor efficient in resolving the transmission congestion. The result was an increasing frequency of the deployment of substantial quantities of energy in both the Houston and South Zones up to the point of exhaustion, thereby triggering the maximum shadow prices for these CSCs and associated high balancing energy prices in the South and Houston Zones that approached or even exceeded the system-wide offer cap of $2,250 per MWh.

To address these market and reliability issues, in late May 2008, again in consultation with ERCOT and the PUCT, the IMM submitted Protocol Revision Request ("PRR") 764, which improved the definition of those transmission elements eligible to be designated as CREs. PRR 764 was processed through the ERCOT committees on an expedited basis and was implemented on June 9, 2008.

While PRR 764 effectively resolved the issues encountered in April through June 2008 within the context of the zonal market model framework, the implementation of the nodal market will eliminate the current bifurcated congestion management process by providing simultaneous, unit-specific solutions that will always present the most effective and efficient congestion management alternatives to the system operator. As a point of comparison, Figure 50 shows the pricing implications of unresolved North to South congestion under the zonal model, and Figure 51 shows the pricing implications of the same unresolved constraint under a nodal market model.

Figure 50: Pricing Contours of Unresolved Congestion in the Zonal Market

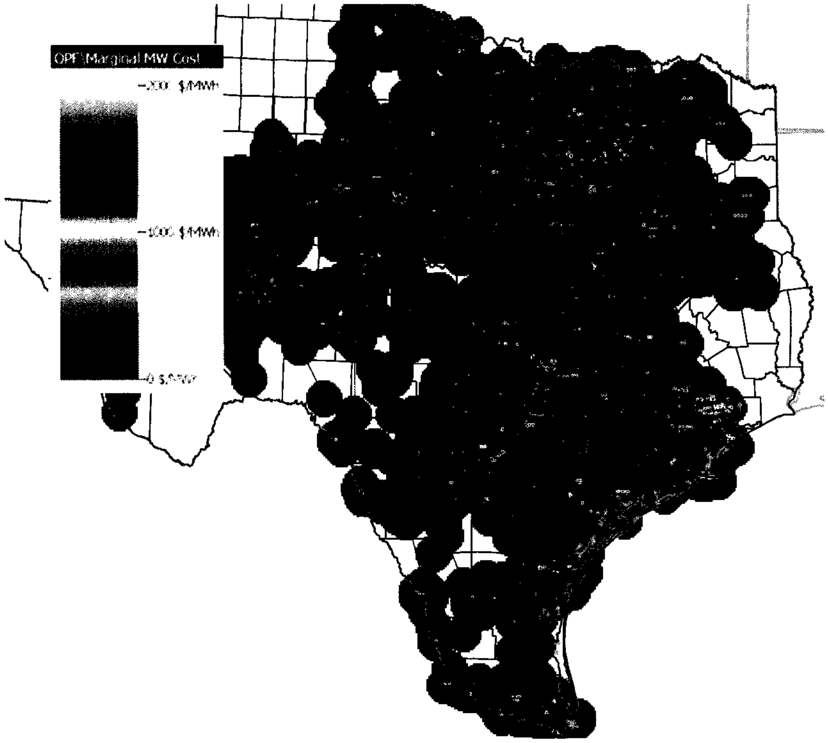
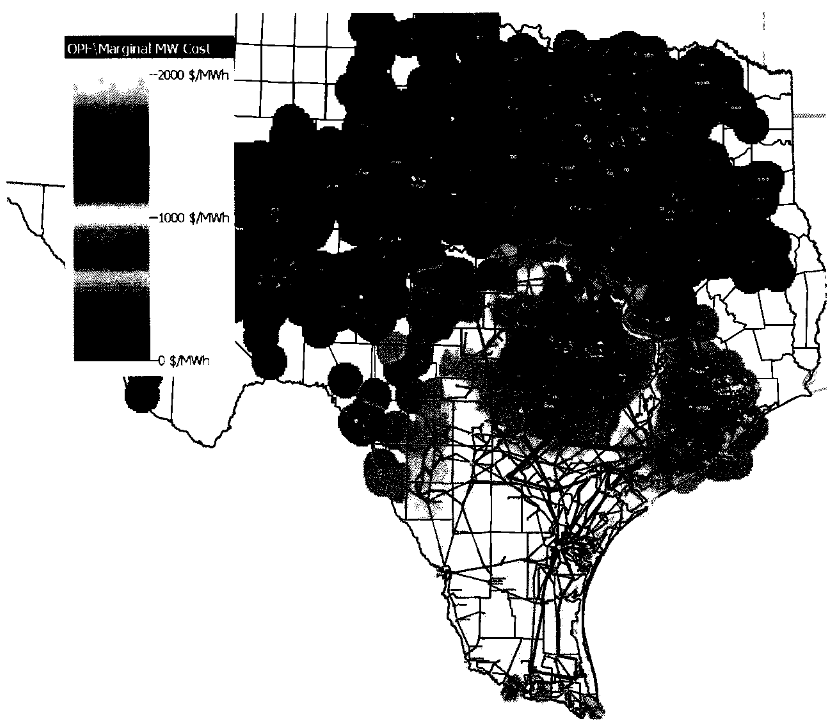


Figure 51: Pricing Contours of Unresolved Congestion in the Nodal Market



These graphics show that an unresolved constraint produces extremely high market clearing prices that are very widespread under the zonal model. In contrast, for the same unresolved constraint, the resulting high prices remain much more localized under the nodal model. These differences in pricing outcomes are due to the use of zonal average shift factors under the zonal model, as previously discussed, compared to the use of location-specific shift factors under the nodal model. Further, because the nodal market employs unit-specific offers and dispatch, the control of power flows on the system is much more flexible and precise than under the zonal model. Hence, it is much less likely under nodal dispatch to even encounter unresolvable constraints such as those experienced on the North to South and North to Houston interfaces in 2008.

In consideration of these differences, we have estimated the benefits that the nodal market would have produced by allowing more efficient resolution of the congestion on the North to South and North to Houston interfaces in 2008. To produce this estimate we assume that a nodal market model would not have eliminated the existence of congestion, but would have been able to resolve the congestion more efficiently without the frequent exhaustion of all available resources. Additionally, consistent with the results shown in Figure 50 and Figure 51, we recognize that the distribution of extremely high market-clearing prices due to congestion is much more localized under the nodal model than the widespread distribution under the zonal model, thereby affecting the average load zone price to a lesser degree under the nodal model. To account for these factors, in each interval that either the North to Houston or the North to South CSC was binding, we limited the price in the import-constrained zones (Houston and South) to a maximum of 20 times the price of natural gas and recalculated the annual average balancing energy market price for the Houston and South Zones in 2008.

This analysis indicates that the annual average balancing energy market price in the Houston and South Zones would have been reduced by approximately $10.42 per MWh under these assumptions. With approximately 168 million MWh consumed in the Houston and South Zones in 2008, the implied value of this reduction is approximately $1.75 billion in 2008. However, because of existing contracts and the fact that the 2008 congestion excursions were contained to a relatively short period of time and not expected to recur in the future, not all customers were directly affected by these wholesale price increases. Assuming that only 5 to 10 percent of customers in the South and Houston Zones were directly affected by the significant price increases in the balancing energy market and short-term bilateral markets associated with the North to Houston and North to South congestion, this analysis indicates that the efficiencies of the nodal market, had it been in place, could have reduced the annual costs for customers by $87 to $175 million in 2008. This analysis estimates only the savings that could have occurred through more efficient congestion management during the periods of acute North to Houston and North to South congestion, and does not include the benefits that the nodal market will provide more generally with respect to congestion management and other dispatch efficiency improvements.

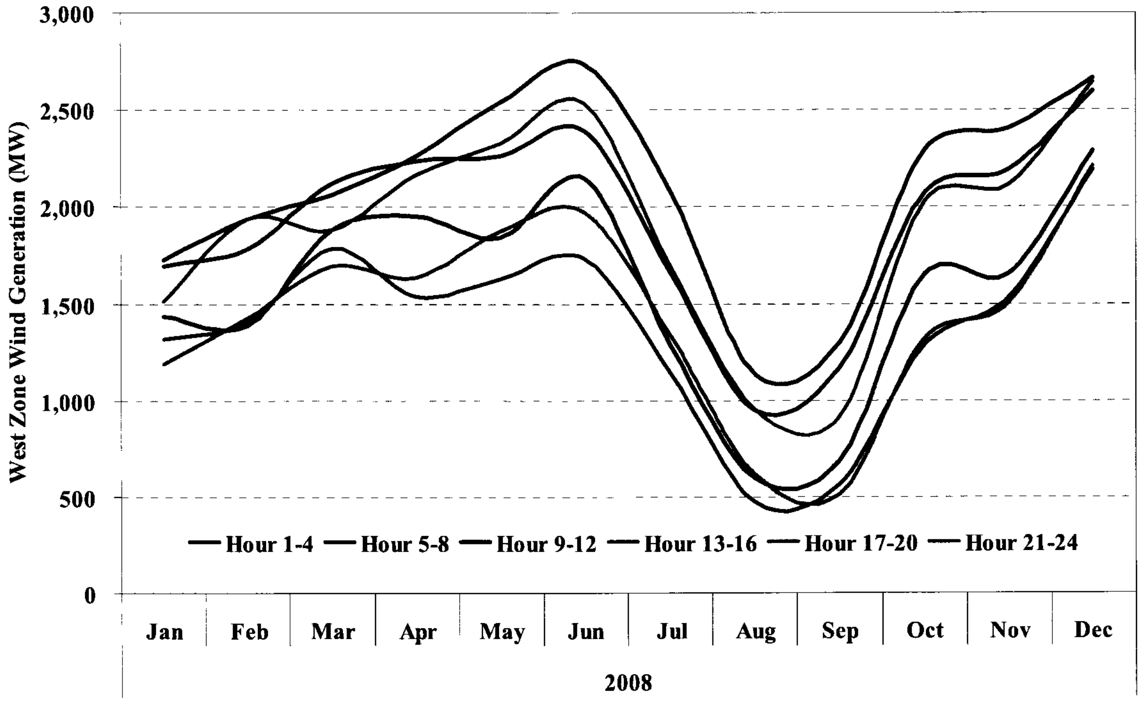
2. Congestion on West to North CSC

In 2008 the West to North CSC was binding in 5,320 intervals (15 percent). This was more frequent than any other CSC in 2008 and 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 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. The installed wind capacity in ERCOT grew from approximately 4.5 GW at the beginning of the year to approximately 8.1 GW by December 2008, with more than 90 percent of wind capacity located in the West Zone.

Average load in the West Zone was 2,547 MW in 2008, with a minimum of 1,828 MW and a maximum of 3,910 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 52 shows the average West Zone wind production for each month in 2008, with the average production in each month shown separately in four hour blocks.[20]

20 Figure 52 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 52 are lower than the production levels that would have materialized absent transmission constraints.

Figure 52: Average West Zone Wind Production



Depending on load levels, transmission system topology and other system operating conditions, up to 4 GW of wind generation in the West Zone can be reliably produced before encountering a transmission export constraint on the West to North CSC. Figure 53 shows the actual flows and the physical limit for the West to North CSC in 2008 for intervals in which the CSC was binding.

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

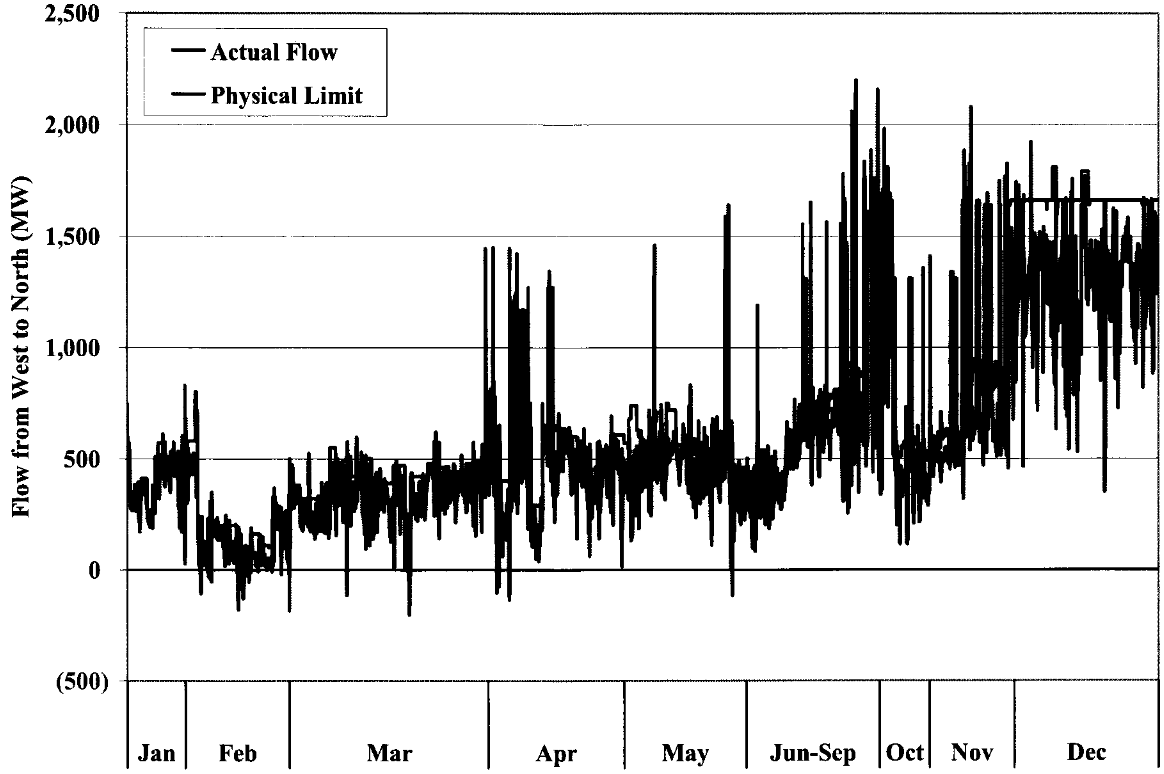


Figure 53 shows that over one-half of the binding intervals occurred in January through May 2008, with relatively fewer binding intervals in the summer months, and congestion activity increasing again in October through December.[21] Hence, although Figure 52 shows actual wind production in January through June that was comparable or lower than October through December, West to North congestion activity was more frequent in the first half of the year.

This higher frequency of congestion on the West to North CSC in the early part of the year was affected by two factors. First, the West to North CSC was subject to similar CRE issues previously described for the North to South and North to Houston CSC during this timeframe, although to a much lower degree. While the majority of the congestion on the North to South and North to Houston CSCs was related to constraints that were not included as CREs with the implementation of PRR 764, the majority of the congestion on the West to North CSC in the first

21 The West to North CSC consists of thermal transfer limits and a dynamic stability limit. Both are represented in Figure 53. Generally, in intervals with a physical limit less than 1,000 MW, the thermal limit was binding, and in intervals with a physical limit greater than 1,000 MW, the dynamic stability limit was binding.

part of the year was related to constraints that remained as CREs following the implementation of PRR 764. The second and more significant explanation for the increased congestion on the West to North CSC in the first part of the year is that a major double-circuit 345 kV transmission line was out of service for maintenance for most of February and March, thereby significantly reducing the West to North transfer capability. However, even with the changes implemented with PRR 764 and most major transmission lines in service, Figure 53 shows that congestion activity increased again in October through December 2008, and that this increase in congestion activity corresponds to the increase in wind production shown in Figure 52 for these months.

Although plans exist through the Competitive Renewable Energy Zone ("CREZ") project to significantly increase the transmission export capability from the West Zone, it is likely given the current transmission infrastructure and the level of existing wind facilities in the West Zone that the quantity of wind production that can be reliably accommodated in the West Zone will continue to be significantly limited for several years until such transmission improvements can be completed.

3. Congestion on North to West CSC

The North to West CSC was binding in 278 intervals (0.8 percent) in 2008. The North to West CSC was primarily binding under system conditions where wind generation in the West Zone is very low and few of the thermal generating units in the West Zone are online. Figure 54 shows the actual flows and the physical limit for the North to West CSC in 2008 for intervals in which the CSC was binding.

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

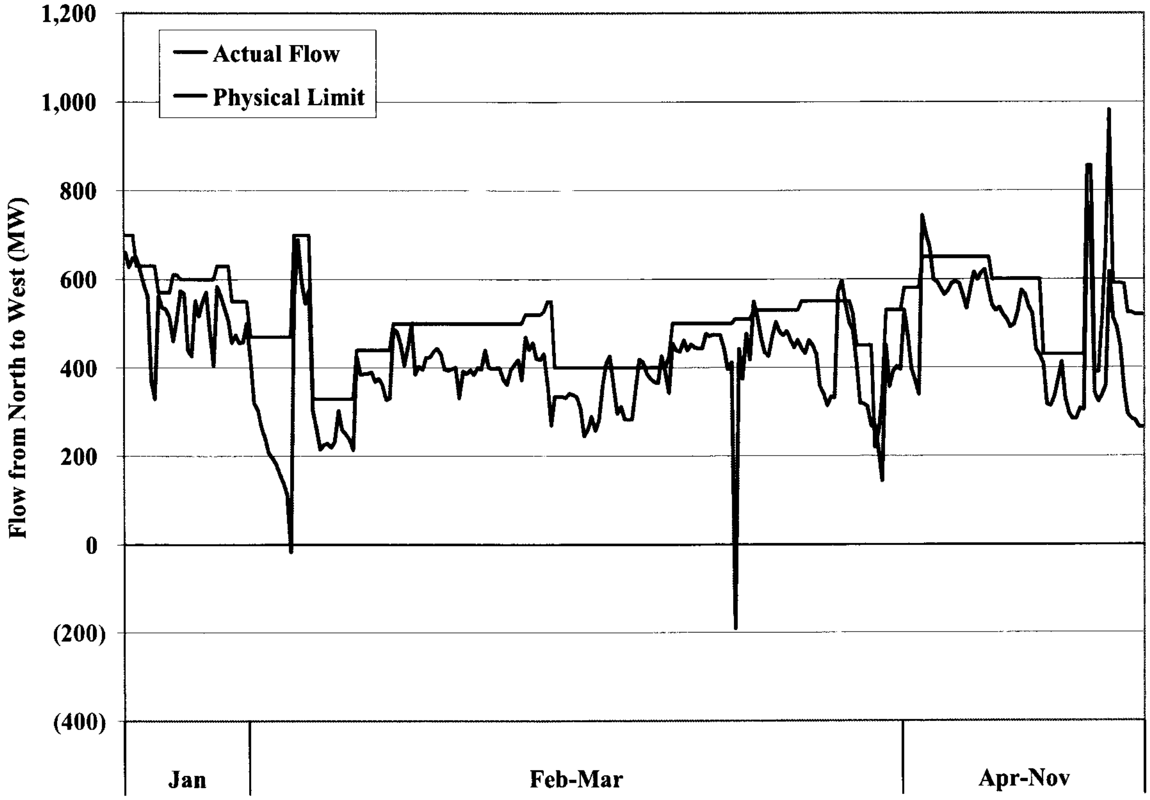


Figure 54 shows that the majority of the binding intervals on the North to West CSC occurred during the first quarter of 2008. The transmission outages discussed in relation to the West to North CSC also affected the import capability for North to West in the February and March timeframe. Additionally, as with the North to Houston and North to South CSCs, some of congestion on the North to West CSC was related to constraints that were removed as eligible zonal constraints with the implementation of PRR 764 in June 2008, although the magnitude of congestion on the North to West CSC associated with these issues was much less than for the North to Houston and North to South CSCs.

4. Congestion on South to North CSC

The South to North CSC was binding in 571 intervals (1.6 percent) in 2008. Figure 55 shows the actual flows and the physical limit for the South to North CSC in 2008 for intervals in which the CSC was binding

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

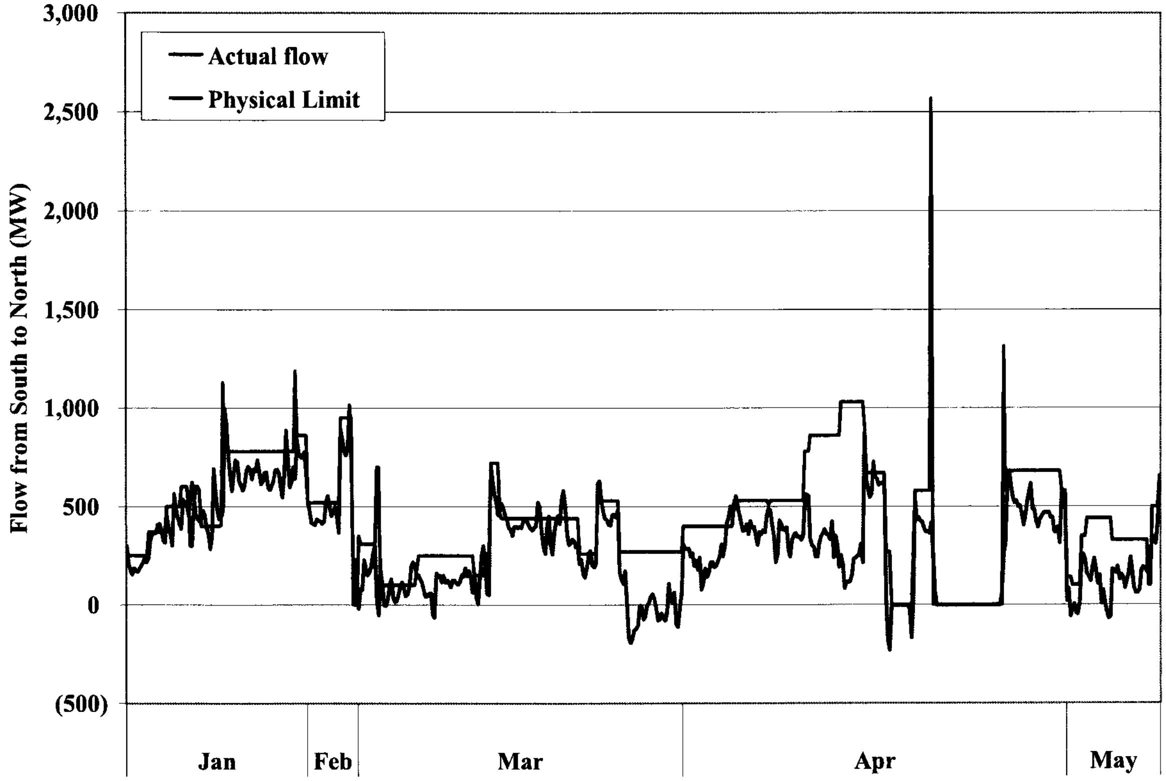


Figure 55 shows that all of the binding intervals on the South to North CSC occurred during the months of January through May 2008. For the South to North CSC, practically all of the congestion during this timeframe was related to constraints that were removed as eligible zonal constraints with the implementation of PRR 764 in June 2008.

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 2008.

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

CSC 2008	Average Physical Limit (MW)	Average Actual Flow (MW)	Avg. Physical Limit - Avg. Actual Flow (MW)
North to South	234	139	94
North to Houston	1,948	1,655	293
South to North	444	297	148
West to North	733	568	165
North to West	516	419	97

Table 4 shows that, for all CSCs in 2008, the average actual flow was considerably less than the average physical limit. For all CSCs combined, the average actual flow was 21 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.

The nodal market will provide many improvements, including unit-specific offers and shift factors, 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 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 2008. Figure 56 shows the average number of TCRs and PCRs awarded for each of the CSCs in 2008 compared to the average SPD-modeled flows during the constrained intervals.

Figure 56: Transmission Rights vs. Real-Time SPD-Calculated Flows Constrained Intervals - 2008

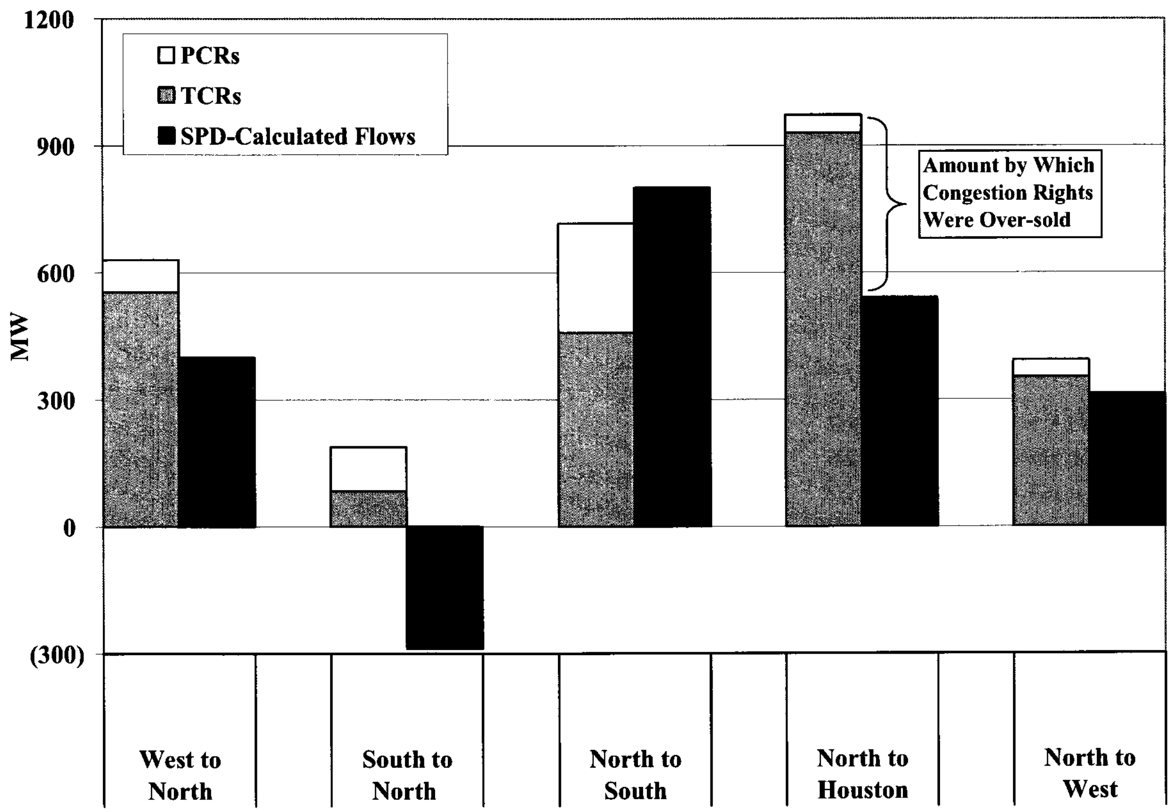


Figure 56 shows that total congestion rights (the sum of PCRs and TCRs) on all the interfaces exceeded the average real-time SPD-calculated flows during constrained intervals except for the North to South CSC. These results indicate that the congestion rights were oversold in relation to the SPD-calculated limits for most CSCs, even though fewer TCRs were awarded in 2008 than in 2007. For example, congestion rights for the North to Houston CSC were oversold by an average of 432 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 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.

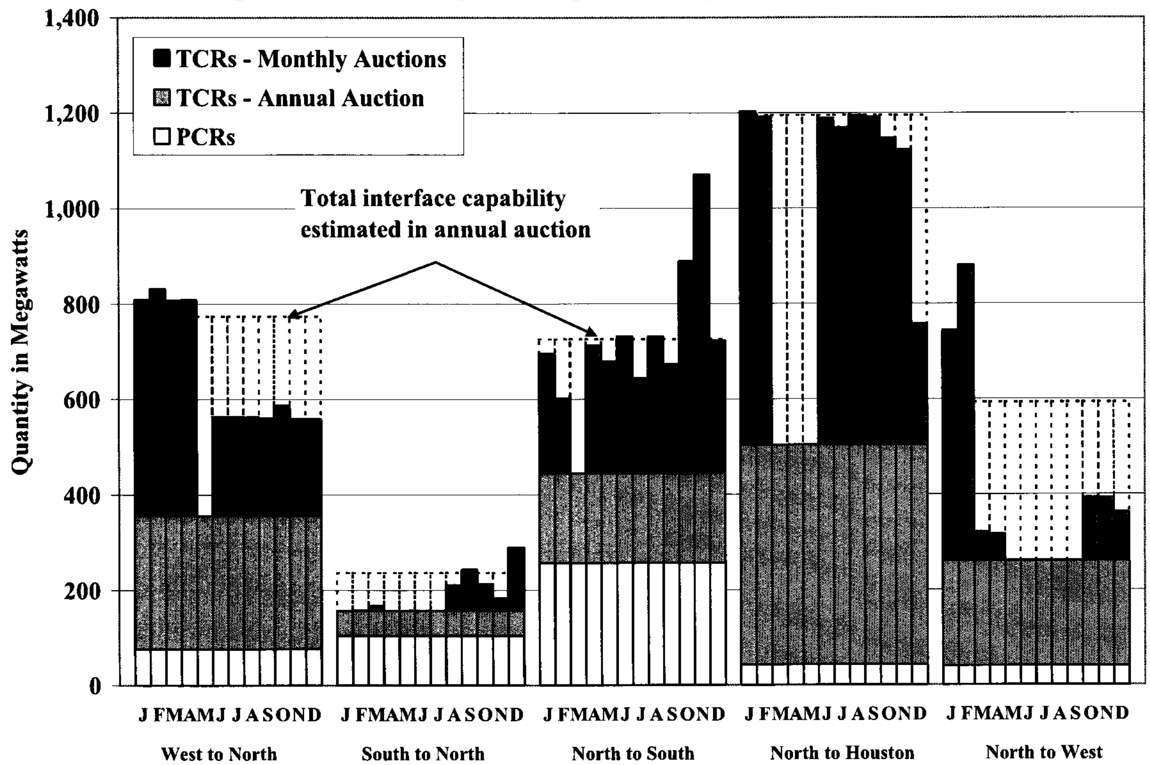
As an example, suppose the SPD-calculated flow limit is 300 MW for a particular CSC during a constrained interval and that holders of congestion rights own a total of 800 MW over the CSC. ERCOT will receive congestion rents from the balancing energy market to cover precisely 300 MW of the 800 MW worth 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 that cause the actual physical flow to be significantly different from the SPD modeled flow. As discussed above, ERCOT operators may need to respond by lowering the SPD-modeled flow limits to manage the actual physical flow. Accordingly, it is likely that the quantity of congestion rights 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 57 shows the quantity of each category of congestion rights for each month during 2008. The quantities of PCRs and annual TCRs are constant across all months and were determined before the beginning of 2008, while monthly TCR quantities can be adjusted monthly.

Figure 57: Quantity of Congestion Rights Sold by Type

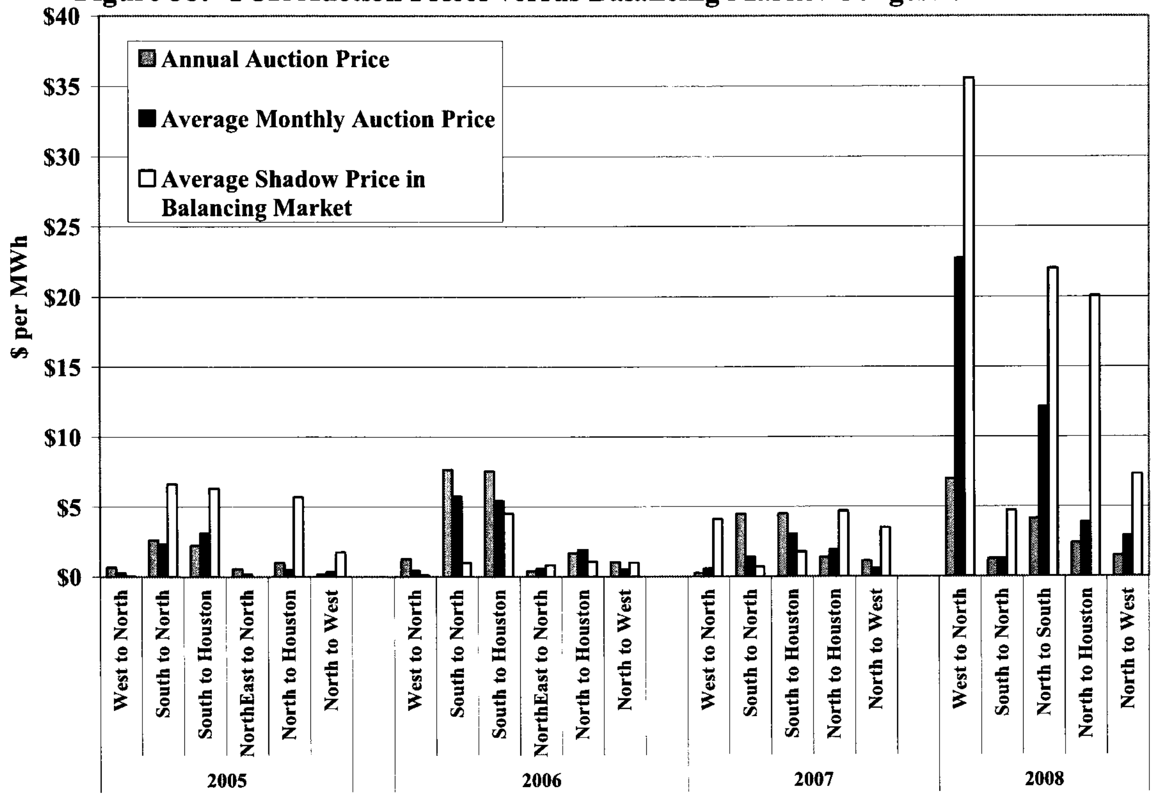


When the monthly planning studies indicate changes from the summer study, revisions are often made to the estimated transmission capability. Therefore, the auctioned congestion rights may increase or decrease relative to the amount estimated in the summer study. The shadow boxes in the figure represent the capability estimated in the summer study that is not ultimately sold in the monthly auction. When there is no shadow box in Figure 57, 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, North to West 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 and North to West 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 58 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.

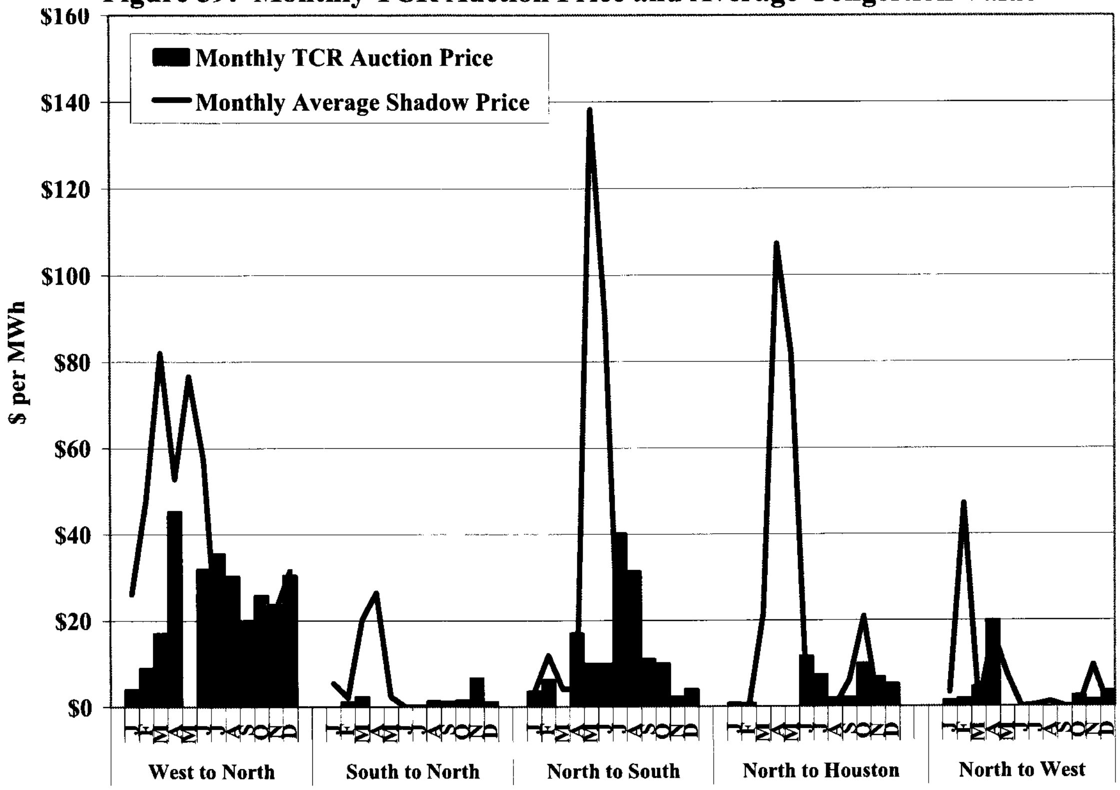
Figure 58: TCR Auction Prices versus Balancing Market Congestion Prices



This figure shows that the TCR annual auction prices were significantly lower than the value of congestion in real-time in 2008. This suggests that participants are not able to forecast annual interzonal congestion costs and accurately value the TCRs in the annual auction, and instead rely more upon historical market outcomes. Monthly TCR auction prices were more consistent with real-time congestion prices; however, real-time shadow price exceeded the average monthly auction price for all the five CSCs in 2008.

Figure 59 compares monthly TCR auction prices with monthly average real-time CSC shadow prices from SPD for 2008. The TCR auction prices are expressed in dollars per MWh.

Figure 59: Monthly TCR Auction Price and Average Congestion Value



The significant divergence in the monthly TCR auction prices and the real-time shadow prices indicates that market participants did a poor job predicting and valuating the real-time cost of zonal congestion in 2008. These outcomes appear to reflect the existence of flaws in the congestion management procedures that were not recognized until ERCOT had difficulty managing interzonal congestion in the spring of 2008. Just as ERCOT was unaware of these flaws, market participants could not anticipate that problems with the congestion management system would lead to the high shadow prices that occurred in April through June 2008.

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.[22] 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 59 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.

22 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 60: TCR Auction Revenues, Credit Payments, and Congestion Rent

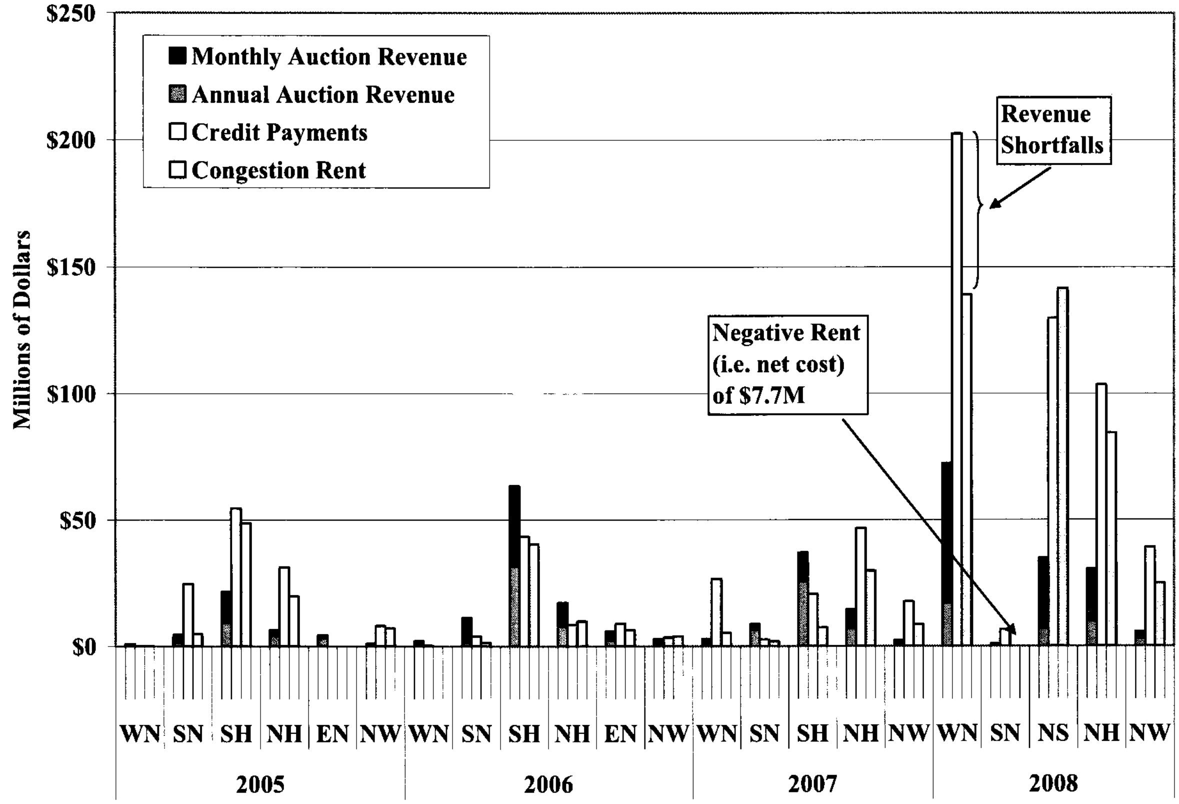


Figure 60 shows that the West to North, North to Houston and North to West CSCs each had significant revenue shortfalls in 2008. When congestion rents fall significantly below payments to TCR holders, it implies that the SPD-calculated flows across constrained interfaces have been systematically lower than the amount of TCRs sold for the interfaces. In contrast, the North to South CSC produced a slight surplus in 2008 (*i.e.*, congestion rents exceeded TCR credit payments), which indicates that the North to South CSC had SPD-calculated flows that were generally greater than the amount of TCRs sold. The South to North CSC actually experienced a negative $7.7 million congestion rent in 2008, which not only produces a revenue shortfall, but also shows that the average SPD-calculated flow across this constraint when it was binding was negative when the actual flows were positive, which is a product of the issues associated with the use of zonal average shift factors as described in the first part of this Section.

Figure 60 also shows that payments to TCR holders have consistently exceeded the congestion rents that have been collected from the balancing market since the creation of the TCR market. Congestion rents covered 68, 90 and 47 percent of payments to TCR holders in 2005, 2006 and

2007, respectively. Congestion rents covered 79 percent of the payments to TCR holders in 2008, with an annual net revenue shortfall of $99 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 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.

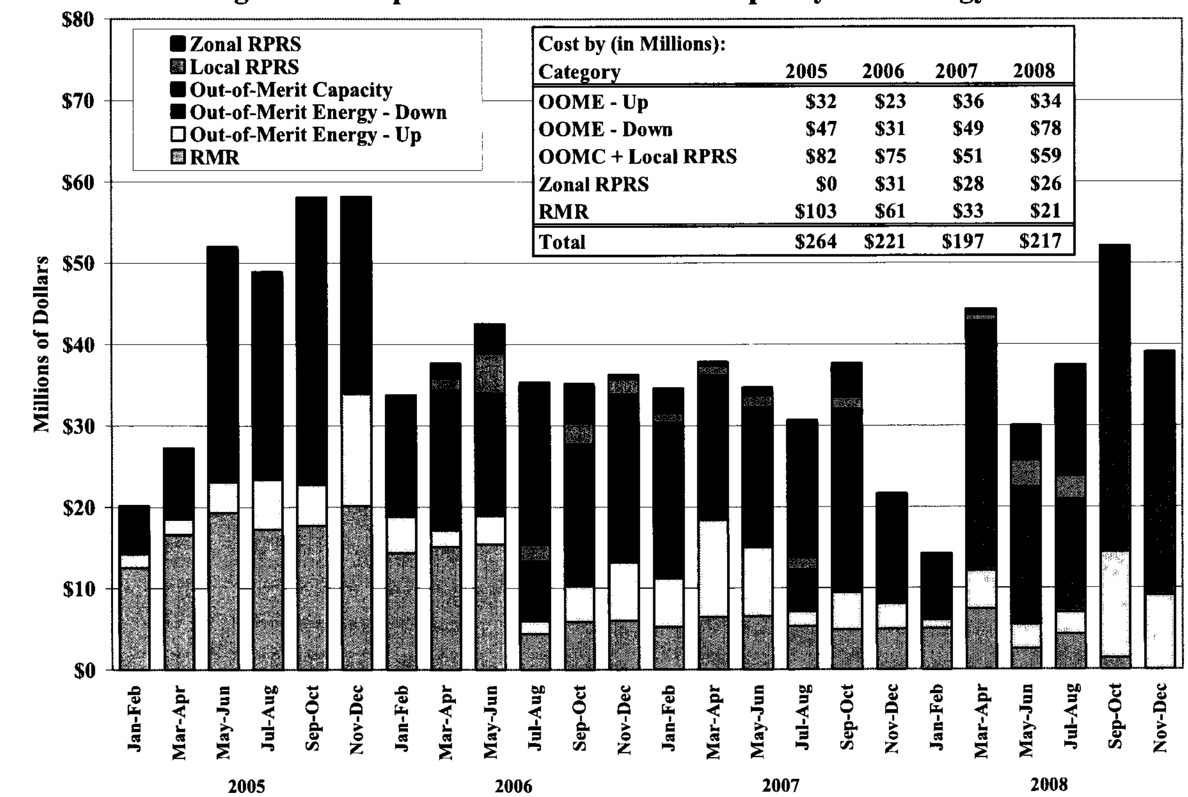
When a unit is dispatched out of merit (OOME up or OOME down), the unit is paid for a quantity equal to the difference between the scheduled output based on the unit's resource plan and the actual output resulting from the OOME instruction from ERCOT. The payment per MWh for OOME is a pre-determined 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. However, as a part of the RMR exit strategy process, all units were removed from RMR status by October 2008. Units contracted to provide RMR service to ERCOT are compensated for start-up costs, energy costs, and are also paid a standby fee. Figure 61 shows each of the four categories of uplift costs from 2005 to 2008.

Figure 61: Expenses for Out-of-Merit Capacity and Energy



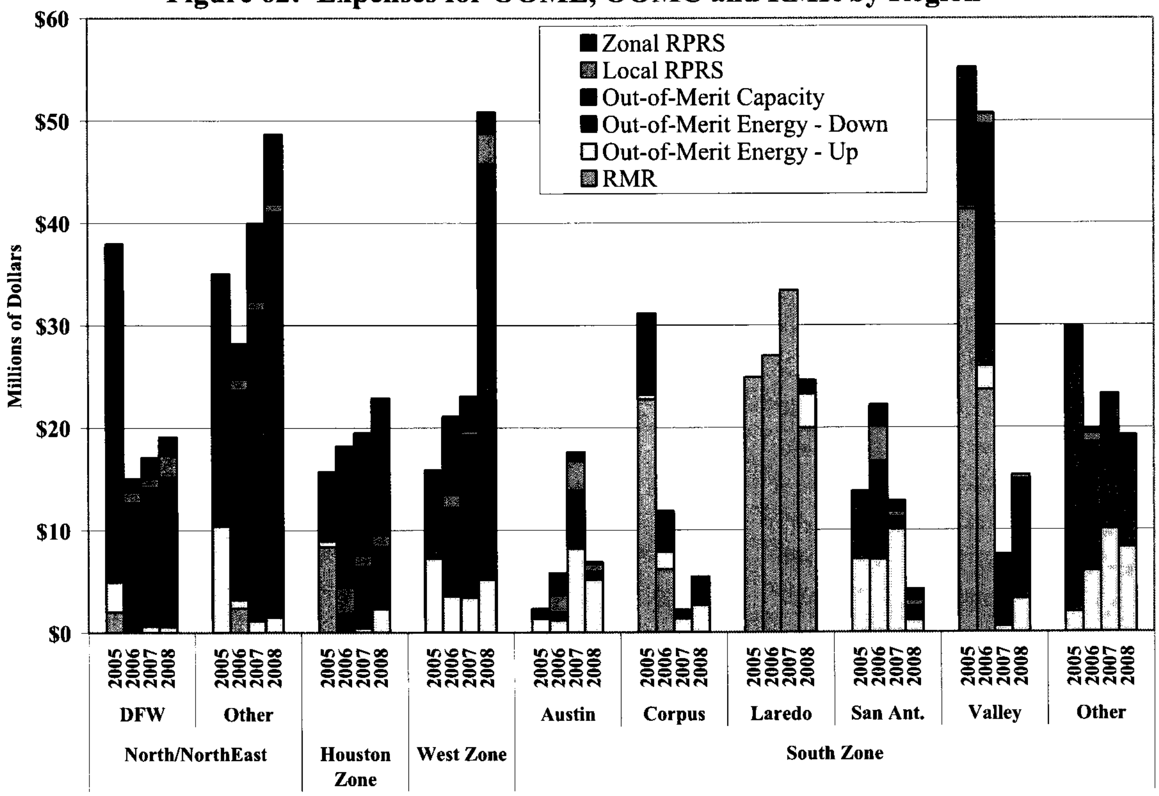
The results in Figure 61 show that overall uplift costs for RMR units, OOME units, OOMC/Local RPRS and Zonal RPRS[23] units were $217 million in 2008, which is a $20 million increase over the $197 million in 2007. OOME Down costs accounted for the most significant portion of the change in 2008, increasing from $49 million in 2007 to $78 million in 2008.

[23] 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. The RPRS procurement tool was not in production in 2005, thus all capacity procurements were conducted via OOMC in 2005.

OOMC/Local RPRS costs increased from $50 million in 2007 to $60 million in 2008, and RMR costs decreased from $33 million in 2007 to $20 million in 2008. Figure 61 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 reliability. The rest of the analyses in this section evaluate in more detail where these costs were caused and how they have changed between 2005 and 2008. Figure 62 shows these payments by location.

Figure 62: Expenses for OOME, OOMC and RMR by Region



The most significant changes in 2008 compared to 2007 (*i.e.*, an increase or decrease of more than $5 million in a category by location) shown in Figure 62 are as follows:

- OOME Down costs in the West Zone increased by $21 million in 2008. This increase was associated with the significant addition of wind capacity in the West Zone.
- OOMC and OOME Down costs in the North Zone each increased by $5 million in 2008. These increases were associated with local reliability requirements near the load centers