2008 Assessment of the Electricity Markets in New England

By:

POTOMAC ECONOMICS

David B. Patton, Ph.D.
Pallas LeeVanSchaick, Ph.D.
Jie Chen, Ph.D.

Independent Market Monitoring Unit
for
ISO New England Inc.

June 2009

Table of Contents

List of Figures

List of Tables

Guide to Abbreviations

ASM	Ancillary Services Market
ERS	External Reserve Support
FCA	Forward Capacity Auction
FCM	Forward Capacity Market
FTR	Financial Transmission Rights
GW	Gigawatt (1 GW = 1,000 MW)
HHI	Herfindahl-Hirschman Index, a standard measure of market concentration
ISO	Independent System Operator
LFRM	Locational Forward Reserve Market
LMP	Locational Marginal Price
LOC	Lost Opportunity Cost, a component of the regulation price
MMbtu	Million British Thermal Units, a measure of energy content in natural gas
MMU	Market Monitoring Unit
MW	Megawatt
MWh	Amount of energy equal to producing 1 MW for a duration of one hour
NCPC	Net Commitment Period Compensation
NEMA	North East Massachusetts
NERC	North American Electric Reliability Corporation
NPCC	Northeast Power Coordinating Council, Inc.
RAA	Reserve Adequacy Assessment
RCP	Regulation Clearing Price
RCPF	Reserve Constraint Penalty Factors
RMR	Reliability Must-Run
RTO	Regional Transmission Organization
SEMA	South East Massachusetts
SCR	Special Constraint Resources
SMD	Standard Market Design
TMNSR	Ten-minute non-spinning reserves
TMOR	Thirty-minute operating reserves
TMSR	Ten-minute spinning reserves
UDS	Real-time dispatch software

Preface

Potomac Economics serves as the Independent Market Monitoring Unit for ISO New England. In this role, we are responsible for evaluating the competitive performance, design, and operation of the wholesale electricity markets operated by ISO New England. In this Annual Assessment, we provide our annual evaluation of the ISO's markets for 2008 and our recommendations for future improvements. This report complements the State of the Market Report produced by the Internal Market Monitoring Unit, which provides its evaluation of the market outcomes in 2008.

We wish to express our appreciation to the Internal Market Monitor and other staff of the ISO for providing the data and information necessary to produce this report.

I. Executive Summary

This report assesses the efficiency and competitiveness of New England's wholesale electricity markets in 2008. Since ISO New England began operations in 1999, it has made significant enhancements to the energy market and introduced markets for other products that have improved overall efficiency. ISO New England's markets currently include:

- *Day-ahead and real-time energy*, which coordinate commitment and production from the region's generation and demand resources, and facilitate wholesale energy trading;
- *Financial Transmission Rights ("FTRs")*, which allow participants to hedge the congestion costs associated with delivering power to a location that is constrained by the limits of the transmission network;
- *Forward and real-time operating reserves*, which are intended to ensure that sufficient resources are available to satisfy demand when an outage or other contingency occurs;
- *Regulation*, which allows the ISO to instruct specific generators to increase or decrease output moment-by-moment to keep system supply and demand in balance; and
- *Forward Capacity Market ("FCM")*, which is intended to provide efficient long-term market signals to govern decisions to invest in new generation and demand resources and to maintain existing resources.

These markets provide substantial benefits to the region by ensuring that the lowest cost supplies are used to meet demand in the short-term and by establishing transparent, efficient price signals that govern investment and retirement decisions in the long-term. Although it is difficult to quantify the benefits that result from market coordination, good coordination is essential due to the physical characteristics of electricity and the transmission network used to deliver it to customers. This coordination affects not only the prices and production costs of electricity, but also the reliability with which it is delivered.

A. Introduction and Summary of Findings

In addition to providing a summary of market outcomes in 2008, this report includes findings in two primary areas: the competitive performance of the market and the operational efficiency of the market. The broad findings in each of these areas are discussed below.

Competitive Performance of the Market

Based on our evaluation of the markets in New England (in both constrained areas and the broader market), we find that the markets performed competitively in 2008. Our analysis raised no competitive concerns that suppliers withheld resources to raise prices in the New England markets. Although energy prices increased in 2008, this was due primarily to increases in fuel prices. Because fuel costs constitute the vast majority of the marginal costs of generation, higher fuel costs translate to higher offer prices and market clearing prices.

However, we find that frequent supplemental commitment encouraged some generators to raise offers above competitive levels (i.e., above marginal cost). Generators committed for local reliability often do not face meaningful competition and may have local market power.[1] The market power mitigation measures have generally limited the ability of suppliers to exercise market power.[2] However, due to the high frequency of some local reliability commitments, the mitigation measures were not fully effective at addressing certain conduct. In particular, conduct by a large supplier in Boston resulted in higher Net Commitment Period Compensation ("NCPC") payments until April 2008, when the ISO revised the reliability requirements that had previously necessitated frequent local reliability commitments.[3] To address similar issues in the future, we worked with the Internal MMU to file proposed changes to the mitigation rules with the Federal Energy Regulatory Commission.

Operational Efficiency of the Markets

In general, the day-ahead and real-time markets operated efficiently in 2008 with prices that reflected underlying market fundamentals. Electricity prices in New England have been strongly

---

1 When local reliability requirements are satisfied outside the normal market process, the suppliers are generally paid their offer price. This gives them incentives to submit offers above their marginal costs (i.e., "pay-as-bid" incentives), even when they face competition. Hence, it can be difficult to distinguish economic withholding from a competitive outcome when suppliers have pay-as-bid incentives.

2 Market power mitigation measures are contained in ISO New England's Tariff. They address potential abuses of market power by allowing the ISO to modify suppliers' offers when certain criteria are satisfied.

3 NCPC payments are made to ensure a supplier recovers its full offer costs.

correlated with changes in underlying fuel prices, as one would expect in a well-functioning market. To maintain reliability in constrained areas, the ISO has taken appropriate actions to supplement the commitments from the day-ahead market with additional resources. However, these supplemental commitments tend to diminish energy and ancillary services price signals in constrained areas and increase NCPC costs, which are difficult for load-serving entities to hedge. This issue is common to all electricity markets and arises when the markets do not fully satisfy the reliability needs of the system. Additionally, the limited number of fast-start units in New England increases the need to commit larger, slower-starting generation to assure reliability in constrained areas. Several changes have been implemented or are planned that significantly reduce the need for supplemental commitments, including:

- ISO New England was able to revise its reliability requirements for Boston-area voltage support following transmission upgrades in 2007. This led to a dramatic reduction in supplemental commitment in Boston.
- New transmission investment into Southwest Connecticut reduced supplemental commitment in Connecticut.
- Planned transmission investment in Southeast Massachusetts is expected to address the cause of most of the remaining supplemental commitment for local reliability.

These changes reduce the costs of the local requirements that exist in the energy, operating reserve, and capacity markets. The fact that all of these markets now have local requirements is important because markets with local requirements reduce the need for manual actions by operators, lower uplift costs, and improve economic signals. The improved economic signals should reduce New England's heavy reliance on reliability agreements (used to ensure that units needed for reliability remain in operation). Reliability agreements are poor substitutes for transparent market prices and do little to facilitate efficient investment.

In addition, ISO New England implemented a forward capacity market that procures capacity three years forward on a locational basis. The FCM is intended to facilitate the entry of new supply and demand resources. The first two auctions were conducted successfully in 2008 to meet the capacity requirements from June 2010 to May 2011 and from June 2011 to May 2012.

Recommendations

Overall, we conclude that the markets performed competitively in 2008 and were operated well by the ISO. Based on the results of our assessment, however, we offer some recommendations to further improve the performance of the New England markets. They are listed in a table at the end of this executive summary. The following sections summarize the findings of the report.

B. Energy Prices and Congestion

Although electricity prices increased by more than 18 percent in 2008, we conclude that the market performed competitively.[4] The increase in electricity prices was primarily due to increases in natural gas and oil prices. Correlation between natural gas prices and electricity prices is consistent with a well-performing market, because fuel costs constitute the vast majority of most generators' marginal costs and natural gas-fired units frequently set the market price in New England.

There were no significant price spikes or capacity deficiencies in 2008 as peak demand levels were considerably lower than expectations and the system was operated effectively. The peak load was 26.1 GW in 2008, substantially lower than the summer peak load forecast of 28.0 GW and the all-time peak load of 28.1 GW, which occurred in 2006.

Congestion and Financial Transmission Rights

Under SMD, New England has experienced relatively little congestion in historically-constrained areas such as Boston and Connecticut. In fact, a large portion of the price separation between net exporting regions and net importing regions has been due to transmission losses, rather than transmission congestion. In 2008, the Lower Southeast Massachusetts area ("Lower SEMA") had the most congestion of any area in New England. Energy prices averaged $10 per MWh higher in Lower SEMA than at the New England Hub in the day-ahead market. This congestion is due primarily to the relatively high production costs of generation in Lower SEMA. New transmission should reduce congestion and the other costs to maintain reliability in Lower SEMA.

4 These are load-weighted average prices at the New England Hub.

Congestion into and within Connecticut was relatively modest in 2008. The average congestion price difference between the New England Hub and Norwalk-Stamford (the most import-constrained area in Connecticut) was just $5 per MWh in 2008. The low levels of congestion have resulted primarily from transmission additions under Phase I and Phase II of the Southwest Connecticut 345 kV Transmission Project in recent years.

The ISO operates annual and monthly markets for FTRs.[5] FTRs are invaluable in a locational energy market because they allow participants to hedge the congestion and associated basis risk on the network. Since FTR auctions are forward financial markets, efficient FTR prices should reflect the expectations of market participants regarding congestion in the day-ahead market. Our analysis of FTR prices indicates:

- FTR prices were generally consistent with the congestion that prevailed in the energy markets in 2008, which suggests that the markets have sufficient liquidity to produce efficient FTR prices.
- The consistency of FTR prices and congestion improved from the annual auction to the monthly auctions. This result is expected because participants gain additional information about market and system conditions after the annual auction.

Although congestion was relatively mild in 2008, the ISO collected a total of $121 million in congestion revenue from the day-ahead and real-time markets, most of which was paid out to the holders of FTRs.

Day-Ahead to Real-Time Price Convergence

We evaluated price convergence at the New England Hub, which is broadly representative of prices outside of transmission-constrained areas. The differences between average day-ahead and average real-time prices were modest, indicating good overall convergence. This is important because day-ahead prices that accurately reflect expected real-time conditions facilitate efficient commitment of generating resources. However, prices were not well-arbitraged at one import-constrained location (Lower SEMA) due to the higher price volatility and supplemental commitment in the area.

5 FTRs entitle the holder to the congestion price difference between the FTR's sink and source in the day-ahead market (i.e., the congestion price at the sink minus the congestion price at the source).

C. External Interface Scheduling

In this report, we evaluated transaction scheduling between New England and the three adjacent regions: Quebec, New Brunswick and New York.

Quebec and New Brunswick Interfaces

Power is usually imported from Quebec and New Brunswick. Average net import levels range from 1,500 MW during peak hours to 1,000 MW during off-peak hours in 2008, which is consistent with the management of hydroelectric resources in Canada.

New York Interface

New England and New York are connected by one large interface between western New England and eastern Upstate New York, and by two small interfaces between Connecticut and Long Island. Exports are consistently scheduled from Connecticut to Long Island over the smaller interfaces (averaging 300 MW in 2008), while participants schedule power flows in both directions on the large interface between the markets depending on the relative prices. In 2008, an average of 220 MW was exported to New York during peak hours and 325 MW was imported from New York during off-peak hours.

Market participants should arbitrage the prices in the two areas by scheduling power from the low-priced market to the high-priced market. However, uncertainty and long scheduling lead times have prevented participants from fully utilizing the interfaces with New York. Explicit coordination of the physical interchange of power between the markets is likely needed to achieve efficient utilization of the interfaces between New York and New England. We employed simulations to estimate the benefits of optimal hourly scheduling of the primary interface between New England and New York in 2006 and 2007. The simulations indicated that consumers in New England would have saved $83 million over the two-year timeframe. However, these estimates will rise sharply in the future if the frequency of operating reserve shortages increases because full utilization of the interface can prevent such reserve shortages.

D. Reserve Markets

The ISO operates a forward reserve market where reserves are procured in seasonal auctions with locational requirements. It also operates a real-time reserve market where reserves are scheduled with local requirements and co-optimized with the real-time energy market. The reserve markets provide market-based mechanisms for the wholesale market to meet the reliability needs of the system, thereby reducing the need for out-of-market actions by the operators. This section summarizes our evaluation of the reserve markets.

Real-Time Reserve Market Results

By co-optimizing the scheduling of energy and reserves, the market is able to reflect in the clearing prices of both energy and reserves the redispatch costs that are incurred to maintain reserves. When available reserves are not sufficient to meet the required levels, the real-time model will be short of reserves and set the reserve clearing price at the level of the Reserve Constraint Penalty Factor ("RCPF"). This method of shortage pricing was recently endorsed in FERC Order No. 719.[6]

Reserve clearing prices were relatively low in the real-time market in 2008. Average clearing prices were less than $2 per MWh for all classes of reserves, with little variation by location. However, we also found that actions taken by the ISO to maintain local reserves are often more costly than the local RCPF, which indicates that the local RCPFs are set inefficiently low. Setting RCPFs at appropriate levels is important because:

- RCPFs contribute to setting prices in the reserve markets and the energy market when the reserve requirements cannot be met;
- RCPFs cause the market to utilize all available resources and reduce the need for market operators to take actions outside of the market process to maintain reliability.

Recognizing these concerns, the ISO has proposed to revise the local RCPFs to $250 per MWh. Our analysis indicates that this level would likely have been sufficient to maintain reserves in local areas.

6 *Wholesale Competition in Regions with Organized Electric Markets*, Order No. 719, 73 *Fed. Reg.* 64100 (October 28, 2008), FERC Stats. & Regs. ¶ 31,281 (2008) ("Order No. 719").

Finally, the report evaluates the designation of the local reserve zones, which initially were defined to include Boston, Southwest Connecticut, and Connecticut. There are five other local areas that require the commitment and dispatch of resources to meet local second contingency protection requirements.[7] Because they are not designated as local reserve zones, the ISO relies entirely on imported reserves to maintain reliability in these areas. Designating new reserve zones for these areas would: a) allow the model to satisfy the requirements with the least-cost mix of internal resources and imported reserves; and b) produce reserve clearing prices for the areas, which provide short-term and long-term price signals to prospective suppliers of reserves. However, the benefits from designating new reserve zones would be relatively modest under current market conditions due to the low frequency of transmission constraints to satisfy local reserve requirements, although the benefits would become more significant if the frequency of such constraints increases.

Forward Reserve Market Results

The Locational Forward Reserve Market ("LFRM") is a seasonal auction held twice a year where suppliers sell reserves that they are obligated to provide in real-time. LFRM obligations must be provided from an online resource with unused capacity or an offline resource capable of starting quickly (i.e., fast-start generators). The auction procures 10-minute non-spinning reserves ("TMNSR") for all of New England and 30-minute operating reserves ("TMOR") for All of New England, Boston, Connecticut, Southwest Connecticut, and Rest of System.

This report evaluates the results of recent forward reserve auctions and examines how suppliers satisfied their obligations in the real-time market. In the two Forward Reserve Auctions that were held in 2008, prices cleared at the $14 per kW-month price cap in Connecticut where the supply of fast-start generation was not sufficient to meet the local requirements. In Boston, TMNSR and TMOR cleared at $14 per kW-month in the Summer 2008 auction, while TMNSR cleared at $7.30 per kW-month and TMOR cleared at $5.55 per kW-month in the Winter 2008/09 auction. Prices cleared at $14 per kW-month in the Summer 2008 auction when the

7 These areas are Norwalk-Stamford, West Connecticut, Lower Southeast Massachusetts, Western New England, and Maine.

available supply of fast-start generation was not sufficient to meet the local requirement. However, Boston prices dropped considerably in subsequent auctions as the local requirement decreased to reflect the increased transmission capability into Boston. Outside the local areas, TMNSR cleared at $8.89 and $6.74 per kW-month, while TMOR cleared at $6.50 and $4.99 per kW-month.[8]

The fact that there is a single cap of $14 for all reserves products has raised the following potential incentive concerns.

- Suppliers with 10-minute reserve-capable units have the incentive to sell 30-minute reserves because there is no incremental revenue for selling higher-quality reserves.
- Likewise, suppliers with reserves in narrower constrained areas (e.g., Southwest Connecticut) have the incentive to sell their reserves in broader areas (e.g., Connecticut).

Both of these behaviors have been observed in the forward reserve markets. Modifying the price cap to differentiate the payment for higher quality reserves or reserves in more constrained areas would address these incentive issues.

E. Regulation Market

The regulation market performed competitively in 2008 with an average of approximately 700 MW of available supply competing to serve 120 MW of regulation demand.[9] The significant excess supply generally limits competitive concerns in the regulation market. However, regulation supply was sometimes tight in low-demand periods when many regulation-capable resources were offline, leading to transitory periods of high regulation prices.

Regulation market expenses increased from $44 million in 2007 to $51 million in 2008, partly due to the increase of natural gas prices. Natural gas-fired combined cycle generators, which

8 Forward reserve clearing prices are affected by the market rule that suppliers do not receive capacity payments for their forward reserve sales. In the summer of 2008, the capacity revenue was from Transition Payments of $3.75 per kW-month. Thus, a seller of TMOR outside the local areas in the Summer 2008 auction would receive $4.99 per kW-month, but also give up $3.75 per kW-month.

9 The average available supply is the average of offered regulation capabilities from committed resources in each hour.

provide most of the regulation service, are usually committed less frequently during periods of high gas prices. This reduces the availability of low-priced regulation offers and leads to higher regulation expenses.

Given the complex interaction of the regulation market with the energy market, significant benefits likely could be achieved by allowing the real-time market to co-optimize the scheduling of regulation with energy and operating reserves.

F. Real-Time Pricing and Market Performance

The goal of the real-time market is the efficient procurement of the resources required to satisfy the reliability needs of the system. To the extent that reliability needs are not fully satisfied by the market, the ISO must procure needed resources outside of the market, consistent with its operating procedures. However, these out-of-market actions tend to undermine the market prices because the prices will not fully reflect the reliability needs of the system. Efficient real-time prices are important because they encourage competitive scheduling by suppliers, participation by demand response, and investment in new resources when and where needed.

It is particularly important for the market to set efficient real-time price signals during shortages of operating reserves. In Order 719, the Commission recognized ISO New England's approach to shortage pricing as a model for other ISOs. ISO New England uses the operating reserve demand curve approach to set real-time clearing prices during operating reserves shortages. We evaluated five aspects of the real-time market related to pricing and dispatch in 2008.

1. *Price Corrections*: We find that price corrections were very infrequent, which reduces uncertainty for market participants transacting in the New England wholesale market.
2. *Real-Time Pricing of Fast-Start Resources:* Prices in the real-time market do not always reflect the high costs of fast-start resources when they are used to manage congestion or satisfy load. This causes real-time prices to be understated and affects participants' short-term and long-term incentives. This issue is common to most RTO markets because:
   - Fast-start resources are generally inflexible, and generators can typically only set prices when they are dispatched in their flexible range; and
   - Market clearing prices are usually set based on generators' incremental energy offers and do not include generators' commitment costs (start-up cost offers and no-load

cost offers). Hence, the full cost of a decision to start a fast-start unit may not be included in the real-time prices.

3. *Real-Time Pricing During Transmission Scarcity:* Local shortages can arise when local generation and transmission capability into an area are not sufficient to meet demand in the area. Although such shortages are uncommon, it is important for markets to set efficient prices that reflect such conditions. The following issues can compromise efficient pricing under these conditions and are addressed by our recommendations:
   - The use of a "relaxation" algorithm that effectively raises the transmission limit when the constraint cannot be managed; and
   - The use of "penalty factors" that far exceed the redispatch costs that would be reasonable to incur to relieve the constraint. Penalty factors should indicate the maximum value the market would incur to redispatch generation to manage a constraint.

4. *Real-Time Pricing During Demand Response Activation:* Real-time demand response has surged in New England, from 530 MW in January 2006 to 2,029 MW in January 2009. While demand response resources provide substantial benefits to the market, they also pose significant challenges for efficient real-time pricing:
   - Most real-time demand response resources are not dispatchable and must be activated in advance of real time. This inflexibility prevents demand response resources from setting prices and can cause the real-time market not to perceive a shortage, which undermines the efficiency of the real-time market signals during shortage conditions. Similar issues are likely to arise when new demand response programs come into effect under the FCM.
   - This issue is not unique to New England because very little of the demand response in any market is dispatchable in real-time markets. ISO New England is taking steps to integrate dispatchable demand in the energy and ancillary services markets. However, most loads are not currently capable of meeting the performance requirements.

5. *Ex Ante and Ex Post Pricing:* Like PJM and the Midwest ISO, ISO New England re-calculates prices after each interval (ex post pricing) rather than using the "ex ante" prices produced by the real-time dispatch model. Our evaluation of New England's ex post pricing results indicates that it:
   - Creates a small upward bias in real-time prices in uncongested areas; and
   - Sometimes distorts the value of congestion into constrained areas.

Conclusions and Recommendations

Efficient prices are a critical priority for the real-time energy market because they provide incentives for suppliers to offer competitively, for demand response to participate in the wholesale market, and for investors to build new resources when and where they are most valuable. These incentives cause participants to assist the ISO in maintaining a reliable system.

Our evaluation leads to four real-time pricing recommendations that should improve the performance of the market in the future. These changes will be increasingly important if certain trends continue, such as the rapid growth in demand response resources. Therefore, it is prudent to begin the work necessary to evaluate and address these issues before they raise more serious concerns.

G. System Operations

The wholesale market provides efficient incentives for participants to make resources available to meet the ISO's reliability requirements. When the wholesale market does not meet all of these requirements, the ISO will commit additional generation or take other actions. In addition to additional NCPC costs of these actions, these commitments result in added supply that lowers real-time prices and reduces incentives for loads to purchase their full needs in the day-ahead market. Hence, such commitments should only be made when necessary. In this section, we evaluate several aspects of the ISO's process for satisfying reliability requirements in 2008.

Commitment for Local Reliability

Overall, supplemental commitment for reliability decreased from a daily average of 1,600 MW in 2007 to 1,000 MW in 2008. In Connecticut and Boston, such commitment declined in 2008 primarily due to the effects of recent transmission upgrades into both areas and the change in conduct of the largest supplier in Boston. However, supplemental commitment continued to be substantial in Lower SEMA, accounting for 60 percent of all capacity committed for local reliability in 2008.

Lower SEMA continues to require one of the two large units in the area to be committed almost continuously for local reliability protection of the Cape Cod area. These units are rarely committed in the day-ahead market for economic reasons and must, therefore, frequently be supplementally committed. Transmission upgrades planned to be in-service by the summer of 2009 should substantially reduce the frequency of these commitments and the resulting uplift costs.

Evaluation of Local Second Contingency Commitments

We evaluated supplemental commitments to determine whether those made for local second contingency protection were necessary to meet forecasted minimum capacity requirements in constrained areas. It is important for the ISO to avoid making excess reliability commitments because this depresses economic signals in constrained areas and leads to inflated uplift costs. Our analysis indicates that 93 percent of the supplemental commitments in 2008 were necessary to meet the ISO's local reliability requirements. This level is not 100 percent because:

- Operators may be concerned about the accuracy of the forecasted peak load in the constrained area, may be uncertain about the status or availability of a key resource in the area, or may doubt the reliability of fuel supplies to some units.
- Long lead times can cause the ISO to commit resources prior to the completion of the day-ahead market, increasing the uncertainty regarding the need for additional resources.

Consistency between the RAA limits and actual limits is important because inconsistencies can lead either to unnecessary commitments for local reliability or to local capacity deficiencies. We found that the transmission limits the ISO uses to hold reserves on the import capability into certain constrained areas were modestly higher in the real-time market than in the RAA process on average. In 2008, ISO New England upgraded the analytical tools used to calculate these transmission limits, which should contribute to greater accuracy of the limits.

Accuracy of Load Forecasting

The day-ahead load forecast is significant because market participants may use it and other available information to inform their decisions regarding fuel procurement, management of energy limitations, formulation of day-ahead bids and offers, and outage scheduling. In addition, the ISO uses the forecast to estimate the amount of resources that will be needed to satisfy the load and reserve requirements of the system. Based on our analysis of ISO New England's daily peak load forecasts, we found that the average day-ahead load forecast was slightly higher than the average real-time load in the peak load hour of each day. Overall, the forecasting was very accurate and generally superior to comparable results in other RTO markets.

Self-Commitment after the Reserve Adequacy Assessment

The last analysis in this section finds that self-commitment after the RAA was the primary cause of excess capacity in Boston in the first three months of 2008. This resulted from the conduct of the largest supplier in Boston:

- Higher day-ahead offers reduced the day-ahead commitment of the supplier’s large units;
- This required the ISO to supplementally commit some of the supplier’s other capacity;
- The supplier frequently self-committed the large economic units when they were not committed by the ISO, leading to substantial excess capacity in the Boston area and rendering the supplemental commitments by the ISO unnecessary in retrospect.

In April 2008, the ISO modified the local reliability requirements for the Boston area (made possible by the new transmission capability added into the area), which removed the incentive to engage in this conduct.

Conclusions – System Operations

In general, we conclude that the ISO’s operations to maintain reliability in 2008 were reasonably accurate and consistent with the ISO’s procedures. However, substantial quantities of supplemental commitment continue to occur in Lower SEMA, which does not have a large quantity of fast-start resources that can help meet local reserve requirements. Supplemental commitments raise efficiency concerns because they:

- Diminish the efficiency of New England’s overall commitment;
- Dampen economic signals to invest in areas that would benefit the most from additional investment in generation, transmission and demand response resources;
- Increase uplift costs that are difficult for participants to hedge, and that can be volatile; and
- Create incentives for units frequently committed for reliability to avoid market-based commitment and seek additional payments through the reliability commitment process.

The ISO has implemented, or is pursuing, several measures to minimize reliance on supplemental commitments in load pockets, including:

- Transmission upgrades into Boston (completed in Spring 2007) and associated changes in the area's local reliability requirements (implemented in early 2008).
- Transmission upgrades into Southwest Connecticut that allow less internal online capacity to maintain reliability (mostly in service in 2008, projected completion in 2009).
- Transmission upgrades into Southeast Massachusetts that enable the ISO to maintain reliability in these areas with less internal capacity (planned for 2009).
- Upgrades to the software tools used to calculate transmission capability into local areas. The new PowerWorld based application improves the accuracy, reliability, and efficiency of the calculations (implemented for 2008).
- Changes to the market power mitigation measures to better address the incentives of suppliers that persistently raise their offers above marginal costs to extract larger NCPC payments when they are committed for local reliability (filed in 2009).

In addition, we recommend two changes listed in the table of recommendations below. These changes, together with the pricing improvements proposed above, should improve the performance of the real-time markets and improve the economic signals that they produce.

H. Competitive Assessment

The final section of the report evaluates the market concentration and competitive performance of the markets operated by ISO New England in 2008. Given the constraints on the transmission network, the most substantial market power exists in constrained areas that can become separate geographic markets with a limited number of suppliers when congestion arises. This assessment evaluates the conduct of market participants in these areas.

The first part of our assessment evaluates each geographic market using a pivotal supplier analysis to determine the demand conditions under which a supplier may have market power. This analysis identifies conditions under which the energy and operating reserve requirements cannot be satisfied without the resources of a given supplier (i.e., the "pivotal supplier").

Based on our analyses in the competitive assessment section of the report, we found:

- The largest suppliers in five of the seven areas are pivotal in a large number of hours.

- However, when we account for the large amounts of nuclear capacity and reliability agreements, we find a pivotal supplier in: (i) Lower SEMA in 78 percent of hours, (ii) Boston in 30 percent of hours, and (iii) All of New England in 7 percent of hours.
- Market power will be a more significant concern in Connecticut once the large quantity of reliability agreements expire. Hence, it will be important to continue to monitor these areas and ensure that the market power mitigation measures are fully effective.

The second part of this assessment examines market participant behavior to identify potential exercises of market power. We analyzed potential economic withholding (i.e., raising offer prices to reduce output and raise prices) and physical withholding (i.e., reducing the claimed capability of a resource or falsely taking a resource out of service). Based on our evaluation in the Competitive Assessment section of this report, as well as the monitoring we performed over the course of the year, we find very little evidence of attempts to exercise market power.

While there is no substantial evidence that suppliers withheld capacity from the market to raise clearing prices, suppliers can also exercise market power by raising their offer prices to inflate the NCPC payments they receive when committed for local reliability. The conduct described above that occurred in the Boston area substantially increased NCPC payments, as did conduct in the Lower SEMA area to a lesser extent. We recently coordinated with the Internal Market Monitoring Unit on an evaluation of the criteria used to mitigate offers that increase NCPC payments. We agree that the mitigation criteria for conduct that affects NCPC payments in constrained areas should be modified, particularly given that the introduction of locational forward reserve markets and forward capacity markets should ensure that resources needed for local reliability recover their fixed costs. Furthermore, we support the recent proposed Tariff modifications filed by the Internal MMU to enhance the mitigation measures that are applicable when generation is committed for local reliability.

I. Forward Capacity Market

The Forward Capacity Market was successfully introduced by ISO New England with no significant procedural problems. The qualification processes and the auctions have occurred on schedule, which is noteworthy for a major wholesale market design initiative. Furthermore, the results of the auctions have been competitive, and sufficient capacity is planned to be in-service to satisfy the needs of New England through May 2012. The use of out-of-market payments by the ISO to retain existing resources has been virtually eliminated. This has significantly improved incentives to capacity suppliers compared with the current reliance on reliability agreements to retain existing capacity.

In the first two FCAs, more than 4 GW of new capacity was procured from generation, demand response resources, and imports.[10] However, most of the new investment in generation under FCM has been motivated by out-of-market payments related to RFPs of the Connecticut PUC. A very small amount of new generation has been directly facilitated by the FCM (i.e., generation that was not already committed to enter or that received an award under the Connecticut RFP). This fact alone does not raise any concerns regarding the FCM because there is a substantial surplus of capacity in New England and the prevailing prices in the FCM are well below most estimates of the entry costs for new generation. It is unlikely that substantial amounts of additional generation investment will occur until capacity clearing prices increase significantly. Therefore, it will be difficult to determine whether the FCM facilitates efficient investment in new generation until the current surplus of capacity diminishes.

Large quantities of demand response resources have entered at prices well below the net entry costs for new generation in the first two auctions. This outcome is efficient as long as the market provides investment incentives to demand resources and supply resources that are unbiased so that the lowest-cost resources enter. However, demand response resources accept different (and potentially less costly) obligations than generation resources or imports. The most important difference is that the Peak Energy Rent ("PER") provisions do not apply currently to demand response resources. This may inefficiently bias investment in favor of demand response resources. The Internal MMU also identified this issue in its recent FCM report, and recommended changes that would make the obligations accepted by demand response resources and generation resources more consistent.[11] We support these recommendations.

---

10 This excludes new resources treated as existing resources because they are already committed to enter.

11 See *Review of the Forward Capacity Market Auction Results and Design Elements* by Dave LaPlante, Hung-Po Chao, Sam Newell, Metin Celebi, and Attila Hajos.

Finally, the results of the first forward capacity auction highlight an issue with the Local Sourcing Requirements ("LSRs"), which are the capacity requirements modeled in the auction for individual zones. Two delist bids (330 MW of capacity) were rejected for Connecticut reliability reasons even though the Connecticut LSR was satisfied with an excess of more than 1 GW. The Internal MMU has identified the issue and recommended that the LSR criteria be made consistent with the criteria that are used in the auction to determine whether a delist bid should be rejected for zone-level reliability. We support this recommendation.

J. Table of Recommendations

RECOMMENDATION	SECTION	HIGH BENEFIT	FEASIBLE IN ST[12]
Energy Markets			
1. Evaluate potential pricing changes that would allow costs of fast-start units to be more fully reflected in real-time prices.	VII.B	✓	
2. Develop rules to allow demand response activation to be reflected in prices when they are needed to avoid a shortage.	VII.D	✓	
3. Consider replacing the current ex post pricing process with one that uses ex ante prices for settlement.	VII.E		
4. Consider providing suppliers with flexibility to modify their offers closer to real-time to reflect changes in marginal costs.	VIII.E		
Ancillary Services Markets			
5. Set the local RCPFs at levels that are more consistent with the costs incurred to meet the local-area reserve requirements.	V.B	✓	✓
6. Create additional local reserve zones in the real-time market to satisfy the local reliability requirements more efficiently.	V.C		✓
7. Eliminate the "Rest of System" TMOR requirement is necessary in the forward reserve market.	V.D		✓
8. Consider replacing the forward reserve market's price cap with a tiered cap to recognize higher-value reserves.	V.D		✓
9. Evaluate the benefits of moving to a regulation market that is co-optimized with the energy and ancillary services markets.	VI.C		
Forward Capacity Market			
10. Improve the consistency of the Local Sourcing Requirements with the reliability criteria used in the auction to determine whether to reject a delist bid.	X.C	✓	
11. Extend PER provisions to demand response resources to provide unbiased investment incentives.	X.C	✓	
System Operations			
12. Develop provisions to coordinate the physical interchange between New York and New England in real-time.	III.C	✓	
13. Discontinue relaxation of violated transmission constraints -- set penalty factors that reflect the value of constraints and allow them to determine LMPs when a constraint is violated.	VII.C		✓

12 *Feasible in Short Term*: indicated if the recommendation is likely to be feasible within one to two years at a reasonable cost. Others likely require study of costs and benefits, or research to identify a feasible approach.

High Benefit: Indicated for recommendations that will likely produce considerable efficiency benefits.

II. Prices and Market Outcomes

In this section, we review wholesale market outcomes in New England during 2008. Our review includes analyses of overall price trends, patterns of transmission congestion, and convergence of prices in the day-ahead and real-time markets.

A. Price Trends

Our first analysis examines day-ahead prices at the New England Hub in 2007 and 2008. The New England Hub is located at the geographic center of New England and is an average of the prices at 32 individual pricing nodes. The New England Hub price has been developed and published by the ISO to disseminate price information that facilitates bilateral contracting. The average New England Hub price increased from about $70 per MWh in 2007 to approximately $83 per MWh in 2008. This 18 percent increase was primarily due to significant increases in fuel prices from 2007 to 2008. The average natural gas price rose from $8.3 per MMbtu in 2007 to $10.1 per MMbtu in 2008, a 21 percent increase. These price changes are evaluated and discussed in more detail below.

Figure 1 shows the average price at the New England Hub in the day-ahead market for each month in 2007 and 2008.[13] The figure also shows the average natural gas price,[14] which is a key driver of electricity prices when the market is operating competitively. Currently, approximately 40 percent of the generating capacity in New England burns natural gas as its primary fuel.[15] Low-cost coal and nuclear resources typically produce at full output, while natural gas-fired resources are on the margin and set the market clearing price in most hours. Therefore, electricity prices should be correlated with natural gas prices. This relationship is evident in Figure 1.

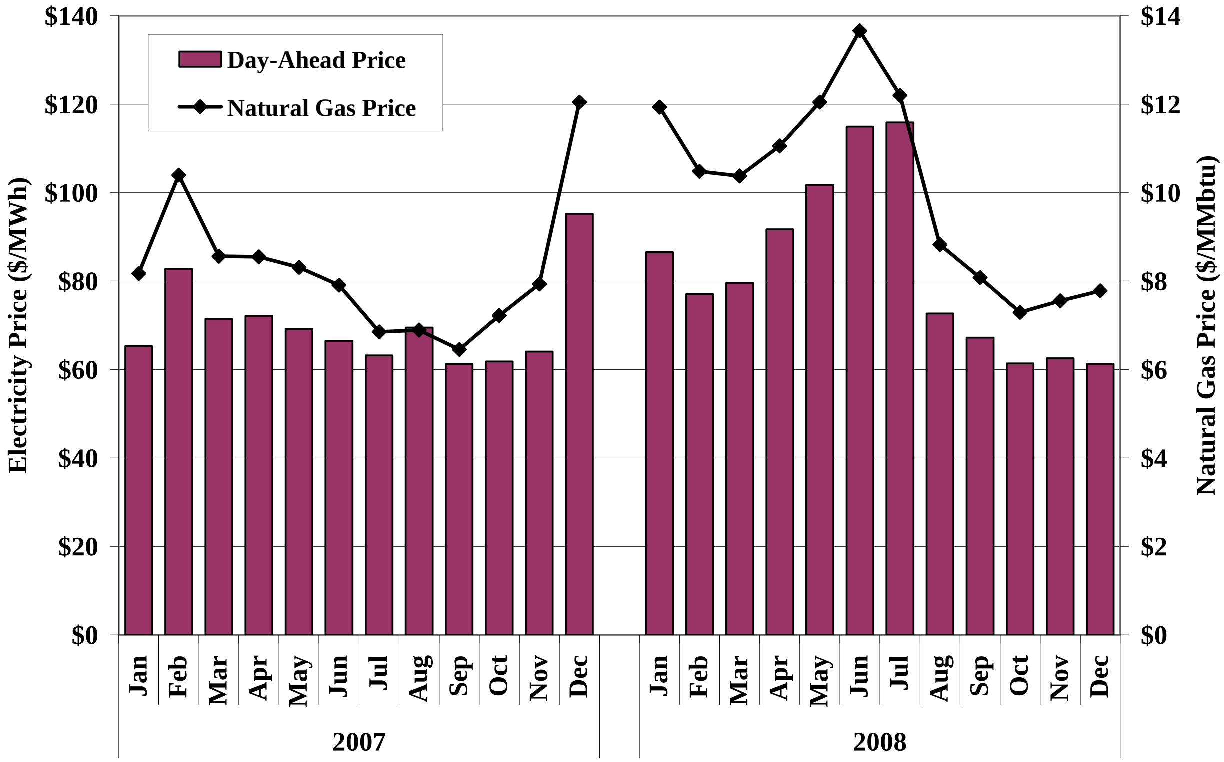
---

13 This average is weighted by the New England load level in each hour.

14 The figure shows the day-ahead price reported for the Algonquin pipeline.

15 ISO New England, "2008-2017 Forecast Report of Capacity, Energy, Loads, and Transmission (CELT) Report," April 2008.

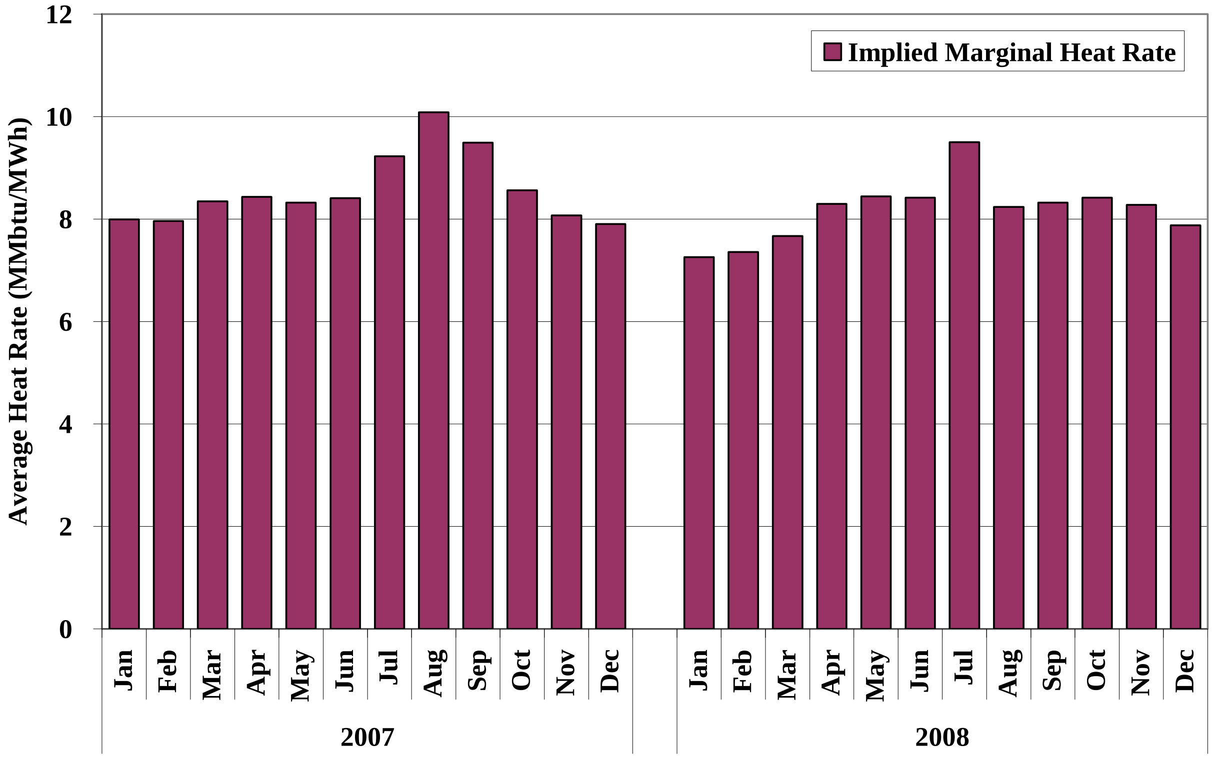
Figure 1: Monthly Average Day-Ahead Prices and Natural Gas Prices New England Hub, 2007 - 2008



As expected, natural gas price fluctuations were the most significant driver of variations in monthly average electricity prices in 2007 and 2008. Natural gas prices increased sharply in February and December 2007, leading to significant increases in average power prices. Similarly in 2008, the average natural gas price increased from March to June and decreased from July to October, coinciding with a rise and fall in power prices over the same period. Power prices usually increase during high summer and winter load periods when the demand for heating and cooling are highest. However, the effects of seasonal changes in demand were smaller than the effects of changes in fuel prices in New England.

To better identify changes in electricity prices that are not related to the fluctuations in natural gas prices, Figure 2 shows the marginal heat rate that would be implied if natural gas resources were always on the margin. The *implied marginal heat rate* is equal to the electricity price divided by the natural gas price measured in MMbtu. Thus, if the electricity price is $72 per MWh and the natural gas price is $9 per MMbtu, this would imply that an 8.0 MMbtu per MWh generator is on the margin. Figure 2 shows the load-weighted average implied marginal heat rate for the New England Hub in each month during 2007 and 2008.

Figure 2: Monthly Average Implied Marginal Heat Rate
Based on Day-Ahead Prices at New England Hub, 2007 - 2008



By adjusting for the variation in natural gas prices, the implied marginal heat rate shows more clearly the seasonal variation in electricity prices. The figure shows that implied marginal heat rates were highest in the summer months (from June to September), due primarily to the higher loads and tighter market conditions that prevail on the hottest days during the summer. The months with the highest average loads were August 2007 and July 2008. Accordingly, the months with the highest average implied marginal heat rates were also August 2007 and July 2008.

The average implied marginal heat rate decreased 5 percent from 2007 to 2008. During the summer months, the implied marginal heat rate averaged 9.3 MMbtu per MWh in 2007 and 8.6 MMbtu per MWh in 2008. The weather was particularly mild in the summer of 2008, which likely contributed to the decline in the implied marginal heat rate. Outside the summer, the average implied marginal heat rate fell slightly from an 8.2 MMbtu per MWh in 2007 to 8.0 MMbtu per MWh in 2008.

B. Prices in Transmission Constrained Areas

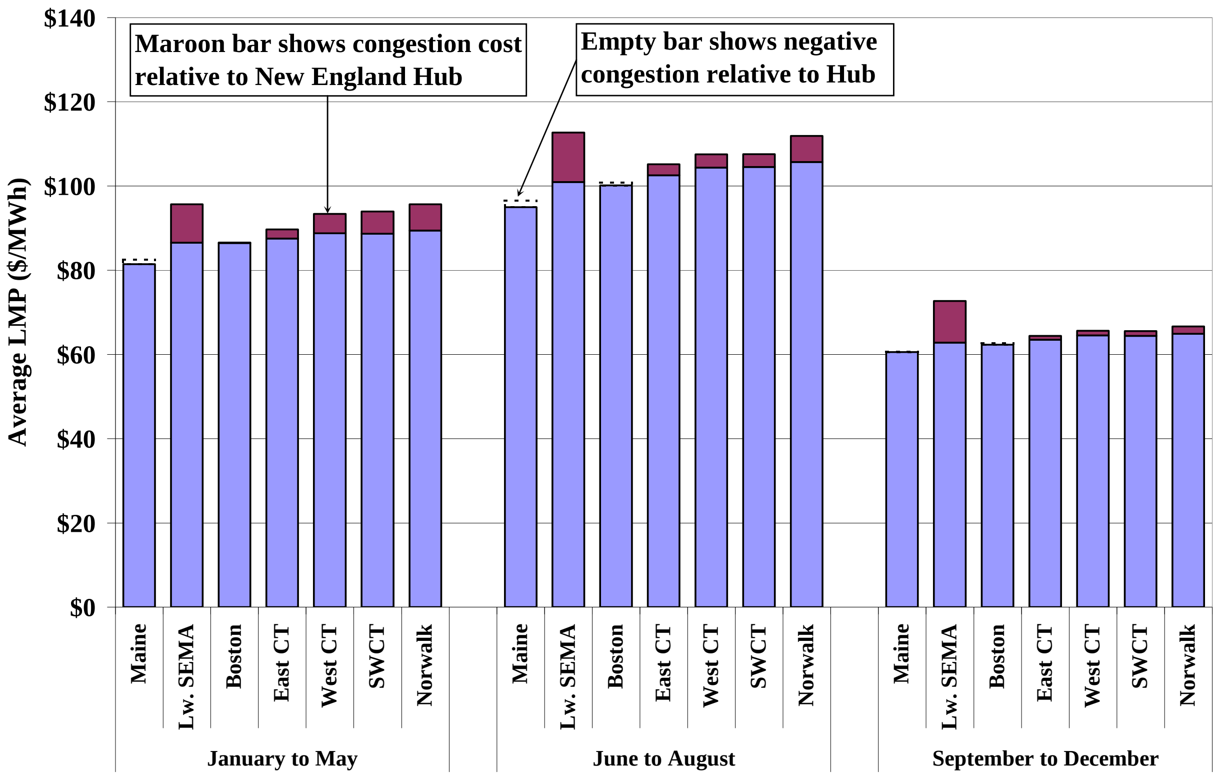
Historically, there have been significant transmission limitations between net-exporting and net-importing regions in New England. In particular, exports from Maine to the rest of New England have frequently been limited by transmission constraints, while Connecticut and Boston were often unable to import enough power to satisfy demand without dispatching expensive local generation in the past. ISO New England uses locational marginal pricing ("LMP") to manage transmission constraints in an efficient manner and to produce local price signals. In LMP markets, the variation in prices across the system reflects the marginal value of transmission losses and congestion and provides incentives for the efficient dispatch of resources.

Losses occur whenever power flows across the transmission network. Generally, transmission losses increase as power is transferred over longer distances and/or at lower voltages. The rate of transmission losses also increases as power flow increases across a particular transmission facility. Transmission congestion arises when the lowest-cost resources cannot be fully dispatched because transmission capability is not sufficient to deliver all of their output to end-users. When congestion occurs, LMP markets establish a spot price for energy at each location on the network that reflects the marginal system cost of meeting load at that location. The LMPs can vary substantially across the system, reflecting the fact that higher-cost units must be dispatched in place of lower-cost units to serve incremental load while not overloading any transmission facilities. This causes LMPs to be higher in "constrained locations" than locations where there is no congestion.

Just as transmission constraints limit the delivery of energy into an area and require higher cost generation to operate in the constrained area, transmission constraints may also require additional operating reserves in certain locations to maintain reliability. Such locational requirements are used in the real-time reserve market to schedule reserves and energy efficiently in local areas, particularly during shortages. When generation is redispatched in real-time to provide additional reserves to a local area, the marginal system cost of the redispatch is reflected in the LMPs. The reserve markets are discussed in Section V.

We analyzed the differences in energy prices between several key locations during the study period. Figure 3 shows load-weighted average day-ahead LMPs on a seasonal basis for the Maine load zone, Lower SEMA, NEMA/Boston load zone, and four areas within Connecticut. Connecticut is divided into: East Connecticut, the portion of West Connecticut that excludes Southwest Connecticut, the portion of Southwest Connecticut that excludes Norwalk-Stamford, and Norwalk-Stamford.

Figure 3: Average Day-Ahead Prices by Location
2008



Note: The average prices reported for SWCT exclude Norwalk-Stamford, and the prices for West CT exclude SWCT and Norwalk-Stamford.

For each location, the load-weighted average LMP (including the effects of marginal transmission losses) is indicated by the height of the solid bars. The maroon portion of the bars indicates positive congestion to the location from the New England Hub, while negative congestion is indicated by the empty bars. Thus, the areas that are import-constrained (e.g., Lower SEMA) exhibit positive congestion from the Hub.

It is notable that there was very little congestion into Boston in 2008. This was due in large part to the substantial increase in import capability associated with the NSTAR 345 kV Transmission Project that was brought in-service in the spring of 2007. As in 2007, Lower SEMA was the area most affected by congestion in 2008. The congestion price difference between the Hub and Lower SEMA averaged $10 per MWh in 2008. Of the areas shown in Figure 3, Norwalk-Stamford experienced the second most congestion in 2008. On average, the congestion price difference between the Hub and Norwalk-Stamford was $5 per MWh in 2008. Congestion into Lower SEMA and Norwalk-Stamford was primarily associated with “proxy second contingency transmission constraints”. These constraints are used to limit the imports over the transmission interface into the local areas from the rest of New England in order to effectively hold operating reserves on the transmission interface and satisfy local second contingency reliability requirements,

Also notable is that congestion declined in the fall 2008 relative to the spring and summer. This is primarily due to the substantial reduction in fuel prices after the summer months. The price of residual fuel oil fell below the price of natural gas from late October through December 2008, leading to more commitment of large thermal generators in local areas that burn residual fuel oil. This commitment in local areas helped reduce congestion into local areas.

C. Convergence of Day-ahead and Real-Time Prices

The day-ahead market allows participants to make forward purchases and sales of power for delivery in real time. The market provides a valuable financial mechanism that allows participants to hedge their portfolios and manage risk. Loads can hedge price volatility in the real-time market by purchasing in the day-ahead market. Suppliers can avoid the risk of unprofitably starting their generators, because the day-ahead market will only accept their offers when they will profit from being committed. However, suppliers that sell day ahead are exposed to some risk because they are committed to deliver energy in the real-time market. An outage can force them to purchase replacement energy from the spot market during a price spike.

In well-functioning day-ahead and real-time markets, we expect that day-ahead and real-time prices will not systematically diverge. If day-ahead prices were predictably higher than real-time prices, buyers would decrease purchases and sellers would increase sales in the day-ahead market. Alternatively, if day-ahead prices were foreseeably lower than real-time prices, buyers would increase purchases and sellers would decrease sales in the day-ahead market.

Historically, average day-ahead prices have been relatively consistent with the average real-time prices in New England, although it has been common for day-ahead prices to carry a slight premium over real-time prices. Predictable day-ahead price premiums encourage speculative market participants to schedule "virtual supply" (i.e., to sell short at the day-ahead price and buy back at the real-time price). This response puts downward pressure on day-ahead prices and tends to limit the size of the average day-ahead premium. Price convergence is desirable because it promotes the efficient commitment of generating resources and scheduling of external transactions.

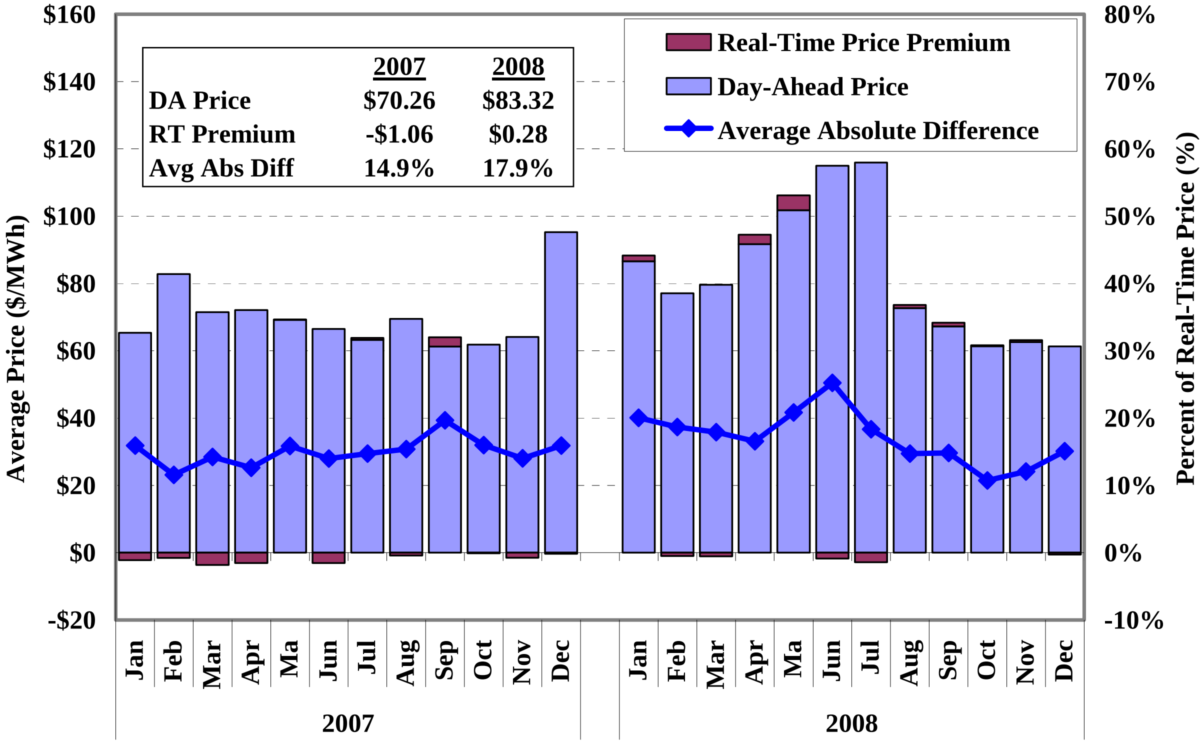
In the remainder of this section, we evaluate the convergence of prices between day-ahead and real-time markets. Section D examines convergence of energy prices at the New England Hub, which is broadly representative of the New England market. Section E examines convergence of energy prices in several areas that are sometimes isolated from the rest of New England by transmission constraints.

D. Price Convergence at the New England Hub

We examine price convergence between the day-ahead and real-time markets at the New England Hub to provide an indication of the overall price convergence in the region. In this section, two measures are used to assess price convergence. The first measure reports the simple difference between the average day-ahead price and the average real-time price. The second measure reports the average absolute difference between day-ahead and real-time prices on an hourly basis. The first measure is an indicator of the systematic differences between day-ahead and real-time prices, while the second measure captures the overall variability between day-ahead and real-time prices over the year.

Figure 4 summarizes day-ahead prices and the convergence between day-ahead and real-time prices at the New England Hub in each month of 2007 and 2008.[16]

Figure 4: Convergence of Day-Ahead and Real-Time Prices at New England Hub 2007 - 2008



The first measure of convergence reported in the figure, the average real-time premium, is equal to the average real-time price minus the average day-ahead price. The sum of the average day-ahead price (blue bar) and the average real-time price premium (maroon bar) is equal to the average real-time price. The second measure of convergence, the average absolute difference between day-ahead and real-time prices, is shown by the blue line, and it is reported as a percentage of the average day-ahead price in the month.

In electricity markets, average day-ahead prices tend to be slightly higher than average real-time prices, partly because many buyers are willing to pay a small premium to purchase at day-ahead prices, which are less volatile than real-time prices. In addition, load settling in the real-time

16 These are averaged on a load-weighted basis.

market is usually subject to higher NCPC allocations. This is consistent with the difference between average day-ahead and average real-time prices that was observed in 2007 when most months exhibited a day-ahead price premium. In 2008, however, average real-time prices were slightly higher than average day-ahead prices (by $0.28 per MWh) with some months exhibiting a day-ahead premium and other months exhibiting a real-time premium. Overall, the differences between day-ahead and real-time prices are modest and indicate good overall convergence.

The second measure of price convergence evaluated in the figure above is the average absolute difference between day-ahead and real-time prices. This measure is calculated by averaging the absolute value of the hourly differences between day-ahead and real-time prices. As a percentage of the average day-ahead price in each year, the average absolute difference increased from 14.9 percent in 2007 to 17.9 percent in 2008. The average absolute difference was particularly elevated during the period from January to July 2008 when natural gas and oil prices rose to unusually high levels. Convergence between day-ahead and real-time prices is diminished by fuel price volatility, which increases the uncertainty faced by market participants in the day-ahead market.

E. Price Convergence in Transmission Constrained Areas

When the transmission system is unconstrained, all buyers and sellers effectively participate in a single, regional market. Hence, resources throughout the system are utilized to respond to unexpected changes in load or available supply, which diminishes the price effects from these events. When transmission constraints are binding, such events can have a much greater effect in the congested area. This section examines price convergence in locations that are most frequently isolated from the rest of New England by congestion.

The following table summarizes convergence between day-ahead and real-time prices at the New England Hub, one frequently export-constrained location (Maine), and several frequently import-constrained locations.[17] Connecticut is divided into four regions due to the various internal constraints affecting flows within the state. Convergence is measured in each area using: (i) the

[17] The average day-ahead and real-time prices are load-weighted average prices.

difference between the average day-ahead and average real-time prices and (ii) the average absolute difference between hourly prices in the day-ahead and real-time markets as a percentage of the average real-time clearing price. The difference in average prices shows whether prices over the entire study period were higher in the day-ahead or real-time market. The average absolute difference shows the size of the hourly price variations.

Table 1: Convergence between Day-Ahead and Real-Time Prices by Region 2007 - 2008

	Real-Time Clearing Price ($/MWh)		Day-Ahead - Real-Time Price Difference ($/MWh)		Hourly Absolute Price Difference (percent of RT Price)	
Region	**2007**	**2008**	**2007**	**2008**	**2007**	**2008**
New England Hub	$69.20	$83.61	$1.06	-$0.28	15%	17%
Maine	$65.91	$78.00	$0.45	$0.45	15%	17%
Lower Southeast Massachusetts	$70.80	$88.28	$3.79	$4.65	18%	20%
Boston	$68.07	$83.35	$0.76	-$0.80	15%	17%
Areas in Connecticut:						
East Connecticut	$73.34	$85.13	-$1.34	$0.67	18%	18%
West CT (excluding SWCT)	$75.97	$87.11	-$1.38	$1.20	18%	18%
SWCT (excluding Norwalk)	$75.45	$86.97	-$0.41	$1.58	18%	18%
Norwalk-Stamford	$76.40	$88.67	$0.46	$2.12	19%	19%

Table 1 shows that price convergence was generally better in the less congested locations, reflecting that unforeseen market events tend to have larger price effects in isolated areas. The difference between the average day-ahead price and average real-time price was highest in Lower SEMA in both 2007 and 2008. The average absolute difference was also higher in the import-constrained locations than at the Hub.

Changes in the commitment of key generators after the day-ahead market contribute to poor convergence between the day-ahead and the real-time market in some load pockets. This was evident in Lower SEMA, which exhibited day-ahead premiums of $3.79 per MWh in 2007 and $4.65 per MWh in 2008. On many days, the majority of the generation in these areas was committed after the day-ahead market. As a result, congestion in the day-ahead market was often based purely on scheduled load bids and virtual transactions with no physical resources scheduled.

Notwithstanding the relatively poor convergence in Lower SEMA, we find that the overall convergence between day-ahead prices and real-time prices in New England was good. The average differences are very similar to those in other RTO markets and the average absolute differences are the lowest of any of the RTOs in the eastern Interconnect. We attribute the latter result to the relatively low real-time price volatility in New England.

III. External Interface Scheduling

This section examines the scheduling of imports and exports between New England and adjacent regions. New England receives imports from Quebec and New Brunswick in most hours, which reduces wholesale power costs for electricity consumers in New England. Between New England and New York, power flows alternate directions depending on market conditions. Overall, New England exported more power to New York than it imported in 2008. The transfer capability between New England and adjacent control areas is large relative to the typical load in New England, making it particularly important to schedule interfaces efficiently.

Consumers benefit from the efficient use of external transmission interfaces. The external interfaces allow low-cost external resources to compete to serve consumers who would otherwise be limited to available internal resources. The ability to draw on neighboring systems for emergency power, reserves, and capacity also helps lower the costs of meeting reliability standards in the interconnected system. Wholesale markets facilitate the efficient use of both internal resources and transmission interfaces between control areas.

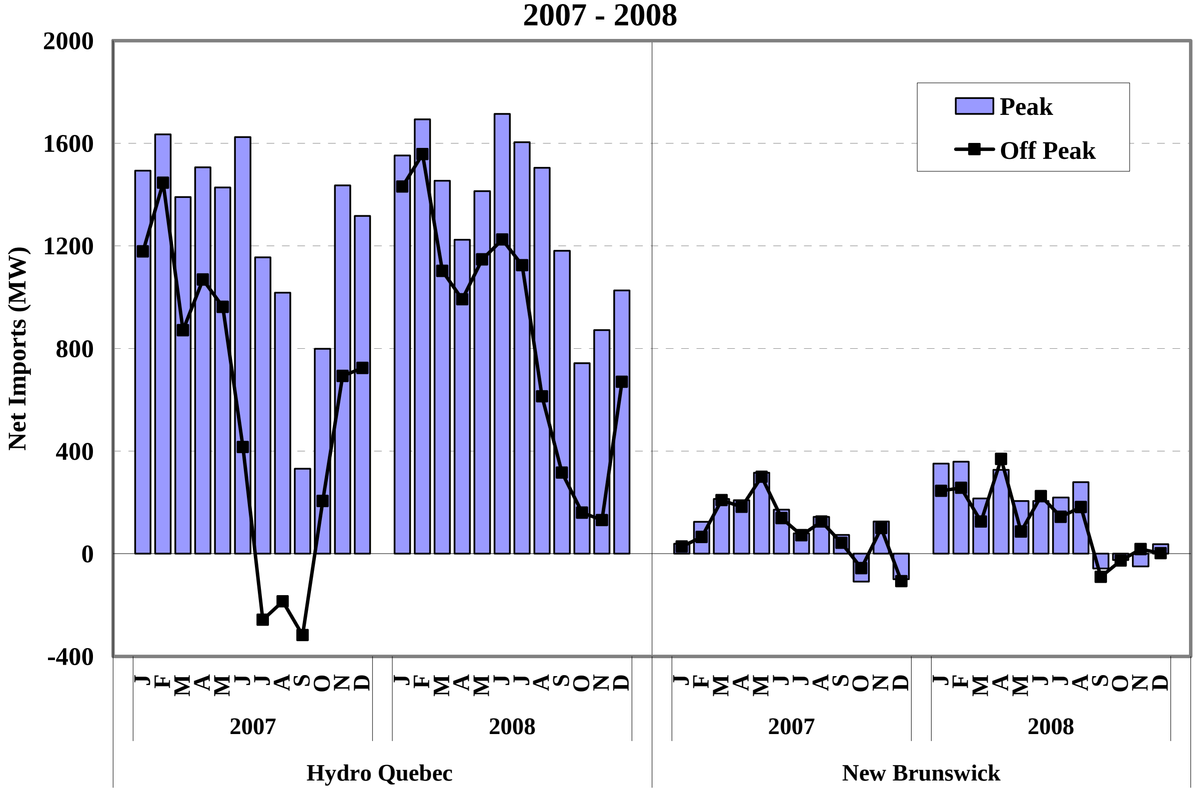
ISO-NE is interconnected with three neighboring control areas: the New York ISO, TransEnergie (Quebec), and the New Brunswick System Operator. New England and New York are interconnected by three interfaces: the Roseton Interface, which includes several AC tie lines connecting Upstate New York to Connecticut, Massachusetts, and Vermont; the 1385 Line, a controllable AC interconnection between Norwalk and Long Island; and the Cross-Sound Cable, a DC interconnection between Connecticut and Long Island. New England and Quebec are interconnected by two interfaces: Phase I/II, which is a large DC interconnection, and the Highgate Interface, which is a smaller AC interconnection between Vermont and Quebec. New England and New Brunswick are connected by a single interface.

This section evaluates several aspects of transaction scheduling between New England and adjacent control areas. Section A summarizes scheduling between New England and adjacent areas in 2007 and 2008. Section B evaluates the efficiency of scheduling by market participants between New York and New England. Section C presents an estimate of the benefits that would result from efficient coordination of interchange between New York and New England by the ISOs. This section also discusses efforts to reduce barriers to efficient scheduling and identifies additional changes that could further improve scheduling across the "seams" between New England and the adjacent markets.

A. Summary of Imports and Exports

The following two figures provide an overview of imports and exports by month for 2007 and 2008. Figure 5 shows the average net imports across the three interfaces with Quebec and New Brunswick by month, for peak and off-peak periods.[18] The net imports across the two interfaces linking Quebec to New England are combined together.

Figure 5: Average Net Imports from Canadian Interfaces 2007 - 2008



18 Peak hours include hours-ending 7 to 22, Monday through Friday, and the remaining hours are included in Off-Peak.

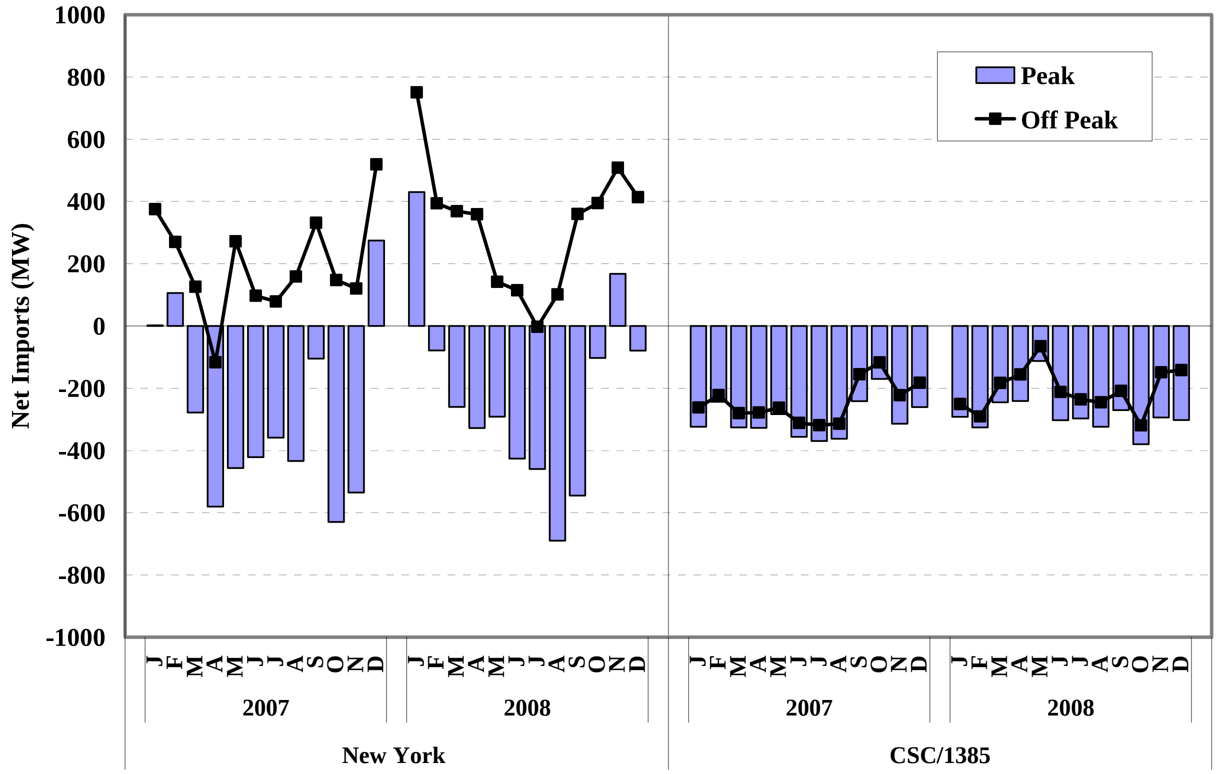
Figure 5 shows that power is generally imported from Quebec and New Brunswick. Across the two interfaces with Quebec, average net imports were higher during peak hours than during off-peak hours by 700 MW in 2007 and by 460 MW in 2008. These patterns reflect the tendency for hydro resources in Quebec to store water during low demand periods in order to make more power available during high demand periods. Net imports over the New Brunswick interface were much lower than the Quebec interfaces and did not vary significantly from peak to off-peak hours in either year. Overall, net imports from Canada increased modestly from 2007 to 2008. Average net imports increased from 1,030 MW in 2007 to 1,180 MW in 2008 from Quebec and from 100 MW in 2007 to 160 MW in 2008 from New Brunswick.

Figure 6 shows average net imports across the three interfaces with New York by month in 2007 and 2008 for peak and off-peak periods. The net imports across the Cross-Sound Cable and the 1385 Line are combined together. However, the 1385 Line was treated as a part of the New York AC interface prior to June 27, 2007, so the figure does not report flows across the 1385 Line before this date.[19] After this date, the 1385 Line was out of service for most of 2007 and 2008 due to cable replacement work and problems with the phase shifter. The 1385 Line came back into service at the beginning of January 2009. Due to the outages of the 1385 Line, the combined net imports reported in the figure primarily reflect flows across the Cross-Sound Cable interface.

Across the primary interface with New York, Figure 6 shows that power usually flowed into New York during peak periods and into New England during off-peak periods. However, the level of flows varied considerably during 2007 and 2008. Overall, New England was a net exporter in both 2007 and 2008.

19 The 1385 Line can carry up to 100 MW between Norwalk and Long Island.

Figure 6: Average Net Imports from New York Interfaces
2007 - 2008



The figure shows that flows were relatively consistent from New England to Long Island across the Cross-Sound Cable, averaging approximately 300 MW in 2007 and 2008. Hence, the Cross-Sound Cable, which has a transfer capability of 330 MW, was fully utilized to export power to Long Island in a large share of the hours in 2007 and 2008.

B. Interchange with New York

The performance of wholesale electricity markets depends not only on the efficient use of internal resources, but also the efficient use of transmission interfaces with adjacent areas. This section evaluates the efficiency of scheduling between New England and New York. Since both regions have real-time spot markets, market participants can schedule market-to-market transactions based on transparent price signals in each region. Based on the prevailing prices in each market, we can evaluate whether the interface is scheduled efficiently.

Trading between neighboring markets tends to bring the prices in the two markets closer together. When an interface is used efficiently, prices in adjacent areas should be consistent unless the interface is constrained. For example, when prices are higher in New England than in New York, imports from New York should continue until prices have converged or until the interface is fully scheduled. A lack of price convergence indicates that resources are being used inefficiently. In other words, higher-cost resources are operating in the high-priced region that could have been supplanted by increased output from lower-cost resources in the low-priced region. It is especially important to schedule flows efficiently between control areas during peak demand conditions when small amounts of additional imports can substantially reduce prices.

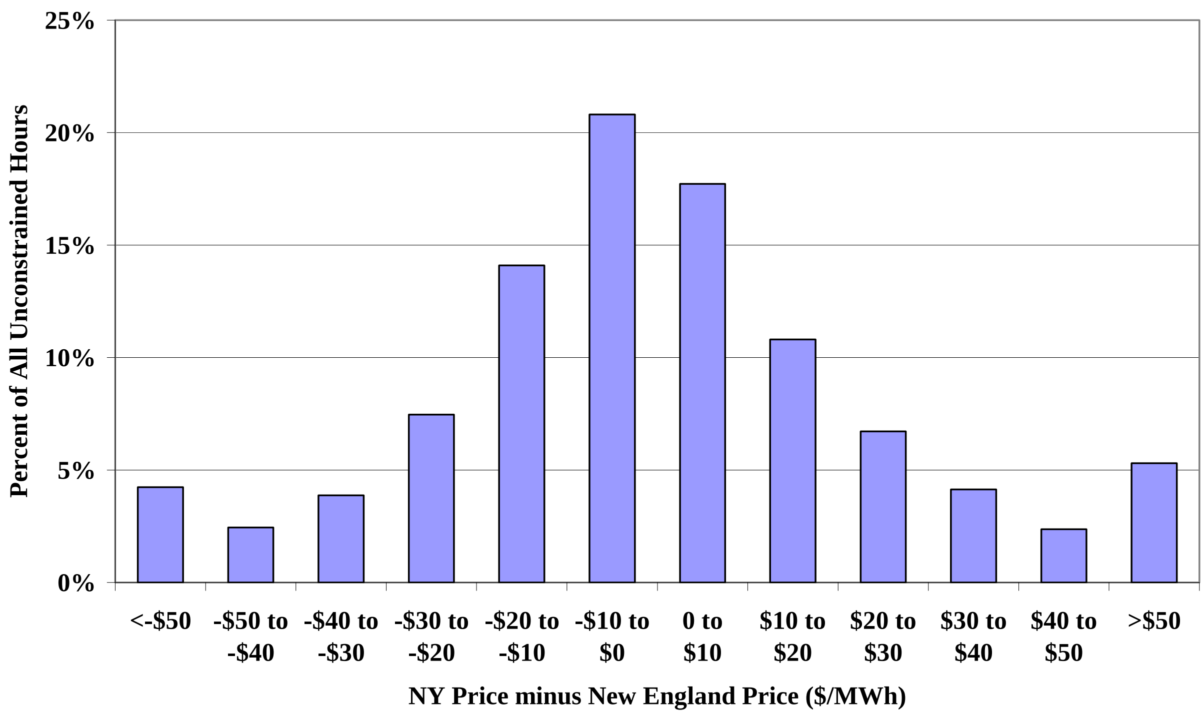
Several factors prevent real-time price differences between New England and New York from being fully arbitraged. First, market participants may not be able to predict which side of the interface will have a higher real-time price when transaction bids and offers must be submitted. Second, differences in the procedures and timing of scheduling in each market serve as barriers to full arbitrage. Third, there are transaction costs associated with scheduling imports and exports that diminish the returns from arbitrage. Participants will not schedule additional power between regions unless they expect a price difference greater than these costs. Last, risks associated with curtailment and congestion can reduce the incentives of participants to schedule external transactions when the expected price difference is small. Given these considerations, one cannot reasonably expect that trading by market participants will fully optimize the use of the interface.

The following figures focus on the efficiency of scheduling across the primary interface between New England and New York. The Cross-Sound Cable is not evaluated in the following figures because it is scheduled under separate rules.[20] The 1385 Line is also not included because it was

20 Service over the Cross-Sound Cable is provided under the Merchant Transmission Facilities provisions in Schedule 18 of ISO New England's Tariff, which is separate from the transmission service provisions governing use of the Pool Transmission Facilities. Access to the MTF requires Advance Reservations on the CSC, recommended to be acquired in advance of submitting transactions to the day-ahead market, and energy transactions accepted in ISO New England and NYISO market systems. Scheduling limits restrict the ability to use the CSC interface for short-run arbitrage transactions between Connecticut and Long Island.

out of service in all but one month of 2008. Figure 7 shows the distribution of real-time price differences across the primary interface between New England and New York in hours when the interface was unconstrained. While the factors listed above prevent complete arbitrage of price differences between regions, trading still helps keep prices in the neighboring regions from diverging excessively. Nonetheless, Figure 7 shows that more than 60 percent of the unconstrained hours have real-time price differences of greater than $10 per MWh. In almost 10 percent of the hours, the price difference is greater than $50/MWh.

Figure 7: Real-Time Price Difference Between New England and Upstate New York Unconstrained Hours, 2008



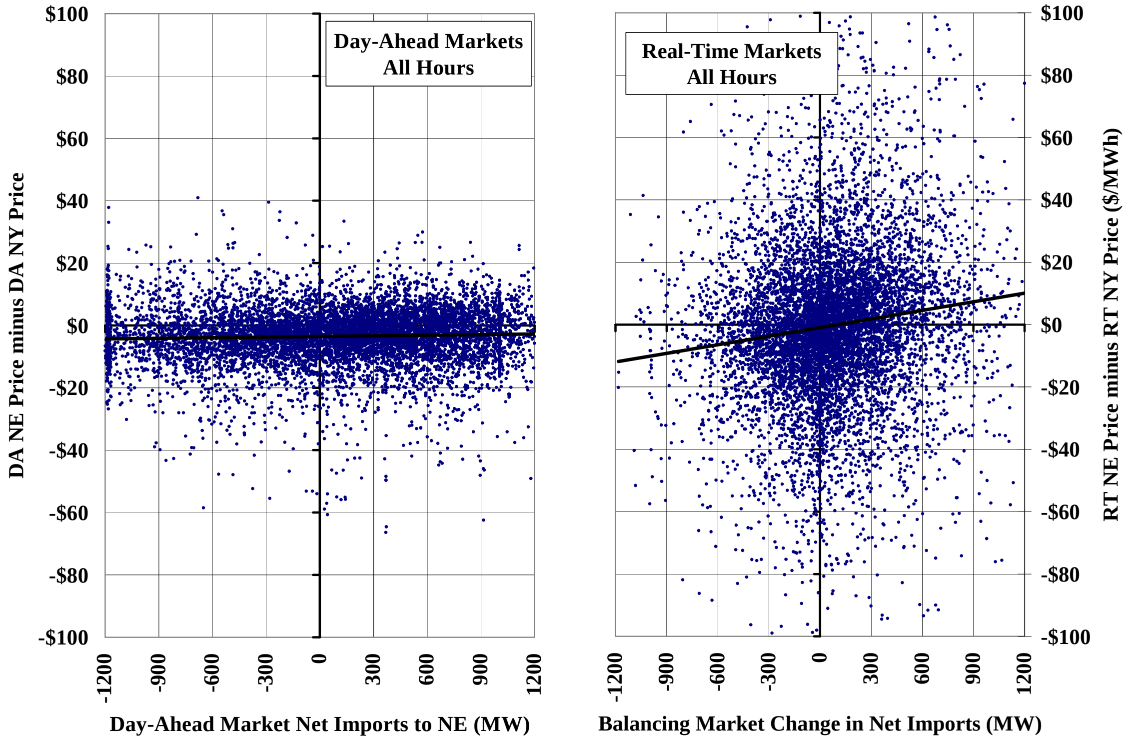
The results shown in the figure indicate that the current process does not fully utilize the interface. Given the size and frequency of price differences shown, there is a substantial number of hours when increasing flows from the lower priced region into the higher priced region would have had significant beneficial effects on the efficiency of clearing prices and production in both regions.

In real-time, it has proven difficult for the adjacent markets to achieve price convergence by relying on transactions scheduled by market participants. Uncertainty, imperfect information, and offer submittal lead times limit the ability of participants to capitalize on real-time arbitrage opportunities. This failure to fully arbitrage the interfaces leads to market inefficiencies that could be remedied if the ISOs were to coordinate interchange.

Although scheduling by market participants has not fully arbitraged the interfaces between New York and New England, the following analysis demonstrates that scheduling by market participants has incrementally improved price convergence. Figure 8 shows, for each hour in 2008, the net scheduled flow across the interface versus the difference in prices between New England and Upstate New York. The left side of the figure shows price differences in the day-ahead market on the vertical axis versus net imports scheduled in the day-ahead market on the horizontal axis. The right side of the figure shows hourly price differences in the real-time market on the vertical axis versus the *change* in the net scheduled imports after the day-ahead market on the horizontal axis. For example, if day-ahead net scheduled imports for an hour are 300 MW and real-time net scheduled imports are 500 MW, the change in net scheduled imports after the day-ahead market would be 200 MW.

The trend line in the left panel does not show a positive or negative correlation between the price difference and direction of scheduling in the day-ahead market. However, the trend line in the right panel does show a statistically significant positive relationship between the price difference and the direction of scheduling in the real-time market. This positive relationship indicates that the scheduling of market participants generally responds to price differences by increasing net flows scheduled into the higher priced region.

Figure 8: Efficiency of Scheduling in the Day-Ahead and Real-Time Interface Between New England and New York, 2008



Total net revenues from cross-border scheduling in 2008 were $10 million in the real-time market (not accounting for transaction costs).[21] The fact that significant profits were earned from the external transactions provides additional support for the conclusion that market participants generally help improve market efficiency by facilitating the convergence of prices between regions, although the arbitrage of prices is far from complete.

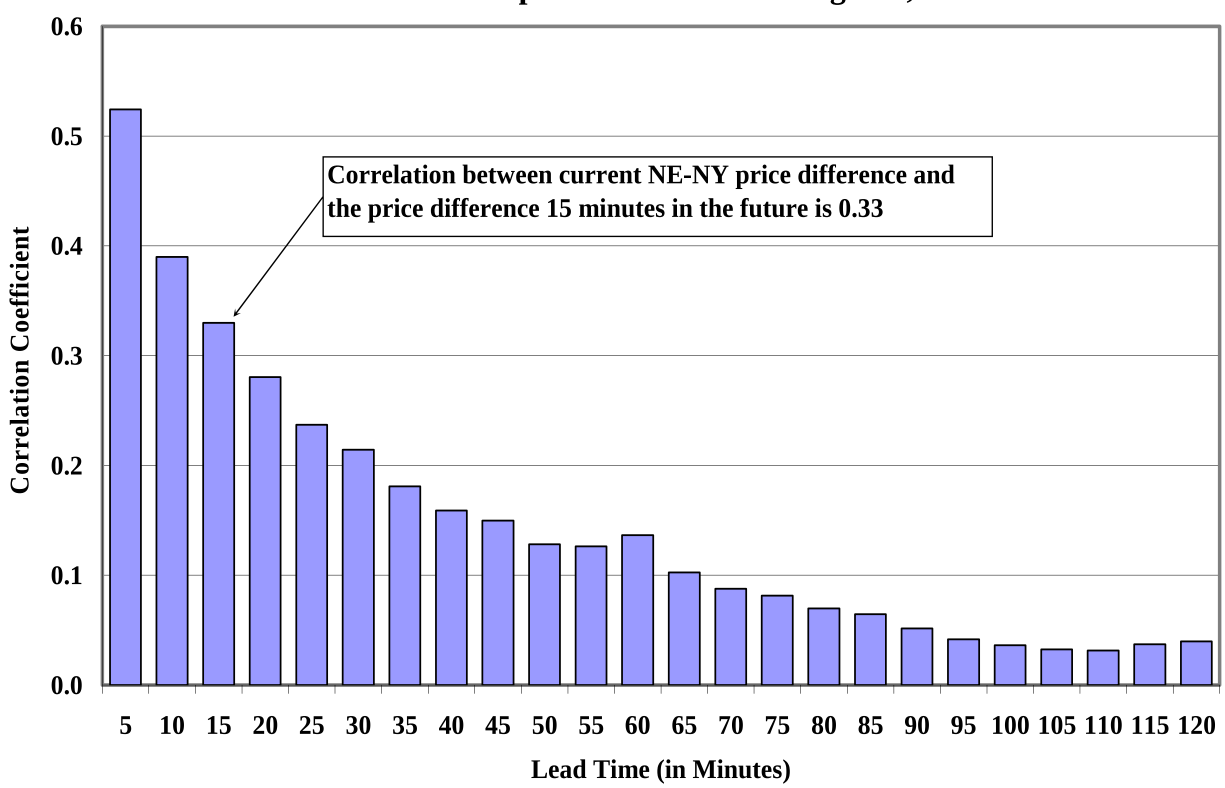
Nevertheless, the plots show a wide dispersion of points on the right that reflects the difficulty of predicting changes in market conditions in real-time. Forty-four percent of the points in the real-time market panel are in unprofitable quadrants – upper left and lower right – indicating hours when the net real-time adjustment by market participants shifted scheduled flows in the

21 This likely underestimates the actual profits from scheduling because it assumes that day-ahead exports from one market are matched with day-ahead imports in the other market. However, market participants have other options such as matching a day-ahead export in one market with a real-time import in the other market. This flexibility actually allows participants to earn greater profits from more efficient trading strategies than those represented in the figure.

unprofitable direction (increasing output in the high-priced market and reducing output in the low-priced market). Although market participant scheduling has helped converge prices between adjacent markets, Figure 8 shows that there remains considerable room for improvement.

The next analysis examines the correlation between the lead times for scheduling transactions and the predictability of price differences between adjacent markets. Figure 9 reports the correlation coefficient of the real-time price difference between New England and Upstate New York between the current period and each subsequent five-minute period over two hours. For example, the correlation of the price difference at the current time and the price difference 15 minutes in the future was 0.33 in 2008.

Figure 9: Correlation Between Price Differences and Lead Time
Interface between Upstate NY and New England, 2008



Not surprisingly, Figure 9 shows that actual price differences are more strongly correlated to price differences in periods near in time than to price differences in periods more distant in time. Currently, to schedule transactions between New York and New England, market participants must submit their offers 75 minutes before the start of an hour, which is 75 to 135 minutes before the power actually flows since transactions are scheduled in one-hour blocks at the top of the hour. This analysis suggests that reducing the lead times for scheduling would enable market participants to schedule more efficiently.

C. Coordination of Interchange by the ISOs

Incomplete price convergence between New England and New York suggests that more efficient scheduling of flows between markets would lead to production cost savings and substantial benefits to consumers. Although past efforts to reduce barriers to market participant scheduling between regions have improved the efficiency of flows, and additional such efforts would lead to further improvements, uncertainty and risk are inherent in the market participant scheduling process. Hence, even with improvements, one cannot reasonably expect the current process to fully utilize the interface. As is the case for efficient scheduling of the transmission capability within ISO regions, optimal use of transmission capability between ISO regions requires explicit coordination of interchange by the ISOs.

We employed simulations to estimate the benefits of optimal hourly scheduling of the primary interface between New England and New York in 2006 and 2007. The benefits of efficient scheduling include reduced production costs and lower prices for consumers. The production cost net savings represent the increased efficiency of generator operations in both regions as additional production from lower-cost generators displaces production from higher-cost generators. The net consumer savings arise because improved coordination between the ISOs tends to lower prices on average in both regions.

Our prior simulations showed that consumers in both regions together would have saved $320 million in 2006 and 2007, $83 million of which would have been saved in New England. New England's lower share of consumer savings is attributable to the lower frequency of shortages in New England than New York in those years.

Shortage pricing provisions in both the New York and New England markets have contributed to more efficient energy pricing when reserve shortages occur. Coordination of the physical interchange between the ISOs can be especially important in avoiding or resolving shortage conditions. Our analysis suggests that ISO coordination of external flows during reserve shortages would have accounted for almost 40 percent of the potential consumer savings in 2007. Hence, as capacity margins decrease and the frequency of shortages increase, the total savings for New England customers could increase substantially.

Our simulations showed that production costs would have fallen $38 million in 2006 and 2007. These production cost savings, while not insignificant, naturally tend to be smaller than consumer net savings. Better coordination of flows between regions would not affect the bulk of the generators operating in both systems. In most cases, a few higher-cost generators in the higher-price region would be displaced by a few lower-cost generators in the lower-priced region. Hence, the producer cost effects are smaller than the price effects.

D. Conclusions and Recommendations

Efficient use of transmission interfaces between regions allows customers to be served by lower-cost external resources. New England imports large amounts of power from Quebec and New Brunswick, which reduces wholesale power costs for electricity consumers in New England. Power flows in both directions between New England and New York, depending on market conditions in each region.

Our evaluation of external transactions between New York and New England indicates that scheduling by market participants does not fully utilize the external interfaces or achieve all of the potential benefits available from inter-regional trading. Improving the efficiency of flows between regions is particularly important during shortages or very high-priced periods because modest changes in the physical interchange can substantially affect the market outcomes in both New England and New York. Over the past several years, efforts have been made to reduce barriers to scheduling external transactions. Although the external transaction scheduling process is functioning properly and scheduling by market participants tends to improve convergence, significant opportunities remain to improve scheduled interchange between regions.

Proposals have been made to allow the physical interchange to be adjusted within an hour when prices diverge at the interface between the two markets. By reducing scheduling lead times, such a change would facilitate more efficient interchange and reduce inefficiencies caused by poor convergence. Moreover, better arbitrage would cause prices in both regions to be less volatile and lower overall.

Elimination of remaining barriers to market participant scheduling between regions, while desirable, would not achieve full utilization of the external interfaces. Uncertainty, imperfect information, and a lack of coordination limit the ability of market participants to arbitrage fully the prices between regions. Hence, we continue to recommend that the ISOs develop provisions to coordinate the physical interchange between New York and New England in real-time. Some have argued that this would constitute involving the ISOs in the market, but this is not the case. The ISOs would rely upon bids and offers submitted by participants in each market to establish the optimal interchange between the markets in the same way that they establish optimal power flows across each transmission interface inside both markets.

IV. Financial Transmission Rights

A key function of LMP markets is to set efficient energy prices that reflect the economic consequences of binding transmission constraints. These prices guide the short-term dispatch of generation and establish long-term economic signals that govern investment in new generation and transmission assets. Hence, a primary focus of this report is to evaluate locational marginal prices and associated congestion costs.

Congestion costs are incurred in the day-ahead market based on the modeled transmission flows resulting from the day-ahead energy schedules. These costs result from the difference in prices between the points where power is consumed and generated on the network. A price difference due to congestion indicates the gains in trade between the two locations if additional transmission capability were available. Hence, the difference in prices between the locations represents the marginal value of transmission. The differences in locational prices caused by congestion are revealed in the congestion component of the LMP at each location.[22]

Financial Transmission Rights ("FTRs") can be used to hedge the congestion costs of serving load in congested areas or as speculative investments for purchasers who forecast higher congestion revenues between two points than the cost of the associated FTR. An FTR entitles a participant to payments corresponding to the congestion-related difference in prices between two locations in a defined direction. For example, a participant that holds 150 MW of FTRs from point A to point B is entitled to 150 times the locational energy price at point B less the price at point A (a negative value means the participant must pay) assuming no losses. Hence, a participant can hedge the congestion costs associated with a bilateral contract if it owns an FTR between the same receipt and delivery points as the bilateral contract.

Through the auctions it administers, the ISO sells FTRs with one-year terms ("annual FTRs") and one-month terms ("monthly FTRs"). The annual FTRs allow market participants greater

22 The congestion component of the LMP represents the difference between the marginal cost of meeting load at that location versus the marginal cost of meeting load at a reference location, not including transmission losses.

certainty by allowing them to lock-in congestion hedges further in advance. The ISO auctions 50 percent of the forecasted capacity of the transmission system in the annual auction, and all of the remaining capacity in the monthly auctions.[23] FTRs are auctioned separately for on-peak and off-peak hours.[24]

In this section, we assess two aspects of the performance of the FTR markets. First, we evaluate the net payments to FTR holders. The net payments to FTRs holders decreased approximately 15 percent from 2007 to 2008 due to a modest decline in congestion on paths corresponding to the FTRs in the day-ahead market. Payments to FTR holders are funded by the congestion revenue collected by the ISO. In 2008, the congestion revenue collected by the ISO was sufficient to satisfy 100 percent of the obligations to FTR holders (referred to as the "target payment amount"). This is an improvement over 2007 when FTR holders were paid an average of 94 percent of the target payment amount.

Second, we compare FTR prices with congestion prices in the day-ahead and real-time markets. Since FTR auctions are forward financial markets, FTR prices should reflect the expectations of market participants regarding congestion in the day-ahead market. In 2008, FTR prices were reasonably consistent with congestion values in the day-ahead and real-time markets given the volatility of congestion patterns. Furthermore, the consistency of FTR prices with congestion values improved substantially from the annual auction to the monthly auctions as market participants gained additional information about market conditions.

A. Congestion Revenue and Payments to FTR Holders

As discussed above, the holder of an FTR from point A to point B is entitled to a payment equal to the value of the congestion between the two points. The payments to FTR holders are funded from the congestion revenue fund, which is primarily generated from congestion revenue

---

23 In the annual auction the ISO awards FTRs equivalent to 50 percent of the predicted power transfer capability of the system, and in the monthly auctions the ISO awards FTRs equivalent to 100 percent of the remaining predicted power transfer capability after accounting for planned transmission outages. See generally, the *ISO New England Manual for Financial Transmission Rights*, Manual M-06.

24 On-peak hours include hours-ending 7 to 22, Monday through Friday. Off-peak includes all other hours.

collected in the day-ahead market. The congestion revenues are collected in the following manner:

- Day-ahead congestion revenue is equal to the megawatts scheduled to flow across a constrained interface times the shadow price (i.e., the marginal economic value) of the interface.
- Real-time congestion revenue is equal to the change in scheduled flows (relative to the day-ahead market) across a constrained interface times the real-time shadow price of the interface.
  - ✓ Consequently, when real-time scheduled flows are lower than day-ahead scheduled flows across an interface that is constrained in the real-time market, it results in *negative* congestion revenue.[25]
  - ✓ When congestion revenue collected by the ISO is not sufficient to satisfy the targeted payments to FTR holders, it implies that the quantities sold in the FTR auctions exceeded the capability of the transmission system.

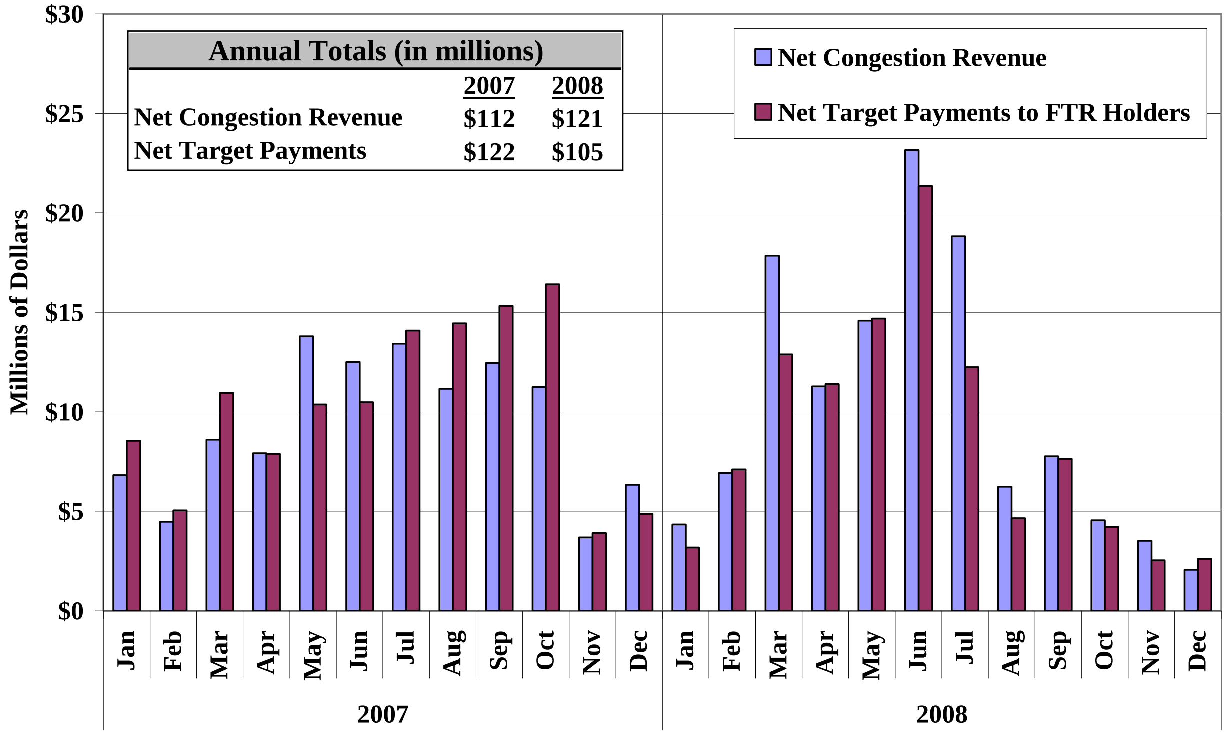
If the ISO does not collect enough congestion revenue to pay the targeted amounts to FTR holders during a month, the unpaid amounts are accrued until the end of the year. At the end of the year, any excess congestion revenues remaining from months with a surplus are used to pay amounts accrued, plus interest, from months with a shortage. If the end-of-year surplus is less than the total accrued shortfall amounts, the end-of-year payments on shortfall amounts are discounted *pro rata*. If the surplus is greater than the total accrued shortfall amounts, the excess congestion revenues after all FTR entitlements have been satisfied are returned to transmission customers per the tariff.

Figure 10 compares the net congestion revenue collected by the ISO with the net target payments to FTR holders in each month of 2007 and 2008. Net congestion revenue includes the sum of all positive and negative congestion revenue collected from the day-ahead and real-time markets. Net target payments to FTR holders include the sum of all positive target payments to FTR

25 For example, suppose 100 MW is scheduled to flow across a particular interface in the day-ahead market, and the interface is constrained when 90 MW is scheduled to flow across it in the real-time market (due to a reduction in capacity after the day-ahead market). If the real-time shadow price of the constraint is $50 per MWh, the 10 MW flow reduction from the day-ahead to the real-time market will result in negative $500 of congestion revenue.

holders and all negative target payments (i.e., payments from FTR holders). The table in the figure reports these quantities by year.

Figure 10: Congestion Revenue and Target Payments to FTR Holders 2007 - 2008



From 2007 to 2008, the figure shows that net congestion revenue rose modestly, from $112 million in 2007 to $121 million in 2008, while net target payments to FTR holders declined from $122 million in 2007 to $105 million in 2008. Congestion was lower in the last two years than in previous years due to a number of factors, including transmission additions in Southwest Connecticut and Boston, as well as the expiration of PUSH bidding rules. The patterns of congestion are evaluated in greater detail in Section B.

The figure shows that the number of months when net congestion revenues were less than the net target payments to FTR holders declined from eight months in 2007 to just two months in 2008. As a result, the net congestion revenues were sufficient to fund the net target payments in 2008. Conversely, in 2007, net congestion revenues were not sufficient to fund the net target payments, so positive target payments to FTR holders were reduced to the sum of (a) net congestion revenues and (b) the payments from FTR holders (i.e., negative target payments). In 2008, the ISO collected $121 million in net congestion revenue and $61 million from FTR holders that were obligated to make payments to the ISO. The total collection of $182 million was $17 million more than the $165 million needed to cover positive targeted payments to FTR holders, so FTR holders received 100 percent of the positive target payments in 2008 compared to 94 percent in 2007.

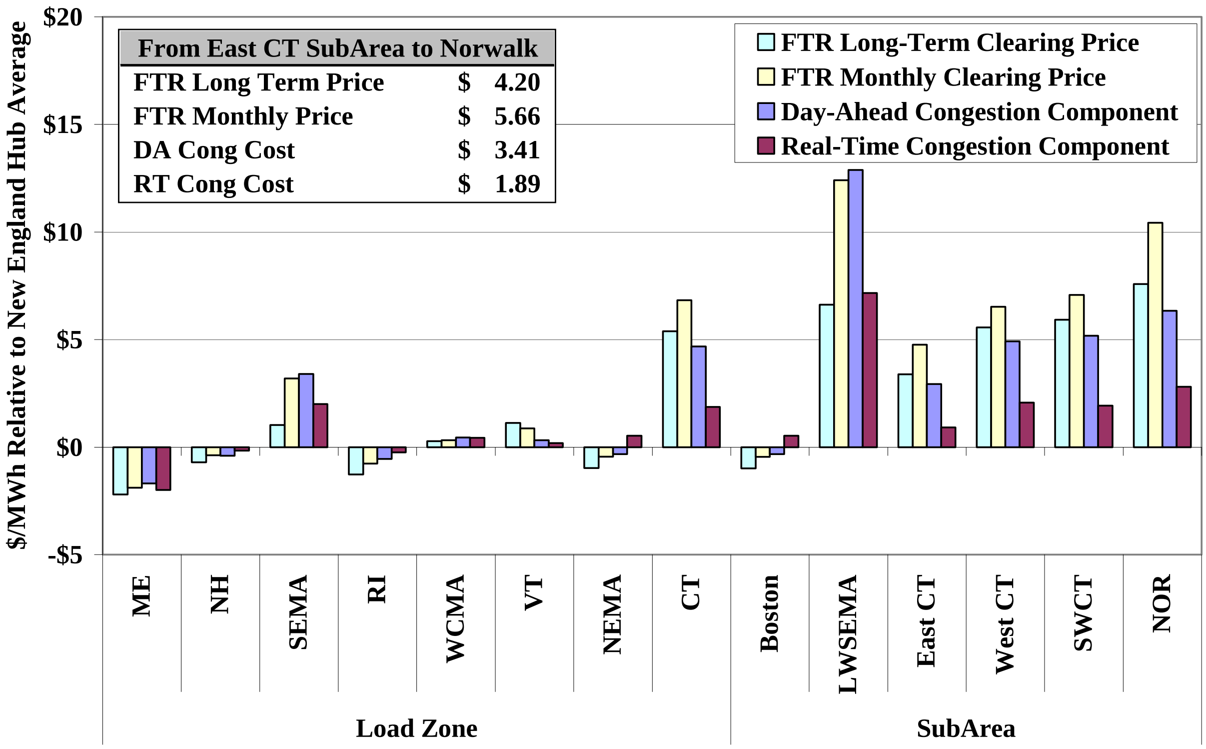
Finally, the figure shows two months (March and July) when congestion revenues substantially exceeded the net target payments to FTR holders. This can occur when the amount of FTRs purchased along a congested transmission corridor is smaller than the actual transfer capability in real-time. For example, in March 2008, there was a spike in export congestion across the primary interface between New England and New York, increasing the LMP at the export node by an average of $14/MWh in the day-ahead market. During congested hours, an average of 685 MW was scheduled to export at this location in the day-ahead market, however, the net amount of FTRs in circulation that sinked at the export node was only 171 MW in on-peak hours and 78 MW in off-peak hours. Consequently, the ISO collected substantially more congestion revenue from exporters than it paid to FTR holders.

B. Congestion Patterns and FTR Prices

In this section, we compare the FTR prices to the congestion prices in the day-ahead and real-time markets. FTR auctions take place in the prior month (for monthly auctions) or at the end of the preceding year (for annual auctions). Prices in the FTR auctions reflect the expectations of market participants regarding congestion in the day-ahead market. As expected, FTR prices from the annual auction were less accurate predictions of day-ahead congestion than the FTR prices from the monthly auction. Furthermore, the FTR prices from the annual auction were more consistent with congestion patterns in the previous year than FTR prices from the monthly auction.

Figure 11 shows day-ahead and real-time congestion prices and FTR prices for each of the eight New England load zones and six sub-areas of interest in 2008. The congestion prices shown are calculated for on-peak hours relative to the New England Hub. Hence, if the congestion price in the figure indicates $4 per MWh, this is interpreted to mean the cost of congestion to transfer a megawatt-hour of power from the New England Hub to the location averaged $4 per MWh during on-peak hours. The congestion price difference between any two points shown in the figure is the congestion price at the sink location less the congestion price at the source location. For example, a -$2.50 per MWh FTR price for Maine and $10 per MWh FTR price for Connecticut would indicate a total price for an FTR from Maine to Connecticut of $12.50 per MWh. Aside from the eight load zones, the figure shows prices for Boston, Lower SEMA, and four areas within Connecticut. Connecticut is divided into: East Connecticut, the portion of West Connecticut that excludes Southwest Connecticut, the portion of Southwest Connecticut that excludes Norwalk-Stamford, and Norwalk-Stamford. Also, the table in the figure provides an example of the FTR prices and congestion prices from East Connecticut to Norwalk-Stamford.

Figure 11: FTR Auction Prices vs. Day-Ahead and Real-Time Congestion Locational Averages Shown Relative to New England Hub Price Weekdays 6 AM to 10 PM, 2008



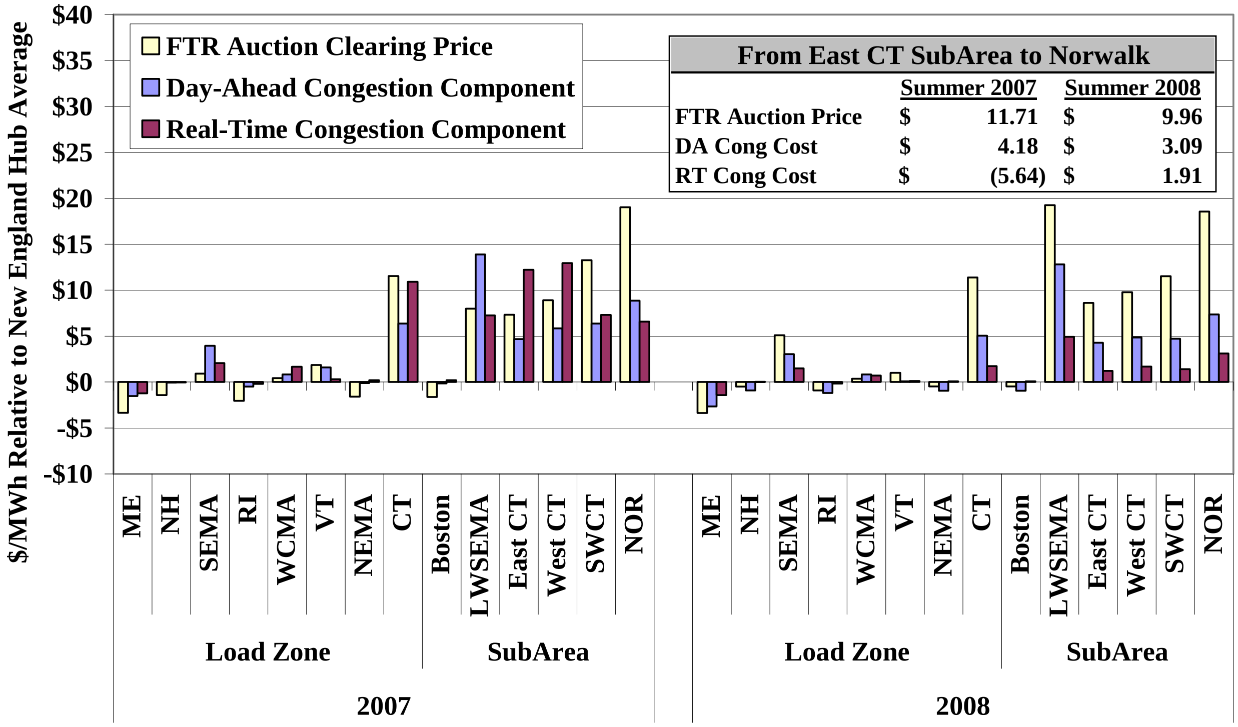
For each location, the figure shows the forward auction prices in chronological order, leading up to real-time from left-to-right. The annual FTR auction occurs first, then the monthly FTR auction, and then the day-ahead market auction. In most areas, monthly FTR prices were relatively consistent with congestion prices in the day-ahead market, while annual FTR prices were less closely correlated with day-ahead congestion prices. However, the monthly FTR prices tended to overestimate congestion into and within Connecticut, while the annual FTR prices were more accurate. The figure also shows that congestion into Boston was virtually non-existent due to recent transmission upgrades into the area.

The annual FTR price from the New England Hub to Lower SEMA was substantially lower than day-ahead congestion prices, while the monthly FTR price was similar to day-ahead congestion prices. This pattern is expected because market participants face greater uncertainty and have less information in the annual auction regarding likely congestion levels than at the time of the monthly auctions. Congestion into Lower SEMA increased from 2007 to 2008, partly due to the increased spread between the prices of natural gas and residual oil, coupled with the heavy reliance of generation in Lower SEMA on residual oil.[26] Since most market participants did not foresee (at the time of the annual auction for 2008 FTRs) the increased spread between the prices of natural gas and residual oil, the annual FTR clearing price from the New England Hub to Lower SEMA was lower than the average day-ahead congestion price in 2008.

The next analysis presents the same results for the summer months in 2007 and 2008. The analysis focuses on the summer because system peaks generally occur in the summer due to cooling demand and the transmission system is under the greatest stress. In addition, higher summer loads generally result in higher congestion prices and greater financial risks for market participants, making FTRs most valuable during the summer. Figure 12 shows the average monthly FTR clearing prices, day-ahead congestion price, and real-time congestion price for the peak hours during the summer season for the same locations as the previous figure.

26 ISO New England, "2008-2017 Forecast Report of Capacity, Energy, Loads, and Transmission (CELT) Report," April 2008.

Figure 12: FTR Auction Prices vs. Day-Ahead and Real-Time Congestion Locations Shown Relative to New England Hub Average Price Weekdays 6 AM to 10 PM, June to August, 2007 - 2008



In the summer of 2008, monthly FTR auction prices exceeded the day-ahead congestion prices from the New England Hub into the areas in Connecticut and into Lower SEMA. This suggests that market participants' expectations in the monthly auctions were substantially higher than the expected congestion in the day-ahead market. Similarly, the day-ahead congestion prices exceeded the real-time congestion prices into those areas by a significant margin, suggesting that participants in the day-ahead market were expecting congestion to affect real-time prices more than it did in the summer of 2008. The monthly FTR auction prices correctly predicted that there would be very little day-ahead or real-time congestion into the Boston area.

Given the volatile nature of congestion patterns, we found that FTRs were valued reasonably well in the FTR auctions. As expected, the monthly auctions generally exhibited more accurate valuations than the 12-month auction. Thus, the FTR market showed signs of adapting to changes in patterns of day-ahead congestion during the study period.

V. Reserve Markets

This section evaluates the operation of the reserve markets, which consist of a real-time reserve market and a forward reserve market. The real-time reserve market has system-level and locational reserve requirements that are integrated with the real-time energy market. The real-time market software co-optimizes the scheduling of reserves and energy. This enables the real-time market to reflect the redispatch costs that are incurred to maintain reserves in the clearing prices for both energy and reserves. Energy-only markets (i.e., markets that do not co-optimize energy and reserves) do not recognize the economic trade-offs between scheduling a resource for energy rather than reserves. It is particularly important to consider such trade-offs during tight operating conditions because efficient scheduling reduces the likelihood of a reserve shortage. When available reserves are not sufficient to meet the requirement, the real-time model will be short of reserves and set the reserve clearing price at the level of the Reserve Constraint Penalty Factor ("RCPF").

The forward reserve market enables suppliers to sell reserves into a forward auction on a seasonal basis. Similar to the real-time reserve market, the forward reserve market has system-level and locational reserve requirements. Suppliers that sell in the forward auction satisfy their forward reserve obligations by providing reserves in real-time from online resources with unused capacity or offline resources capable of starting quickly (i.e., fast-start generators). The forward reserve market is intended to attract investment in capacity that is able to provide reserves at relatively low cost, particularly fast-start generation.

This section evaluates the following aspects of the reserve markets:

- Real-time reserve market;
- Reserve constraint penalty factors;
- Local reserve zones; and
- Forward reserve market.

The final part of this section provides a summary of our conclusions and recommendations regarding the reserve markets, and it describes several steps taken by the ISO-NE to enhance the efficiency of the reserve markets.

A. Real-Time Reserve Market

1. Real-Time Reserve Requirements

The real-time market is designed to satisfy the system's reserve requirements, including locational requirements to maintain minimum reserve levels in certain areas. There are four geographic areas with real-time reserve requirements: Boston, Southwest Connecticut, Connecticut, and the entire system (i.e., all of New England). In addition to the different locations, the reserve markets recognize three categories of reserve capacity: 10-Minute Spinning Reserves ("TMSR"), 10-Minute Non-Spinning Reserves ("TMNSR"), and 30-Minute Operating Reserves ("TMOR").

Sufficient reserves must be held in the New England reserve zone to protect the system in case contingencies (e.g., generator outages) occur. The ISO must hold an amount of 10-minute reserves (i.e., TMSR plus TMNSR) equal to the largest generation contingency on the system, which averaged 1,322 MW in 2008. Based on system conditions, the operator determines how much of the 10-minute reserve requirement to hold as TMSR. ISO-NE held an average of 41 percent of the 10-minute reserve requirement in the form of TMSR during intervals with binding 10-minute spinning reserve constraints in 2008.[27]

The ISO must hold an amount of 30-minute reserves (i.e., TMSR plus TMNSR plus TMOR) equal to the largest generation contingency on the system plus half of the second-largest contingency on the system. The 30-minute reserve requirement averaged 1,907 MW in 2008. Since higher quality reserves may always be used to satisfy requirements for lower quality products, the entire 30-minute reserve requirement can be satisfied with TMSR or TMNSR.

27 The 10-minute spinning reserve requirement was binding in 1.8 percent of the intervals in 2008.

In each of the three local reserve zones, the ISO is required to schedule sufficient resources to maintain service in case the two largest local contingencies occur within a 30-minute period, resulting in two basic operating requirements. First, the ISO must dispatch sufficient energy in the local area to prevent cascading outages if the largest transmission line contingency occurs. Second, the ISO must schedule sufficient 30-minute reserves in the local area to maintain service if a second contingency occurs after the largest transmission line contingency. Alternatively, the local 30-minute reserve requirement can be met with 10-minute reserves or by *importing* reserves from the rest of New England, which is accomplished by producing additional energy within the local area in order to unload transmission into the area. Although ISO-NE is not the first RTO to co-optimize energy and reserves in the real-time market, it remains the only RTO to optimize the level of imported reserves to constrained load pockets.

2. Real-Time Reserve Market Design

The real-time market software jointly optimizes reserves and energy schedules. By co-optimizing the scheduling of energy and reserves, the market is able to reflect the redispatch costs incurred to maintain reserves in the clearing prices of both energy and reserves. For example, if a $40 per MWh combined cycle unit is backed down to provide reserves when the LMP is $50 per MWh, the marginal redispatch cost is $10 per MWh and the reserve clearing price will be no lower than $10 per MWh. The marginal system cost used to schedule the reserves and set reserve clearing prices includes both the redispatch cost (if any) and the offer price for the resource. When excess reserves are available without incurring any costs, reserve clearing prices will be $0 per MWh.

Higher quality reserve products may always be used to satisfy lower quality reserve requirements, ensuring that the clearing prices of higher quality products are never lower than the clearing prices of lower quality products. For instance, if TMOR is available to be scheduled at a marginal system cost of $5 per MWh and an excess of TMNSR is available at no cost, the real-time market will fully schedule the TMNSR to meet the 30-minute reserve requirement. If the zero-cost TMNSR is exhausted before the requirement is met, the real-time market will then schedule TMOR and set the clearing prices of TMNSR and TMOR at $5 per MWh.

When multiple reserve constraints are binding, the clearing price of the highest quality product will be the sum of the underlying marginal system costs for each product. For example, suppose the marginal system costs were $3 per MWh to meet the 10-minute spinning reserve constraint, $5 per MWh to meet the 10-minute reserve constraint, and $7 per MWh to meet the 30-minute reserve constraint. In this case, the TMSR clearing price would be $15 per MWh because a megawatt of TMSR would help satisfy all three constraints. Likewise, the TMNSR clearing price would be $12 per MWh because a megawatt of TMNSR would help satisfy two of the constraints.

ISO-NE is the only RTO that includes the level of imported reserves to constrained load pockets in the co-optimization of energy and reserves. Since local reserve requirements can be met with reserves on internal resources or import capability that is not used to import energy, allowing the real-time model to import the efficient quantity of reserves is a substantial improvement over other market designs. This enhancement is particularly important in New England where the market meets a large share of its local area reserve requirements with imported reserves. For example, imported reserves satisfied 72 percent of the Boston requirement and 44 percent of the Connecticut requirement during constrained intervals in 2008.

The marginal system costs that the market incurs to satisfy reserve requirements are limited by RCPFs. There is an RCPF for each real-time reserve constraint. The current RCPFs are:

- $100 per MWh for the system-level 30-minute reserve constraint,
- $850 per MWh for the system-level 10-minute reserve constraint,
- $50 per MWh for the system-level 10-minute spinning reserve constraint, and
- $50 per MWh for the local 30-minute reserve constraints.

These values are differentiated to reflect values of the reserves and the reliability implications of shortages in the various classes of reserves. It is important to remember that these values are additive when there are shortages in more than one class of reserves, which assures efficient energy and operating reserve prices during shortages. Since energy and operating reserves are co-optimized, the shortage of operating reserves is reflected in energy clearing prices.[28] Tight operating conditions can result in a shortage of 30-minute reserves, which leads to reserve clearing prices of $100 per MWh or more and a contribution to the energy prices of $100 per MWh. Alternatively, more severe conditions that result in shortages of both 30-minute and 10-minute reserves would produce 10-minute reserve clearing prices of $950 per MWh or more ($100 plus $850 per MWh) and energy prices exceeding $1000 ($950 plus the marginal price of energy).

Hence, the system-level 10-minute reserve RCPF of $850 per MWh, together with the other RCPFs, would likely result in energy and operating reserve prices close to the New England market's energy offer cap of $1,000 per MWh during sustained periods of significant operating reserve shortages. The use of RCPFs to set efficient prices during operating reserve shortages was recently affirmed by FERC in Order No. 719, which identifies ISO New England's approach to shortage pricing as a model for other ISOs.[29]

When available reserves are not sufficient to meet a requirement or when the marginal system cost of maintaining a particular reserve requirement exceeds the applicable RCPF, the real-time model will be short of reserves and set clearing prices based on the RCPF. For example, if the marginal system cost of meeting a local area reserve requirement was $75 per MWh, the real-time market would not schedule sufficient reserves to meet the local requirement and the reserve clearing price would be set to $50 per MWh. This would require the operator to intervene in order to maintain the full level of reserves in the local area. To reduce the frequency of such operator actions, the ISO has proposed to modify the RCPF for local reserve zones to $250 per MWh, which would be sufficiently high to maintain local reserves under such circumstances.[30] The RCPFs are analyzed in greater detail later in Section B.

---

28 This assumes the operating reserve shortage results from a general deficiency of generating capacity.

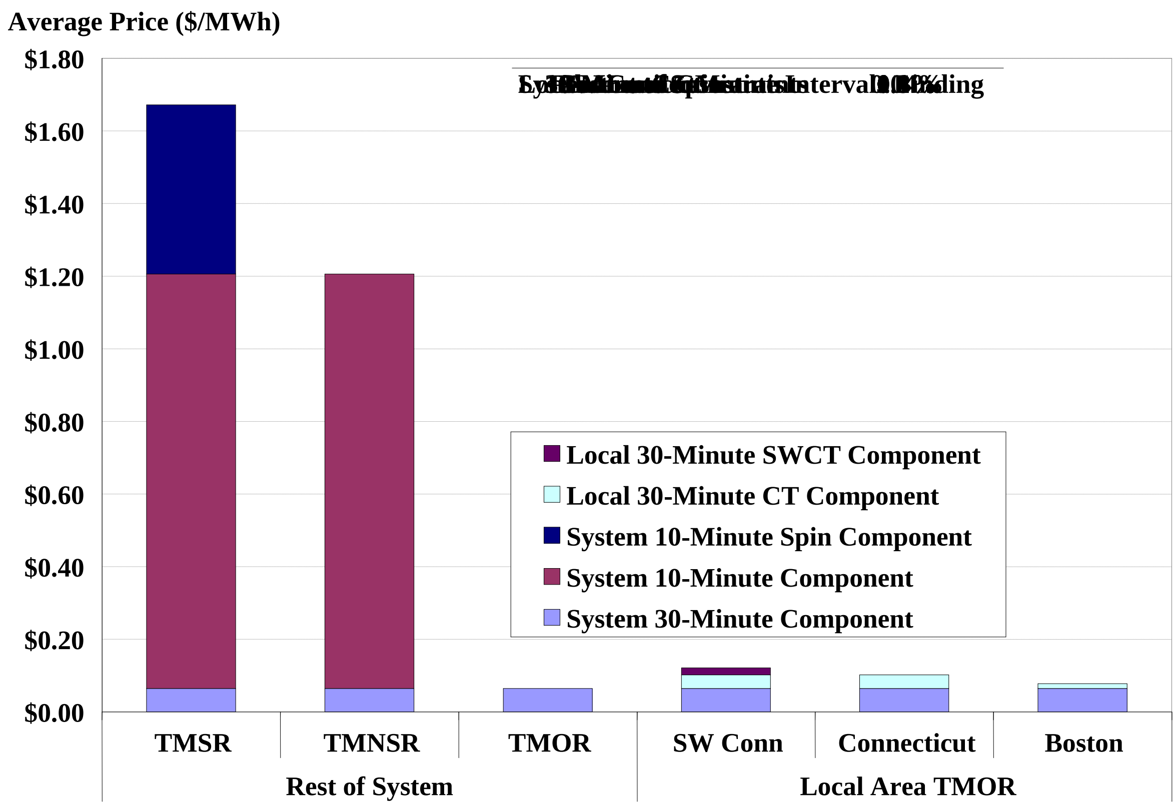
29 *Wholesale Competition in Regions with Organized Electric Markets*, Order No. 719, 73 *Fed. Reg.* 64100 (October 28, 2008), FERC Stats. & Regs. ¶ 31,281 (2008) ("Order No. 719").

30 See NEPOOL Markets Committee meeting materials from May 12 and 13, 2009.

3. Real-Time Reserve Market Results

Figure 13 summarizes average reserve clearing prices during 2008. The left side of the figure shows prices outside the local reserve zones for all three service types. The right side of the figure shows prices in the three local reserve zones for TMOR only. Each price is broken into components associated with the underlying requirements. For example, the Southwest Connecticut price is based on the costs of meeting three requirements: the Southwest Connecticut 30-minute reserve requirement, the Connecticut 30-minute reserve requirement, and the system-level 30-minute reserve requirement. Likewise, the system-level TMSR price is based on the costs of meeting three requirements: the 10-minute spinning reserve requirement, the 10-minute reserve requirement, and 30-minute reserve requirement.

Figure 13: Average Reserve Clearing Prices by Product and Location 2008



Outside the local constrained areas, the average TMSR clearing price was $1.67 per MWh. The $1.67 per MWh for TMSR results from an average of 47 cents per MWh for the 10-minute spinning reserve component, $1.14 per MWh for the 10-minute reserve component, and 6 cents per MWh for the 30-minute reserve component. Although reserve clearing prices were relatively low outside the constrained areas in 2008, they were considerably higher than in 2007 when the average prices for TMSR and TMNSR were both less than 50 cents per MWh. One factor that contributed to the rise in TMSR and TMNSR clearing prices was a decline in TMNSR supply in 2007 and 2008. This decline in TMNSR supply resulted from a performance assessment by the ISO, which found that some peaking resources that had previously offered TMNSR did not meet the applicable performance standard and were only qualified to provide TMOR.

In the local areas, TMOR clearing prices were low in 2007 and declined further in 2008. The highest average TMOR clearing price in 2008 was 12 cents per MWh, which occurred in Southwest Connecticut. This resulted from an average of 4 cents per MWh for the Connecticut 30-minute reserve component, 2 cents per MWh for the Southwest Connecticut 30-minute reserve component, and 6 cents per MWh for the system 30-minute reserve component. TMNSR and TMSR clearing prices are not shown in the local areas because they can also be derived from the underlying requirements. For instance, the average clearing price of TMSR in Southwest Connecticut was $1.73 per MWh. This is composed of $1.67 per MWh for TMSR outside the local areas plus the Southwest Connecticut and Connecticut 30-minute reserve components, which were 2 cents per MWh and 4 cents per MWh.

Average reserve clearing prices were relatively low in 2008 because reserve clearing prices were $0 in the vast majority of real-time intervals. This reflects that there is substantial surplus capacity online during most hours, which is sufficient to meet system-level and local reserve requirements without redispatching generation. Figure 13 indicates that the system-level 10-minute reserve requirement was binding in just 1.5 percent of intervals, implying that the requirement can be met at no cost with surplus capacity in 98.5 percent of intervals. However, when the system-level 10-minute reserve requirement is binding, the clearing price of TMNSR can rise quickly. In 2008, the average TMNSR clearing price was $76 per MWh in intervals when the system-level 10-minute reserve requirement was binding.

B. Reserve Constraint Penalty Factors

In the real-time market, the RCPFs limit the cost that the model may incur to meet the reserve requirements. Consequently, if the cost of maintaining the required level of a particular reserve exceeds the applicable RCPF, the real-time market model will incur a reserve shortage and set the reserve clearing price based on the level of the RCPF.[31] For example, suppose an online generator with a $60 per MWh incremental offer could be backed down to provide reserves when the LMP is $160 per MWh. In this case, the marginal cost to the system of providing reserves from this unit is the opportunity cost of the unit not providing energy at the LMP. This opportunity cost is equal to the difference between the LMP and the incremental offer of the unit, or $100 per MWh in this example ($160 per MWh LMP minus $60 per MWh incremental cost). If the RCPF is $50 per MWh, the market will not back the unit down to provide reserves and the system would be short of reserves since the marginal system cost of doing so ($100 per MWh) exceeds the RCPF ($50 per MWh).

The RCPF levels are important because they determine how the real-time market responds under tight operating conditions. When it is not possible to meet the reserve requirements, the RCPFs prevent the model from incurring extraordinary costs for little or no reliability benefit. However, if RCPFs are not sufficiently high, the model may not schedule all available resources to meet the reliability requirements. In such cases, like the example above, the operator will likely intervene to maintain reserves and significantly affect market prices in the process. Hence, it is important to evaluate the RCPF levels periodically to determine whether modifications are warranted. Based on our evaluation in the 2007 Annual Assessment, we found that the $50 per MW local RCPF was lower than the costs the ISO incurs to maintain the reserves. On this basis, we recommended that the ISO consider increasing the local RCPFs.

31 If only one reserve constraint is binding, the reserve clearing price will be set equal to the RCPF of the reserve that is in shortage. However, if multiple reserve constraints are binding, the reserve clearing price will be set equal to the sum of binding constraint shadow prices. For example, if the market is short of Connecticut reserves and the marginal cost of meeting 30-minute reserves at the system-level is $10 per MWh, the Connecticut reserve clearing price is equal to the sum of the two shadow prices, which is $60 per MWh ($50 per MWh for the Connecticut area and $10 per MWh for the system requirement).

The ISO-NE conducted a review of the RCPF levels that are currently used in the local reserve zones. The ISO-NE concluded that the current local RCPF of $50 per MWh is not sufficiently high to schedule available resources to satisfy the local reserve requirements under some circumstances, leading the operators to intervene to maintain reserves. For this reason, the ISO-NE has proposed to increase the local RCPF levels to $250 per MWh, which it has concluded is high enough to maintain local reserves.

In this section, we evaluate the current RCPF and the proposed RCPF to determine how they compare to the costs that the ISO regularly incurs to maintain local reserves. Specifically, the analysis compares the current and proposed RCPFs to the marginal redispatch costs incurred to maintain local-area reserves in the real-time market in 2008.

1. RCPFs and Real-Time Dispatch

As discussed above, the real-time market may experience a shortage of reserves if the marginal cost of scheduling the available reserves exceeds the RCPF. In such cases, the ISO is required to take additional actions to maintain the required level of reserves if the reserves are available.

There are at least two ways for the ISO to maintain the required level of reserves when the real-time model does not schedule all available reserves. First, the operator can manually adjust the dispatch of certain units in order to provide more reserves in a local area. In the example above, the operator could manually adjust downward the dispatch level of the unit that is capable of providing reserves at an opportunity cost of $100 per MWh. Second, the operator can impose a transmission constraint in the real-time market that forces the model to import a certain amount of reserves (i.e., hold the reserves as import capability on the transmission interface).[32] When possible, the operators use real-time transmission constraints to maintain reserves rather than manual dispatch instructions.

32 This is called a proxy second contingency limit. This type of constraint reduces the limit of the interface below the first contingency limit (the normal limit). The difference between the proxy second contingency limit and the first contingency limit is the amount of reserves that are imported (i.e., held on the interface).

The following analysis in Figure 14 compares the current local RCPF ($50 per MWh) and the proposed local RCPF ($250 per MWh) to the marginal redispatch costs incurred to meet the local-area reserve requirement during 2008. The marginal redispatch costs in this analysis include: (i) the shadow price of the local reserve constraint, which is limited by the RCPF, and (ii) the shadow price of any transmission constraint that is intended to provide imported reserves (i.e., a proxy second contingency constraint). Each bar shows how frequently the marginal redispatch costs were in each range shown on the x-axis.

Figure 14: Marginal Redispatch Costs to Meet Local Reserve Requirements 2008

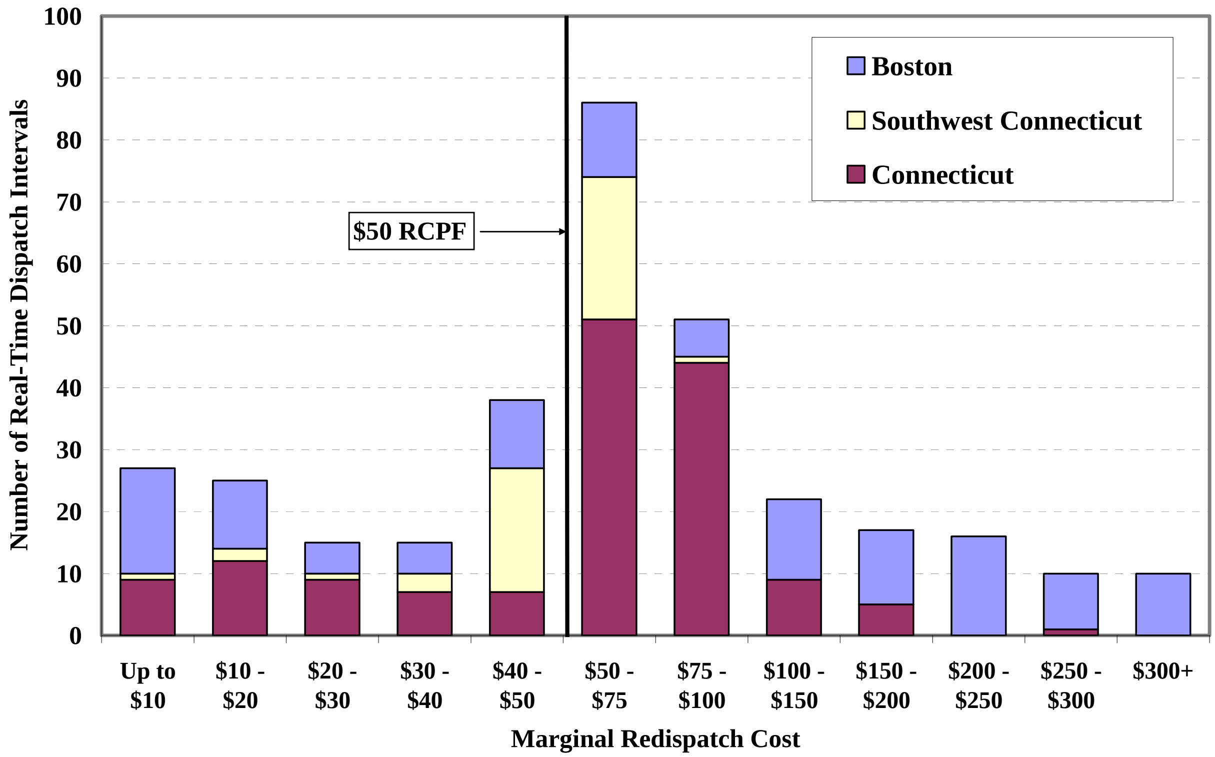


Figure 14 shows that the marginal cost of meeting the local reserve requirements exceeded the current RCPF in more than half of the intervals where redispatch was necessary. The marginal redispatch cost was $50 per MWh or more in 61 percent of the intervals shown for Boston, 71 percent of the intervals shown for Connecticut, and 47 percent of the intervals shown for Southwest Connecticut. This analysis indicates that the RCPF was not sufficiently high to maintain reserves in the local areas in many hours during 2008. As a result, the ISO was compelled to take additional actions to maintain reserves in a substantial number of intervals.

However, the marginal cost of meeting the local reserve requirements rarely exceeded the proposed RCPF of $250 per MWh. The marginal redispatch cost was $250 per MWh or more in just 15 percent of the intervals shown for Boston, 1 percent of the intervals shown for Connecticut, and none of the intervals shown for Southwest Connecticut. The small number of intervals when the redispatch costs exceeded the proposed RCPF occurred May 8 and May 27, 2008. In these intervals, the local reserve requirements were satisfied using proxy second contingency constraints, which led the real-time model to hold reserves as import capability on the transmission interface. This led to higher redispatch costs than would have occurred if the real-time model, operating with a $250 RCPF, had been able to hold the local reserves on offline peaking units.

2. Reserve Constraint Penalty Factors – Conclusions

The previous analysis indicates that the cost of maintaining local reserves frequently exceeds the current local RCPF. This is reflected in real-time operations when the operator must take additional actions to maintain local reserves. There were a substantial number of intervals when the operator attempted to maintain reserves by imposing a proxy second contingency limit, which causes the real-time market model to import reserves. Hence, we support the ISO's proposal to increase the local RCPF to $250 per MWh because this level is consistent with the costs incurred to meet the local-area reserve requirements.

Since the proposed RCPF more accurately reflects the cost of maintaining local reserves in the real-time market, it should increase incentives for more market-based day-ahead commitment in the local areas. This would, in turn, reduce the need for supplemental commitment in the RAA process and shift more of the local reliability costs from NCPC payments to higher market clearing prices. To the extent that a higher RCPF would better reflect the cost of maintaining reserves, increasing the RCPF would improve market-based signals for investment in areas where local reserves are most valuable.

C. Local Reserve Zones

The ISO is required to schedule sufficient resources in local reserve areas to satisfy local second contingency protection requirements (i.e., maintain service in case the largest two local contingencies occur within a 30 minute period). This requires the ISO to schedule local 10-minute and 30-minute reserves and/or import reserves from outside the local reserve area. Three areas are designated as local reserve zones in the real-time reserve market: Boston, Southwest Connecticut, and Connecticut. The ISO schedules reserves in the three local reserve zones by modeling local reserve requirements in the real-time market software and, when sufficient reserves are not available at a cost lower than the RCPF, using proxy second contingency transmission constraints to hold reserves on the transmission interface into the area.

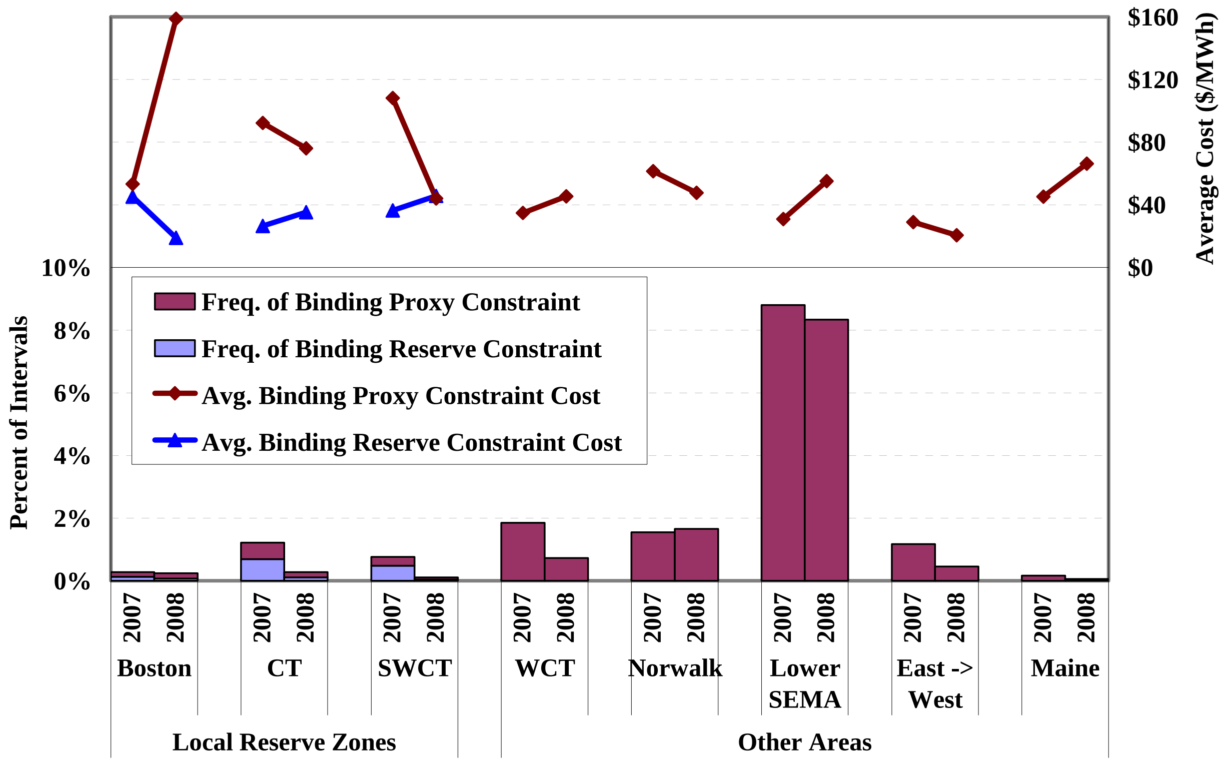
There are other local areas that require the commitment and dispatch of resources to meet local second contingency protection requirements. The ISO's operating procedure for security monitoring discusses the process for maintaining operating reserves in the following eight local areas: the three local reserve zones, as well as West Connecticut, Norwalk-Stamford, Southeast Massachusetts, Western New England, and Maine.[33] In areas requiring local operating reserves that are not designated as Local Reserve Zones, the ISO maintains reserves by modeling proxy second contingency limits in the real-time market software.

The following figure summarizes the actions taken to meet local area reserve requirements in the real-time market in 2007 and 2008. For eight local areas, the figure reports the frequency and average shadow price of binding constraints in the real-time market. Local reserve constraints and proxy second contingency constraints are reported separately.[34]

33 See SOP-RTMKTS.0060.0020, *Monitor System Security*, Section 5.6.

34 The use of proxy second contingency constraints to satisfy local reserve requirements is also discussed in Section B.

Figure 15: Summary of Real-Time Market Constraints to Maintain Local Reserves 2007 - 2008



The figure shows that the ISO redispatched the system to meet local reserve requirements in a very small portion of real-time market intervals in 2007 and 2008. It is notable that maintaining reserves in Norwalk-Stamford, West Connecticut, and Lower SEMA, which are not designated as local reserve zones, is more frequent than maintaining reserves in any of the three designated local reserve zones.

The most efficient way to satisfy local reserve requirements in the real-time market is to explicitly model reserve requirements, rather than to satisfy them by imposing proxy second contingency limits. When local reserve requirements are explicitly modeled in the real-time market software, the software selects the least-cost mix of internal and imported reserves to meet the requirement. Furthermore, the real-time market produces a reserve clearing price, which is a transparent signal of the value of reserves provided by offline and online units. When proxy second contingency limits are modeled in the real-time market software, the software tends to rely too heavily on imported reserves to meet the requirement. This was the case on May 8 and 27, 2008 in Boston as discussed in Section B. Furthermore, the lack of a reserve clearing price in this case causes the market to provide no incentive for suppliers to provide reserves in the local area.

Hence, we recommend that the ISO consider creating additional Local Reserve Zones in order to satisfy local reliability requirements more efficiently in areas that are currently managed using proxy second contingency limits in the real-time market software. Expanding the number of real-time reserve zones would not require the ISO to expand the set of local reserve zones that are modeled in the Forward Reserve Market Auction. Differences already exist between the products that are procured in the Forward Reserve Market and the products that are procured in the Real-Time Reserve Market. The current settlement rules account for such differences.

For example, although there is no TMSR product in the Forward Reserve Market, TMSR is always procured in the Real-Time Market. When suppliers use TMSR capacity to satisfy their TMOR obligations in real-time, they receive a Forward Reserve Payment for TMOR plus the difference between the real-time clearing prices of TMSR and TMOR at their location. Such settlement rules provide suppliers incentives to satisfy their Forward Reserve Obligations with higher quality reserve capacity when it is efficient. Applying similar rules to local reserve zones in the Real-Time Market that are not defined in the Forward Reserve Market would maintain incentives for suppliers to meet their Forward Reserve Obligations in an efficient manner.

D. Forward Reserve Market

Each year, the ISO holds two auctions for Forward Reserves, one for the summer procurement period (the four months from June through September) and one for the winter procurement period (the eight months from October through May). Suppliers that sell in the Forward Reserve auction satisfy their obligations by providing reserves in real-time from online resources or offline resources capable of starting quickly (i.e., fast-start generators). In October 2006, locational requirements were added to the previously existing Forward Reserve Market. Because the cost of maintaining local reserves at reliable levels varies greatly across New England, local requirements improve the economic signals provided by the Forward Reserves Market. For 2008, this section evaluates the results of the forward reserve auctions and examines how suppliers satisfied their obligations in the real-time market.

1. Background on Forward Reserve Market

The ISO purchases several reserve products on behalf of load serving entities in the Forward Reserve Market auction. There are two categories of forward reserve capacity: 10-Minute Non-Spinning Reserves ("TMNSR") and 30-Minute Operating Reserves ("TMOR"). The forward reserve market has five geographic zones: Boston, Southwest Connecticut, Connecticut, Rest of System (i.e., areas outside Connecticut and Boston), and the entire system (i.e., all of New England). With two exceptions, the reserve products sold in the forward reserve market are consistent with the ones sold in the real-time market. First, the forward reserve market has no requirement for 10-Minute Spinning Reserves ("TMSR"). Second, the forward reserve market has a minimum requirement for reserves in Rest of System, while there is no corresponding requirement in the real-time market. The additional reserve zone is intended to ensure that some forward reserves are provided outside local areas.

Forward reserves are cleared through a cost-minimizing uniform-price auction, which sets clearing prices for each category of reserves in each reserve zone. Suppliers sell forward reserves at the portfolio level, which allows them the flexibility to shift where they hold the reserves on an hourly basis. Suppliers also have the flexibility to trade their obligations prior to the real-time market. The flexibility provided by portfolio-level obligations rather than unit-level and bilateral trading enables suppliers to satisfy their obligations more efficiently.

Forward reserve obligations may be satisfied in real-time with reserves of equivalent or higher quality. When obligations are met with reserves of equivalent quality, the reserve provider receives the forward reserve payment instead of real-time market revenue based on the reserve clearing price. When obligations are met with reserves of higher quality, the reserve provider receives the forward reserve payment in addition to real-time market revenue based on the difference in clearing prices between the higher and lower quality products. For example, if Boston TMOR obligations are satisfied in the real-time market with Boston TMSR, the reserve provider will receive the forward reserve payment for Boston TMOR plus the revenue from the price difference between Boston TMSR and Boston TMOR.

2. Forward Reserve Auction Results

Forward Reserve auctions are held approximately one-and-a-half months prior to the first month of the corresponding procurement period. For example, the auction for the Winter 2008/09 procurement period (October 2008 to May 2009) was held in August 2008. Prior to each auction, the ISO sets minimum purchase requirements as follows. For the system-level, the TMNSR requirement is based on 50 percent of the forecasted largest contingency, and the TMOR requirement is based on 50 percent of the forecasted second largest contingency.[35] For Rest of System (i.e., areas outside Connecticut and Boston), the effective TMOR requirement is 798 MW.[36] For each local reserve zone, the TMOR requirement is based on the 95th percentile of the local area reserve requirement in the daily peak hour during the preceding two years, adjusted for major changes in the topology of the system or the status of supply resources. Following significant transmission upgrades in local reserve zones, the portion of the TMOR requirement for the affected local reserve zone that must be satisfied with internal resources is reduced to reflect that a portion of the local requirement will be met by External Reserve Support.

In the Forward Reserve Market auction, an offer of a high quality reserve product is capable of satisfying multiple requirements in the auction. In such cases, the higher quality product is priced according to the sum of the values of the underlying products, although this is limited by the $14 per kW-month price cap. For instance, one megawatt of TMNSR sold in Boston contributes to meeting three distinct requirements: the system-level TMNSR requirement, the system-level TMOR requirement, and the Boston TMOR requirement. The Boston TMNSR clearing price equals the system-level TMNSR clearing price (which incorporates the clearing price of the system-level TMOR) plus the difference between the Boston TMOR clearing price and the system-level TMOR clearing price.

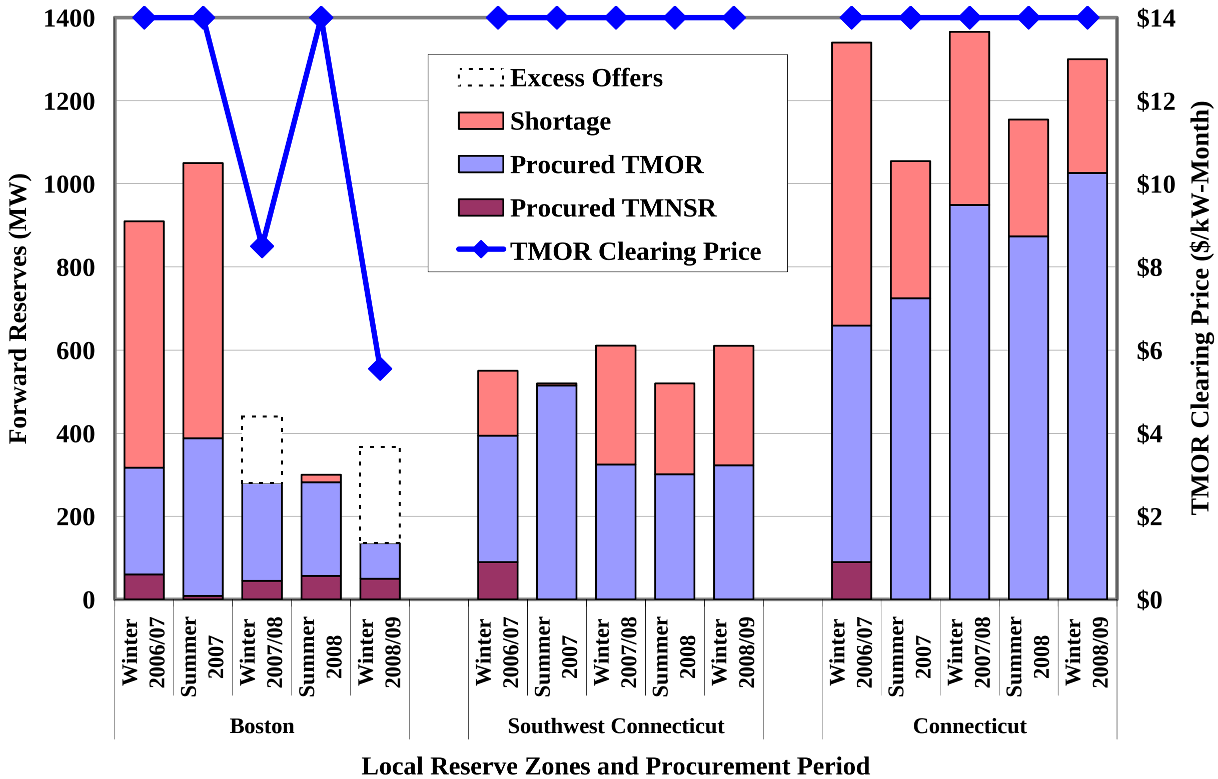
---

35 Usually, the forecasted largest contingency is the HQ Phase I/II Interconnection and the forecasted second largest contingency is the combination of the Mystic 8 and Mystic 9 generating units.

36 The requirement is 600 MW, although this is multiplied by 1.33 to account for the expected performance of offline reserve providers.

The following two figures summarize the quantities purchased in the last five forward reserve auctions towards each requirement. Figure 16 shows auction outcomes for the three local reserve zones, and Figure 17 shows auction outcomes for the system-level and Rest of System requirements. For each local reserve zone in each procurement period, Figure 16 shows the TMOR clearing price, the quantity of TMOR and TMNSR procured, the shortage quantity if the requirement was not met, and the quantity of excess offers if the requirement was met.

Figure 16: Summary of Forward Reserve Auction for Local Areas Procurement for October 2006 to May 2009



The local reserve zone requirement was satisfied in only two of the 15 local reserve procurements shown in Figure 16. A substantial amount of transmission capability was added into the Boston area in 2007, leading the ISO to assume between 900 and 1,065 MW of External Reserve Support in the last three auctions. As a result, the amount of local reserves required from internal Boston resources was reduced to between 135 and 300 MW in the last three auctions. The full Boston requirement was met in the Winter 2007/08 auction, leading the

TMOR price to clear below the cap of $14 per kW-month.[37] The Boston requirement was not satisfied in the Summer 2008 auction when the available supply of fast-start generation was not sufficient to satisfy the local requirement. In the Winter 2008/09 auction, Boston prices dropped considerably as the assumed External Reserve Support increased.

In the 10 local reserve procurements shown in Figure 16 for Connecticut and Southwest Connecticut, the local reserve zone requirement was never satisfied. The shortage quantities ranged from 5 MW for Southwest Connecticut in the Summer 2007 to 681 MW for Connecticut in the Winter 2006/07. The forward reserves procured for Southwest Connecticut are shown both separately and as a subset of the total procurement for Connecticut. Due to recent transmission upgrades into Southwest Connecticut, a large portion of the requirement for Southwest Connecticut will be satisfied by External Reserve Support starting in the Summer 2009 auction.[38] As a result, there will be sufficient internal resources in Southwest Connecticut to satisfy the local requirement in future forward reserve auctions.

Figure 16 shows that a small amount of TMNSR has been sold in Connecticut and Southwest Connecticut, even though there are approximately 200 MW of TMNSR-capable resources there. The low level of TMNSR sales is likely a response to the incentives that arise from the $14 per kW-month price cap. When the local reserve clearing price rises to the price cap, suppliers receive the same compensation for TMNSR and TMOR, even though TMNSR may be more costly to deliver or less easily traded in the bilateral market.

Furthermore, the supplier who sells TMOR in the Forward Reserve Auction will receive a higher real-time settlement than the supplier who sells TMNSR. This is because real-time reserve providers are paid the difference in prices between the product they sold in the Forward Reserve Market and the product they actually provided in real-time. Hence, suppliers with TMNSR-

---

37 However, in the Winter 2007/08 auction, the Boston TMNSR price cleared at the cap of $14 per kW-month, because the combined value of Boston TMOR and system-level TMNSR exceeded the price cap.

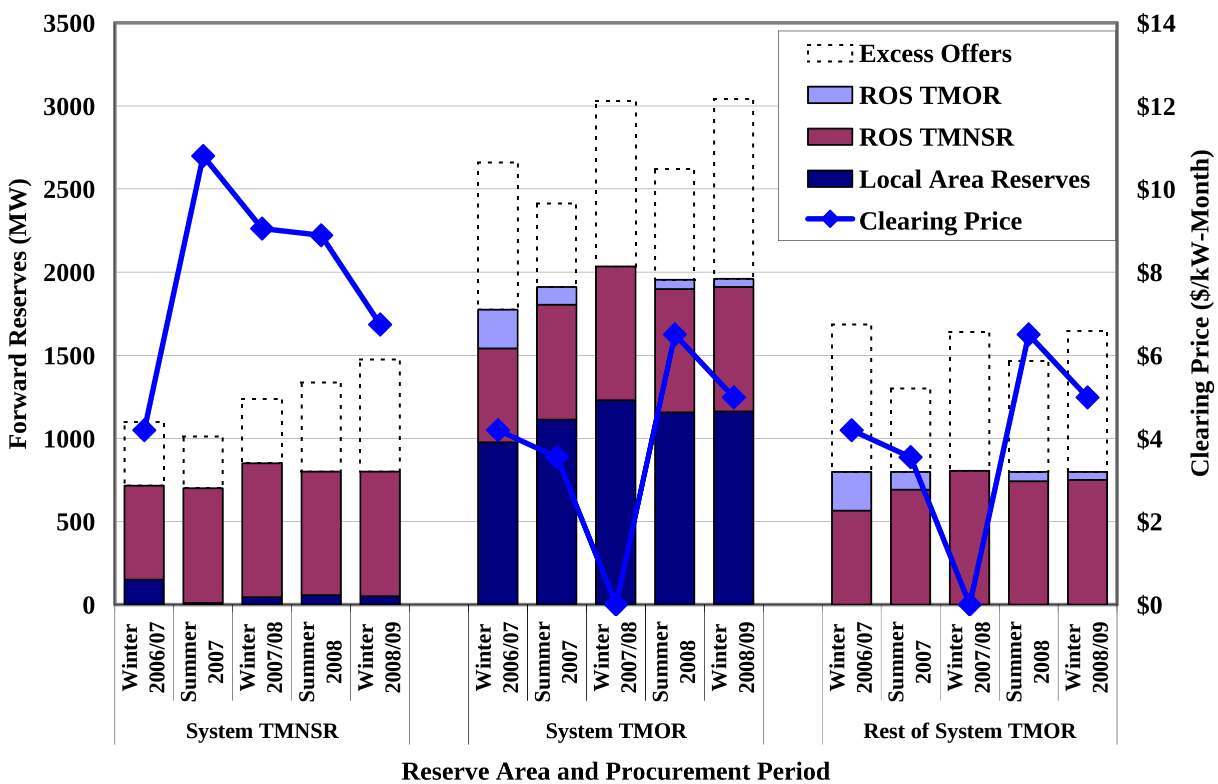
38 In the Summer 2009 auction, all but 25 MW of the 671 MW requirement for Southwest Connecticut will be met with External Reserve Support.

capable resources have a strong incentive to sell TMOR rather than TMNSR in the Forward Reserve Auction when they will receive the price cap of $14 per kW-month in either case.

For similar reasons, the price cap has also discouraged suppliers from selling forward reserves in Southwest Connecticut. Figure 16 shows that the quantity of reserves procured in Southwest Connecticut declined nearly 200 MW after the Summer 2007 auction. This occurred because suppliers with fast-start capacity in Southwest Connecticut began selling Forward Reserves at the Connecticut location. Suppliers with resources in Southwest Connecticut have an incentive to sell at the Connecticut location in the Forward Reserve Auction when they expect to receive the price cap of $14 per kW-month because they cannot receive any additional revenue for selling in Southwest Connecticut.

Figure 17 shows the same analysis for the system-level and Rest of System requirements, showing the prices and products procured outside of the local areas.

Figure 17: Summary of Forward Reserve Auction for Outside Local Areas Procurement for October 2006 to May 2009



Outside of the local reserve areas, the forward reserve requirements were satisfied in each auction. In all five auctions, the system-level TMOR requirement was satisfied by the purchases for other requirements (i.e., no additional costs had to be incurred or purchases made to satisfy the system-wide TMOR requirement). Therefore, the system-wide TMOR price was zero in every auction. However, because all TMOR resources satisfy either a local TMOR requirement or the Rest of System TMOR requirement, suppliers were paid the clearing price for those products rather than zero. The Rest of System TMOR price generally ranged from \$3 to \$7 per MW, but cleared at \$0 in the Winter 2007/08 auction because the requirement was met by procurement for the TMNSR requirement. As noted in the background section, a Rest of System TMOR requirement does not exist in the real-time operating reserve markets, which indicates that TMOR reserves in the local areas are available to respond to congtingencies in the Rest of System area. Given this reliability conclusion, we recommend that the ISO eliminate its Rest of System TMOR requirement in the Forward Reserve Market unless it corresponds to a specific reliability need.

In the Winter 2006/07 auction, TMNSR and TMOR sold at the same price because the TMNSR requirement was met by the combination of TMNSR procured for local area requirements and TMNSR procured to satisfy the Rest of System TMOR requirement. However, in the subsequent four auctions, TMNSR cleared at a premium over TMOR.

Figure 17 shows that a large share of the TMNSR requirement was procured outside of the local areas. For example, just 50 MW of TMNSR was procured in the local areas in the Winter 2008/09 auction, even though approximately 275 MW of TMNSR-capable fast-start capacity exists in the local areas. The low level of TMNSR sales in the local areas is likely a response to the incentives that arise from the \$14 per kW-month price cap. When the local reserve clearing price rises to the price cap, suppliers receive the same compensation for TMNSR and TMOR, providing no incentive to sell TMNSR rather than TMOR. The lack of TMNSR sales in the local areas has resulted in higher clearing prices for TMNSR system-wide. The same incentives also discourage suppliers from selling forward reserves at the Southwest Connecticut location. To address the adverse incentive effects that arise from the price cap, we recommend the ISO evaluate the potential benefits of implementing a tiered price cap. A tiered price cap that allows different price caps for different products could provide suppliers in local areas with better incentives to sell higher-quality forward reserve products than the current market.

3. Forward Reserve Obligations in the Real-Time Market

Forward reserve providers satisfy their obligations in the real-time market by assigning individual resources to provide specific quantities of forward reserves in each hour from 7:00 AM to 11:00 PM, Monday through Friday. Resources assigned to provide forward reserves must be fast-start units or units that are online. These resources must be capable of ramping quickly enough to provide the specified quantity of reserves in 10 minutes for TMNSR and 30 minutes for TMOR. The assigned resources must offer the assigned quantity of incremental energy at a minimum price level.[39] Resources assigned to provide forward reserves forfeit any NCPC payments that they would otherwise receive. Forward reserve providers can arrange bilaterally for other suppliers to meet their obligations, although bilateral trading of obligations between non-affiliated firms was very limited in 2008. Suppliers that do not meet their forward reserve obligations incur a Failure to Reserve Penalty.[40]

There are several types of costs that suppliers consider when assigning units to provide forward reserves. First, suppliers with forward reserve obligations face the risk of financial penalties if their resources fail to deploy during a reserve pick-up.[41] Suppliers can reduce this risk by meeting their obligations with resources that are more reliable. Second, suppliers with forward reserve obligations forego the value of those reserves in the real-time market. For instance,

---

39 This level, known as the "Threshold Price", is equal to the monthly fuel index price posted prior to each month multiplied by a constant of approximately 14.4 MMbtu per MWh. Hence, if the monthly natural gas index price is $8 per MMbtu, it would result in a Threshold Price of approximately $115 per MWh. The month fuel index price is based on the lower of the natural gas or diesel fuel index prices in dollars per MMbtu.

40 The Failure to Reserve penalty is equal to the number of megawatts not reserved times 1.5 times the Forward Reserve Payment Rate, which is the forward reserve clearing price (adjusted for capacity payments) divided by the number of obligation hours in the month.

41 The Failure to Activate penalty is equal to the number of megawatts that does not respond times the sum of the Forward Reserve Payment Rate and the Failure to Activate Penalty Rate, which is 2.25 times the higher of the LMP at the generator's location or the Forward Reserve Payment Rate.

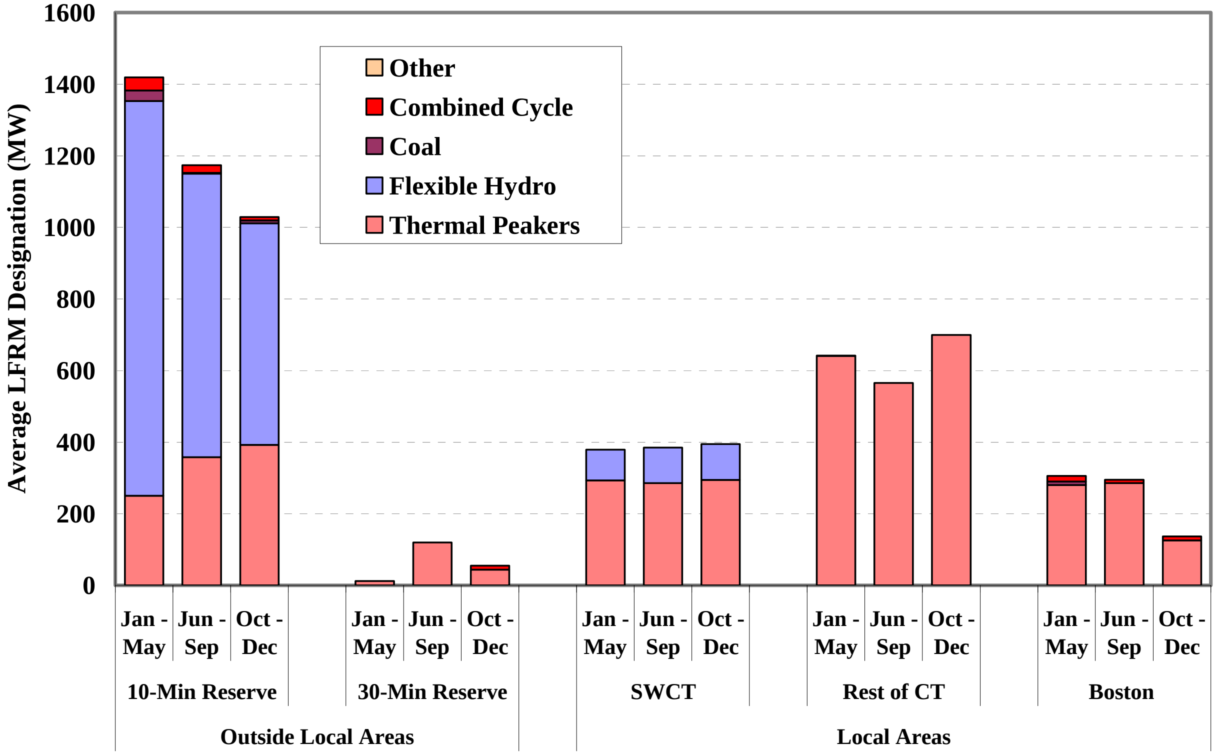
suppose that real-time clearing prices are $10 per MWh for TMOR and $15 per MWh for TMNSR. A supplier that has TMOR obligations would be paid $0 if scheduled for TMOR or $5 per MWh if scheduled for TMNSR. Hence, the supplier would forego the $10 per MWh real-time TMOR price, regardless of whether the supplier is ultimately scheduled for TMOR, TMNSR, TMSR, or energy in the real-time market.

Third, suppliers may forego profitable energy sales as a result of offering incremental energy at the Threshold Price. For instance, suppose the Threshold Price is $130 per MWh and a supplier assigns a generator that has incremental costs of $80 per MWh to provide forward reserves. Because the supplier is required to offer at $130 per MWh, the supplier will not be scheduled to sell energy when the LMP is between $80 per MWh and $130 per MWh. The magnitude of this opportunity cost decreases for units that have high incremental costs (this opportunity cost is zero for units that have incremental costs greater than the Threshold Price).

The previous three cost categories may be incurred by all units that provide forward reserves, but there are additional costs that are only faced by units that must be online to provide reserves. In order to provide reserves from a non-fast-start unit, a supplier may have to commit a unit that would otherwise be unprofitable to commit. This type of cost is zero when energy prices are high and the unit is profitable to operate based on the energy revenues. However, when energy prices are low, the commitment costs incurred by some units may far exceed the net revenue that they earn from the energy market. Because fast-start resources do not face this cost, it is generally most economic to meet forward reserve obligations with fast-start units.

The following analysis evaluates how market participants satisfied their forward reserve obligations in 2008 by procurement period. The figure shows the average amount of reserves assigned in each region by type of resource.

**Figure 18: Forward Reserve Assignments by Resource Type
2008**



The figure indicates that approximately 98 percent of the capacity assigned to provide forward reserves was hydro and thermal peaking capacity capable of providing offline reserves. In some cases, these units were online and providing energy (which is acceptable as long as they offer in accordance with the forward reserve rules). The frequent assignment of fast-start resources to provide forward reserves confirms that it is generally more costly to provide forward reserves from non-fast-start resources.

Combined cycle units were assigned to provide a small portion of the forward reserves in 2008. Combined cycle units are composed of gas turbines and steam turbines where the waste heat from the gas turbines is used to power the steam turbines, thereby increasing the overall efficiency of the unit. Most of the combined cycle units assigned to provide forward reserves in 2008 were ones that are capable of providing offline reserves within 30 minutes.

Several coal-fired steam units were also used to satisfy forward reserve obligations in 2008. Coal units have two characteristics that can make them relatively efficient providers of forward reserves. First, it is nearly always economic to commit coal-fired units so suppliers do not face significant costs from committing them uneconomically. Second, most coal-fired units have a small emergency range that they can use to provide spinning reserves. Production of energy in the emergency range is relatively costly so they do not incur a substantial opportunity cost by offering a small amount of incremental energy at the Threshold Price.

In summary, the preponderance of forward reserves is provided by fast-start units, even in areas where the clearing price rises to the cap of $14 per kW-month. This suggests that many non-fast-start resources do not sell forward reserves because the expected costs of providing forward reserves exceed the price cap. However, non-fast-start units that could provide forward reserves at a cost below the price cap may be discouraged from participating because:

- Units under reliability agreements do not have a financial incentive to participate in the forward reserve market. As these agreements expire, participation in the forward reserve market by non-fast-start capacity may increase.
- Units that are frequently committed for local reliability and receive substantial NCPC payments have disincentives to provide forward reserves because they would be required to forgo the NCPC payments.

There was a hope that the Forward Reserve Market would lower NCPC costs because high-cost units committed for local reliability would sell Forward Reserves. However, this has not occurred.

E. Conclusions and Recommendations

In the real-time market, the scheduling of operating reserves and energy are co-optimized, enabling the real-time model to consider how the cost of energy is affected by the need to maintain operating reserves, and vice versa. ISO-NE is the only electricity market operator to determine in the real-time market the amount of reserves that are held on resources inside a local area versus the amount of reserves that are imported to the area. This innovation helps reduce the overall cost of meeting the local reserve requirements. During reserve shortages, the real-time market sets the reserve clearing prices according to the RCPFs. The use of RCPFs to set clearing prices during reserve shortages provides a robust mechanism for shortage pricing, which was recently endorsed by the FERC in Order No. 719.

In the forward reserve market, clearing prices vary by location, providing signals for investment in capacity that is able to provide reserves at relatively low cost, particularly fast-start generation. Accordingly, we find that 98 percent of the resources assigned to satisfy forward reserve obligations in 2008 were fast-start resources capable of providing offline reserves.

Based on our evaluation of the real-time reserve market, we find that:

- The cost of maintaining operating reserves in local areas frequently exceeds the current local RCPF ($50 per MWh). The RCPF proposed by the ISO ($250 per MWh) would more accurately reflect the cost of maintaining local reserves in real-time prices and improve incentives for market-based day-ahead commitment in the local areas. Increased market-based commitment would help shift some of the local reliability costs from NCPC payments into the market clearing prices.
    - ✓ Hence, we support the ISO's proposal to modify the local RCPFs and set them to levels that are consistent with the costs necessary to meet the local-area reserve requirements.
- There are five areas that are not defined as local reserve zones where the ISO redispatches to maintain local operating reserves. In these areas, local reserve requirements are managed using proxy second contingency transmission limits. It would be more efficient to explicitly model reserve requirements in these areas.
    - ✓ We recommend that the ISO consider creating additional local reserve zones in areas that are currently managed using proxy second contingency transmission limits in the real-time market.

Based on our evaluation of the forward reserve market, we find that:

- In recent forward reserve auctions, prices in the local areas have frequently cleared at the $14 per kW-month price cap. These prices will encourage investment in resources that can provide operating reserves at relatively low cost, such as fast-start generators and qualifying demand response resources.
- Substantial amounts of TMNSR-capable fast-start capacity exist in the local areas, although relatively little has been sold in the forward reserve auctions. This is likely a response to the incentives that arise from the $14 per kW-month price cap. The lack of TMNSR sales in the local areas has resulted in higher clearing prices for TMNSR outside the local areas.
    - ✓ We recommend that the ISO evaluate the potential benefits of implementing a tiered price cap. A tiered price cap (different price caps for different products) would provide suppliers in local areas with better incentives to sell higher-quality forward reserve products in higher value locations.

- The forward reserve market has requirements for TMOR in "Rest of System" and at the system-level. Resources in Boston and Connecticut can satisfy the system-level requirement but not the "Rest of System" requirement.
  - ✓ We recommend that ISO New England eliminate the "Rest of System" requirement given that there is already a system-wide requirement. Such a change would increase the competitiveness of the forward reserve market outside of the local areas and be more consistent with the real-time reserve requirements.

VI. Regulation Market

Regulation is the capability of specially equipped generators to increase or decrease their output on a moment-to-moment basis in response to signals from the ISO. The system operator uses regulation capability to maintain the balance between actual generation and load in the control area. The regulation market provides a market-based system for meeting the system's regulation requirements.

The ISO determines the quantity of regulation capability required to maintain the balance between generation and load based on historical performance and ISO New England, NERC and NPCC control standards. The ISO schedules an amount of regulation capability that ranges from 50 MW to 250 MW depending upon the season, the time of day, and forecasted operating conditions. Historically, the ISO has scheduled 15 to 20 MW more regulation capability in the summer and winter than it has acquired in the spring and fall. During emergency conditions, the ISO may adjust the regulation requirement to maintain system reliability. The ISO periodically reviews regulation performance against the applicable control standards. The high level of performance in recent years has permitted a steady decline in the average quantity of regulation scheduled over the last four years: from 143 MW in 2005 to 134 MW in 2006, 129 MW in 2007, and 121 MW in 2008.

In this section of the report, we evaluate two aspects of the market for regulation: the overall expenses from procuring regulation and the pattern of supply offers from regulation providers.

A. Regulation Market Expenses

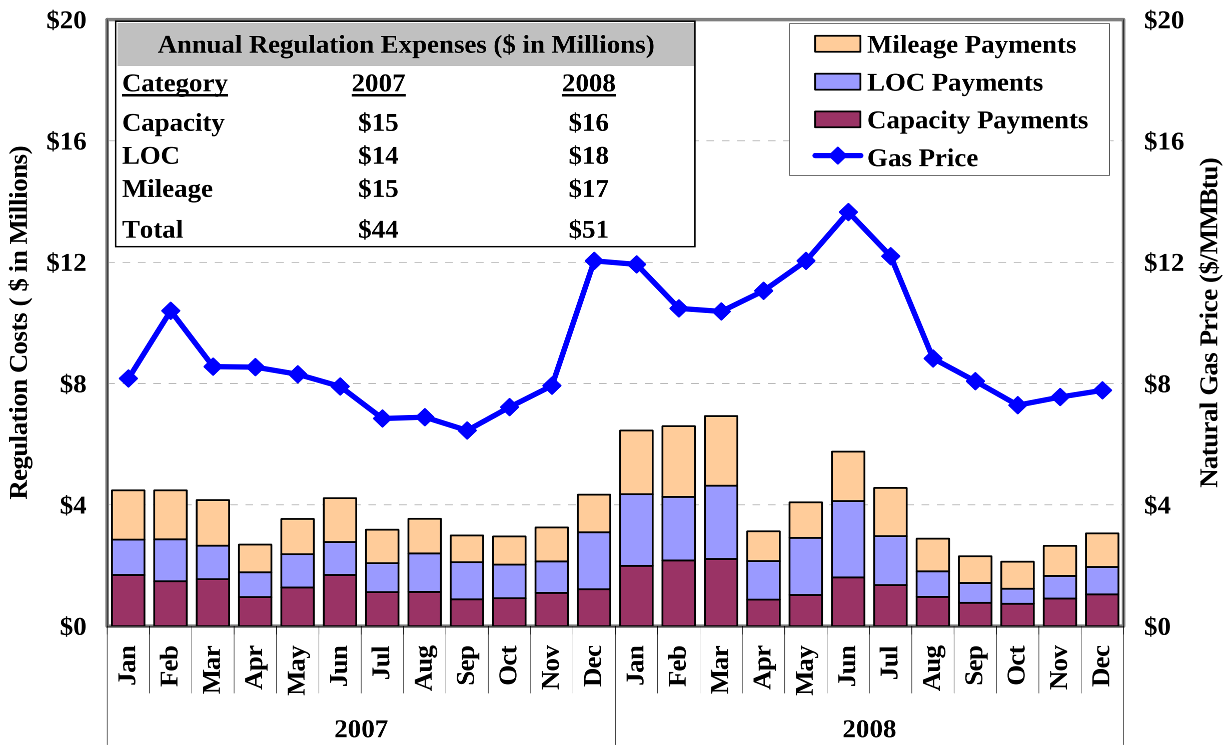
In October 2005, the ISO revised its methods for selecting and paying resources to provide regulation. The current method for selecting resources is described in Section B. Currently, resources providing regulation receive the following payments:[42]

42 In ISO-NE Manual M-11 on Market Operations, the Capacity Payment is called the "Time-on-Regulation Credit", the Mileage Payment is called the "Regulation Service Credit", and the Lost Opportunity Cost Payment is called the "Regulation Opportunity Cost".

- *Capacity Payment* – This equals the Regulation Clearing Price ("RCP") times the amount of regulation capability provided by the resource. The RCP is based on the highest accepted offer price.
- *Mileage Payment* – This is equal to 10 percent of the mileage (i.e., the up and down distance measured in MW) times the RCP. Based on historic patterns of regulation deployment, this formula was expected to generate mileage payments and capacity payments of similar magnitude in the long term.
- *Lost Opportunity Cost ("LOC") Payment* – This is the opportunity cost of not providing the optimal amount of energy when the resource provides regulation service.

A summary of the market expenses for each of the three categories is shown in Figure 19 by month for 2007 and 2008. The figure also shows the monthly average natural gas prices.

Figure 19: Regulation Market Expenses
2007 - 2008



Annual Regulation Expenses ($ in Millions)		
Category	2007	2008
Capacity	$15	$16
LOC	$14	$18
Mileage	$15	$17
Total	$44	$51

This figure shows that each category of expenses accounted for approximately one-third of total regulation expenses in both 2007 and 2008. The total expenses rose from $44 million in 2007 to $51 million in 2008. This 16 percent increase is due partly to the significant increase of natural gas prices in 2008. The figure shows that monthly regulation market expenses were correlated with monthly average natural gas prices in 2007 and 2008.

Input fuel prices can affect regulation market expenses in several ways. First, generators may consume more fuel to produce a given amount of electricity when they provide regulation, leading the costs of providing regulation to be correlated with the price of fuel. Market participants reflect these costs in their regulation offer prices, which directly affect Capacity Payments and Mileage Payments. Second, natural gas-fired combined cycle generators are usually committed less frequently during periods of high gas prices. This reduces the availability of low-priced regulation offers and leads to higher regulation expenses. Third, higher fuel prices normally increase the opportunity costs for units to provide regulation service, which is consistent with the $4 million increase in regulation opportunity cost from 2007 to 2008.

These effects of higher fuel prices translated into changes in offer patterns that explain much of the fluctuations in regulation market expenses in 2007 and 2008. The periods of high expenses during the first three months of 2008 coincided with periods of low offer quantities from online resources and higher offer prices than during other periods. Offer patterns are examined in the following section.

B. Regulation Offer Patterns

Competition should be robust in New England's regulation market in most hours because the amount of capability available in New England generally far exceeds the amount required by the ISO. The regulation market selects suppliers for the upcoming hour with the objective of minimizing consumer payments. Each resource offering to provide regulation is ranked according to the estimated payment it would receive if it were to provide regulation. The model selects the resources with the lowest rank price to provide regulation. The rank price is the sum of the following four quantities:

- *Estimated Capacity Payment* – In the first iteration of the model, this is the offer price of each resource. But since the RCP is set by the highest accepted offer, the subsequent iterations set this equal to the higher of the offer price and the previous iteration's highest priced accepted offer.
- *Estimated Lost Opportunity Cost Payment* – This is the estimated opportunity cost from being dispatched at a level that allows a resource to provide regulation rather than at the most economic dispatch level given the resource's offer prices and the prevailing LMP.

- *Estimated Mileage Payment* – This is equal to the estimated capacity payment.
- *The Look Ahead Penalty* – This is equal to 17 percent of the maximum possible change in the energy offer price within the regulating range. This is included in order to avoid selecting resources that would earn large opportunity cost payments if they were to regulate into a range of their energy offer priced at extreme levels.

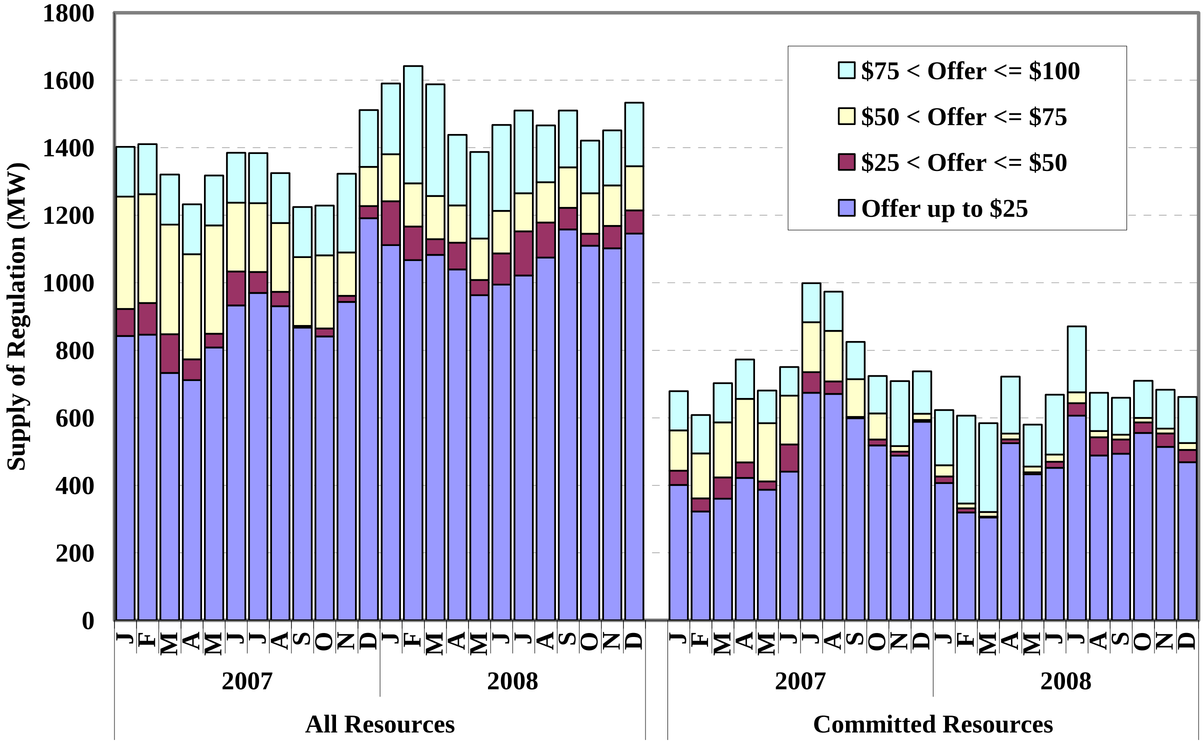
The ranking process iterates until the set of resources selected to provide regulation does not change for two consecutive iterations.[43]

This section evaluates the offer patterns of regulation suppliers. Offline units cannot provide regulation service, so selection of units is limited to units that are online at the time the service is needed. For this reason, we separately examine regulation offers from all resources and from online resources. Figure 20 shows monthly averages of the quantity of regulation offered into the market in 2007 and 2008 for two categories of offers. The left panel in the figure shows offers from all online and offline resources, while the right panel is limited to resources that are actually available to provide regulation. The differing colors on the bars in the chart show the average quantities offered by offer price range.

In Figure 20, the left panel shows that there have been significant variations in the prices and quantities of regulation offers over the past two years. Beginning on December 2007, more regulation capability has been offered into the market. This increase is primarily attributable to one resource that is capable of providing large amounts of regulation capability. The supplier increased its offer quantities from this resource in mid-December 2007, accounting for a 210 MW increase in regulation offers on average. The majority of this additional regulation capability was offered at low prices, which substantially increased the portion of offers below $25 per MW from December 2007 throughout 2008. The right panel shows the changes in offer quantities and prices that more closely correspond to market outcomes since only offers from committed resources can be selected.

43 However, if the RCP rises from one iteration to the next, the model will use the previous iteration to rank resources. For additional details, see Section 3.2.5 of ISO-NE Manual M-11 on Market Operations.

Figure 20: Monthly Average Supply of Regulation
2007 - 2008



On average, approximately half of the regulation offered day-ahead is available to the hourly real-time selection process. Regulation-capable capacity can be unavailable in a given hour because the capacity is on a resource that was not committed for the hour, or because the capacity is held on a portion of a resource that was self-scheduled for energy. More regulation capacity tends to be available during the high-load portion of the day because more units are online. Similarly, more regulation capacity tends to be available during the summer when loads are higher and more generation is committed. Average regulation offers from online resources decreased from 760 MW in 2007 to 670 MW in 2008, which was partly due to elevated gas prices in 2008. Higher gas prices led to less commitment of gas-fired combined cycle units, which provide most of the regulation capability in the market.

During 2007 and 2008, significantly more regulation capability was offered into the market than was actually procured by the ISO. This excess supply generally limits competitive concerns in the regulation market because demand can easily be supplied without the largest regulation supplier. However, sometimes supply is tight in the regulation market when energy demand is high and the regulation market must compete with the energy market for resources. High energy prices during peak-demand periods can lead resources to incur large opportunity costs when providing regulation service, thereby increasing prices for regulation. Likewise, regulation supplies may be tight in low-demand periods when many regulation-capable resources are offline. These conditions can lead to transitory periods of high regulation prices.

C. Conclusions and Recommendations

Overall, the regulation market performed competitively in 2007 and 2008. On average, approximately 700 MW of available supply competes to provide 120 MW of regulation service. The significant excess supply generally limits competitive concerns in the regulation market.

The ISO launched a pilot program on November 18, 2008 that allows new resources with alternative technologies such as energy storage technologies, load response technologies, and other non-generation technologies to participate in the regulation market.[44] The primary objective of this pilot program is to test the performance of alternative control technologies and determine how to utilize them efficiently. The pilot program also evaluates the effects of alternative regulation technologies on the regulation market and other aspects of the wholesale market. The pilot program enrollment is limited to 13 MW of regulating capability, which represents roughly 10 percent of the average hourly regulating requirement. In 2008, one participant enrolled in the pilot program and provided 1 MW regulating capability. This type of program is intended to help integrate new technologies in the wholesale market in order to improve market performance and lower costs to consumers in the long-run.

In the long-run, we recommend that the ISO evaluate potential market design changes that would enhance the performance of the regulation market. Given the complex interaction of the regulation market with the energy market, particularly with respect to commitment decisions made in the day-ahead market, we recommend that the ISO consider day-ahead and real-time regulation markets that are co-optimized with the energy market.

44 See “Appendix J – Alternative Technologies Regulation Pilot Program” of “Market Rule 1” for detail.

VII. Real-Time Pricing and Market Performance

The goal of the real-time market is the efficient procurement of the resources required to meet the reliability needs of the system. To the extent that reliability needs are not fully satisfied by the market, the ISO must procure needed resources outside of the market process. Whenever possible, operations should be performed in a manner that results in efficient real-time price signals. This is because efficient real-time price signals encourage competitive conduct by suppliers, efficient participation by demand response, and investment in new resources or transmission where it is needed most. Hence, it is beneficial to regularly evaluate whether the market produces efficient real-time price signals.

In this section, we evaluate several aspects of the market operations related to pricing and dispatch in the real-time market in 2008. This section examines the following areas:

- Frequency of price corrections;
- Prices during the deployment of fast-start generators;
- Prices during periods of scarce transmission capability;
- Prices during the activation of real-time demand response; and
- The efficiency of the ex post pricing methodology.

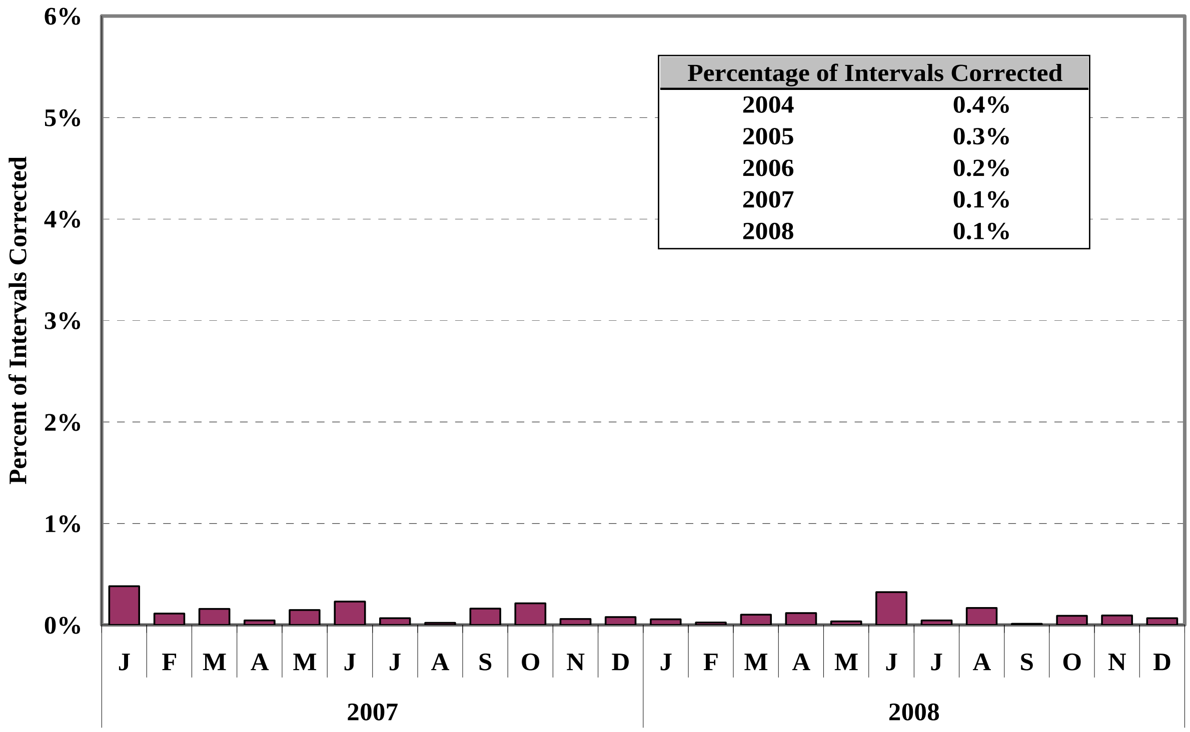
It is also important for the market to set efficient real-time price signals during shortages of operating reserves. This point was recently affirmed by FERC in Order 719, which identifies ISO New England's approach to shortage pricing as an effective method that serves as a model for other ISOs. ISO New England uses Reserve Constraint Penalty Factors ("RCPFs") to set real-time clearing prices during operating reserves shortages. This pricing method is discussed in greater detail in Section V of this report, which evaluates the reserve markets.

A. Real-Time Price Corrections

This sub-section evaluates the rate of real-time price corrections during 2008. Price corrections are necessary to address a variety of issues, including software flaws, operations or data entry errors, system failures, and communications interruptions. Although they cannot be completely eliminated because data and communications errors are an inherent issue in electricity markets, a market operator should aim to minimize price corrections. Substantial and frequent corrections raise ISO and market participant costs and can harm the integrity of the market.

Figure 21 shows the rate of real-time price corrections in New England in each month of 2007 and 2008. The inset table summarizes the annual rate of price corrections in the past five years.

Figure 21: Rate of Real-Time Price Corrections
2007 - 2008



The figure shows that real-time price corrections were infrequent in both 2007 and 2008. The rate was less than 0.4 percent in each month and close to 0.1 percent in most months. The annual rate of price corrections has declined over the past five years. It is notable that approximately 40 percent of the intervals that experienced price corrections in 2008 were due to issues with the real-time software's Dead Bus Logic, which affects the LMPs at very few pricing nodes.[45]

45 Due to equipment outages, the main transmission system may consist of several islands, of which only one is a viable sub-system and the others are considered dead. The market clearing problem is solved only for the viable island and the LMPs are determined in the LMP Calculator. LMPs at dead buses are not directly

Hence, during many of the real-time intervals with price corrections, the effect of the price correction on the market was very limited.

Overall, the frequency of price corrections has been very low over the past five years, supporting the conclusion that the real-time market software for the New England wholesale market has functioned well.

B. Real-Time Commitment and Pricing of Fast-Start Resources

Fast-start generators are capable of starting from an offline status and ramping to their maximum output within 30 minutes of receiving an instruction. This enables them to provide reserves while offline. Areas without significant quantities of fast-start generation must maintain more reserves on online units, which can be very expensive. Another benefit of fast-start units is that they ramp to their maximum output level more quickly than most baseload units, and better enable the system operator to respond rapidly to unexpected changes in load. Such operating conditions can result in especially tight market conditions, making it particularly important to operate the system efficiently and to set prices that accurately reflect the cost of satisfying demand and reliability requirements. This section of the report discusses the difficulties related to efficient real-time pricing when fast-start generators are the marginal supplier of energy in the market. It also evaluates the efficiency of real-time prices when fast-start generators were deployed in merit order in 2008.

1. Treatment of Fast-Start Generators by the Real-Time Dispatch Software

The ISO's real-time dispatch software, called Unit Dispatch Software ("UDS"), is responsible for scheduling generation to balance load and satisfy operating reserve requirements, while not exceeding the capability of the transmission system. UDS provides advance notice of dispatch instructions to each generator for the next dispatch interval based on a short-term forecast of load

available from the LMP Calculator. However, there is need for market settlement purposes to determine the LMPs at dead buses. The algorithm, referred to as LMPc Dead Bus Logic, has been used to facilitate this need.

and other operating conditions.[46] Most commitment decisions are made prior to the operation of UDS in the day-ahead timeframe, so the primary function of UDS is to adjust the output levels of online resources. The only resources that UDS can commit (i.e., start from an offline state) are fast-start generators.[47] Allowing UDS to start fast-start generators is more efficient than relying exclusively on operators to manually commit such units.[48]

When determining dispatch instructions for most online generators, UDS considers only incremental offers. However, for fast-start generators, UDS also considers commitment costs and uses various assumptions regarding the dispatchable range of the generator. The treatment of commitment costs and the dispatchable range have important implications for price setting by the real-time software (i.e., how real-time LMPs are determined). UDS schedules fast-start generators using the following criteria:

- *Offline fast-start generators* – UDS considers commitment costs by adding the amortized start-up and "no-load" offers to the incremental offer.[49] UDS treats the generator as having a dispatchable range from 0 MW to its maximum output level.
- *Online fast-start generators during the minimum run time* – UDS considers only the incremental offer. UDS treats the generator as having a dispatchable range from its minimum output level to its maximum output level.
- *Online fast-start generators after the minimum run time has elapsed* – UDS considers only the incremental offer. UDS treats the generator as having a dispatchable range from 0 MW to its maximum output level.

In the first phase of commitment listed above (when the unit is offline), real-time LMPs usually reflect the full cost of deploying the fast-start generator, partly because UDS considers the no-

46 Generators are usually given instructions 15 minutes in advance, but this can be set higher or lower by the operator.

47 Fast-start units are units that are capable of providing 10-minute or 30-minute non-synchronous reserves and have a minimum run time and a minimum down time of one hour or less.

48 Based on its real-time optimization, UDS recommends that individual fast-start units be started. However, the final decision to start a unit remains with the real-time operator.

49 For example, suppose a 20 MW fast-start unit has an incremental offer of $75 per MWh, a no-load offer of $300/hour, and a start-up offer of $500 (which UDS amortizes over one hour). The average total offer of the unit is $115 per MWh = ($75 per MWh + $300/hour ÷ 20 MW + $500/hour ÷ 20 MW).

load offer and the start-up offer of the generator. Furthermore, UDS allows the fast-start generator to "set price" when the generator is economic to be online by treating the generator as dispatchable between 0 MW and the maximum output level.

However, in the second and third phases of commitment (once the unit is online), real-time LMPs frequently do not reflect the full cost of deploying the fast-start generator, even if the generator is still economic to be online. Since UDS does not consider the start-up and no-load offers, the real-time price-setting logic incorporates only the incremental offer. Furthermore, since the minimum output level of most fast-start generators is within 90 percent of the maximum output level, fast-start generators are frequently dispatched at their minimum output levels where they do not set price during the second phase of commitment. In such cases, the resulting LMP is lower than the incremental offer of the fast-start generator.

The following example illustrates the challenges for efficient pricing when fast-start generators are deployed in merit order. Suppose UDS needs to schedule an additional 15 MW in an import-constrained area and the most efficient way to do so is to start up a fast-start generator with an incremental offer price of $75 per MWh, a no-load offer price of $300/hour, a start-up offer price of $500, a minimum output level of 18 MW, and a maximum output level of 20 MW. In this case, the average total offer of the offline unit is $115 per MWh = ($75 per MWh + $300/hour ÷ 20 MW + $500/hour ÷ 20 MW) when it runs at full output for one hour. This average total offer is used in the price-setting logic during the first phase of commitment.

In the start-up interval, UDS treats the fast-start generator as flexible and schedules 15 MW from the fast-start generator. This generator is the marginal generator and, therefore, sets the LMP at $115 per MWh. Since 15 MW is lower than the minimum output level of the generator, the generator is instructed to produce at its minimum output level.

Once the generating is running (but before its minimum run period has expired) it is no longer possible to schedule 15 MW from the fast-start generator since the minimum output level (18 MW) is enforced. As a result, the fast-start generator is dispatched at 18 MW rather than 15 MW, and the output level of the next most expensive generator is reduced by 3 MW to compensate for the additional output from the fast-start generator. In this case, the fast-start generator is no longer on the margin since it is at its minimum, so the next most expensive generator sets the LMP at a price lower than the incremental offer of the fast-start generator ($75 per MWh).

After the minimum run time elapses, UDS can schedule 15 MW from the fast-start generator if that is most economic, because the minimum output level is not enforced in this phase. In this case, the fast-start generator sets the LMP at its incremental offer of $75 per MWh.

In this example, the fast-start generator is dispatched in merit order, although the full cost of the decision is not reflected in real-time LMPs. The fast-start generator costs $115 per MWh to operate in the first hour and $90 per MWh thereafter, however, the LMP is set to $115 per MWh in the first UDS interval (usually approximately 10 minutes), less than $75 per MWh for the remainder of the first hour, and $75 per MWh thereafter. As a result, the owner of the fast-start unit would receive an NCPC payment to make up the difference between the full offer and the real-time market revenue.

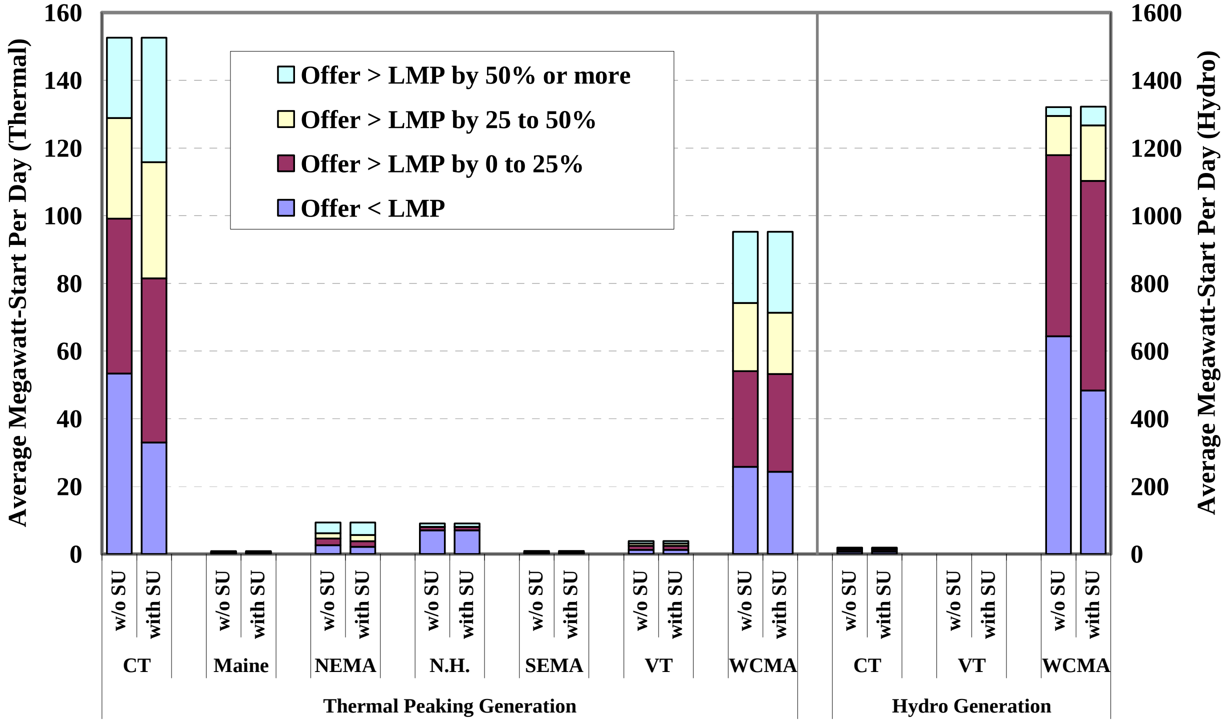
2. Evaluation of Fast-Start Deployments by UDS

The following analysis evaluates real-time LMPs when fast-start generators were deployed in merit order in 2008. Specifically, the figure summarizes the consistency of the average total offer[50] of fast-start generators that were deployed in merit order by UDS with the average real-time LMP over the initial commitment period (usually one hour). When the average real-time LMP is greater than the average total offer, the figure shows the associated capacity in the category labeled "Offer < LMP". However, when the average real-time LMP is less than the average total offer, LMPs do not fully reflect the cost to the system of deploying the fast-start generator. The figure shows such occurrences in three categories according the size of the difference between the average total offer and the average real-time LMP. This comparison is shown separately for each zone for hydro and thermal peaking generators in 2008. To examine

50 The average total offer is the sum of incremental, no-load, and start-up offer components averaged over the economic maximum of the unit for the one hour amortization period.

the significance of the start-up offer in this comparison, the figure shows results including the start-up offer ("with SU") and not including the start-up offer ("w/o SU").

Figure 22: Comparison of Real-Time LMPs to Offers of Fast-Start Generators First Hour Following Start-Up by UDS, 2008



The figure shows that Connecticut accounted for 56 percent of the thermal peaking generation started in 2008, while West-Central Massachusetts accounted for most of the remainder. Across New England, the average amount of thermal peaking generation that was started-up and had an average total offer that was less than the average real-time LMP was 68 MW per day. However, the average total offer exceeded the average real-time LMP for an average of 203 MW per day, and the average total offer exceeded the average real-time LMP by over 50 percent for an average of 67 MW per day. When start-up offers are excluded, the portion of starts that appear to be economic increases. Although thermal peaking generators are deployed in a relatively limited set of hours, they are frequently the marginal source of supply to the system in those hours, making it particularly important to reflect the full cost of their deployment in real-time LMPs.

The figure also shows that flexible hydro generation accounted for the majority of fast-start generation that was started in 2008. This was primarily associated with 1,600 MW of pumped storage capacity located in West-Central Massachusetts. The share of starts where the average total offer was less than the average real-time LMP was higher for the flexible hydro generators than for the thermal peaking generators. This may be due to the fact that they tend to be more flexible than the thermal peaking generators.

3. Conclusions and Recommendations

This analysis shows that fast-start units are routinely deployed in merit order, but real-time prices do not always reflect the underlying costs of the units. This result can lead to inefficient market signals. First, understated real-time prices will reduce the incentives to fully schedule load through the day-ahead market because the day-ahead prices will tend to be higher than the real-time prices if load is fully scheduled day ahead. One consequence of under-scheduling day ahead is that fewer slow-starting units will be committed and the market will rely more heavily on fast-starting resources in real time to meet the incremental load. This pattern reduces the overall market efficiency and increases uplift costs.

Second, it does not provide a correct signal to participants that may import or export power to or from New England. The understated price in this case will lead to fewer net imports and increase New England's deployment of the fast-start resources. Finally, it diminishes the price signals that govern new investment in the long term. Hence, we recommend that the ISO evaluate potential changes in the pricing methodology that would allow the deployment costs of fast-start units to be more fully reflected in the real-time market prices. The Midwest ISO has been engaged in research on this issue, so it may be beneficial for ISO New England to coordinate with the Midwest ISO on this project.

C. Real-Time Pricing During Transmission Scarcity

Local shortages arise when local generation plus the transmission capability into the local area are not sufficient to satisfy demand for energy and reserves in the area. Although such shortages are relatively uncommon, it is important for wholesale markets to set efficient prices that reflect the tight operating conditions during such periods. Efficient prices provide generation and demand response resources incentives to respond to maintain reliability. Efficient prices also provide signals that attract new investment when and where needed.

Under the ISO's current operating procedures, UDS redispatches generation when a transmission limit is binding so that flows do not exceed the limit. UDS can use nearly all available resources to manage transmission flows.[51] On occasion, the marginal redispatch cost (i.e., shadow price) necessary to manage the flow over a transmission facility reaches extraordinary levels (e.g., greater than $10,000 per MWh). When UDS does not have sufficient resources to reduce the flow under the limit, a violation occurs and the shadow price of the constraint is set by the marginal available resource(s).

The current procedures have functioned effectively, allowing the ISO to redispatch the lowest-cost offers available to maintain reliability under tight operating conditions. Yet, it is important to assess the efficiency of price signals under such conditions because prices give generators and demand response resources incentives to respond and attract investment. This section provides an assessment of market outcomes during periods of acute transmission constraints, focusing on:

- The efficiency of the prices when resources are insufficient to manage the constrained transmission line; and
- Whether excessive redispatch costs are incurred to manage the constraints (i.e., redispatch costs that exceed the value of reducing the flow on the constrained transmission line).

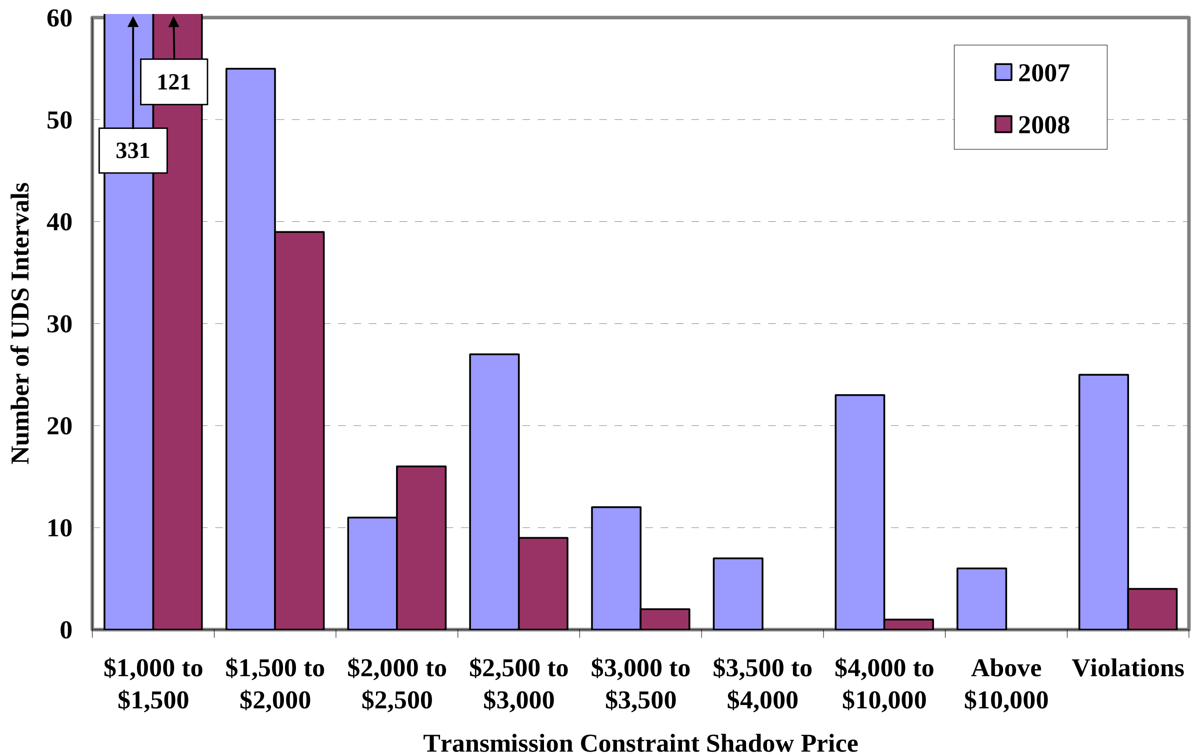
Regarding the second issue, whether such costly redispatch is warranted depends on the reason why the transmission limit was initially imposed. Some transmission limits may be safely violated for an extended period with no substantial effect on reliability while other violations

51 UDS will not use redispatch options that exceed the level of the Transmission Constraint Penalty Factor, although the penalty factor is ordinarily set to a very high level. Also, UDS will not redispatch resources that have a sensitivity factor with a magnitude of less than 2 percent relative to the flow over the transmission facility.

may necessitate immediate curtailment of firm load to maintain reliability.[52] Hence, it would be beneficial to develop procedures that distinguish between these two situations and only incur extraordinary redispatch costs under more severe conditions.

Figure 23 illustrates the significance of periods when transmission congestion was particularly acute in 2007 and 2008.

Figure 23: Frequency of UDS Intervals with High Redispatch Costs 2007 - 2008



The figure shows that the frequency and the cost of acute congestion declined substantially from 2007 to 2008. There were 497 UDS intervals in 2007 when either a transmission constraint shadow price exceeded $1,000 per MWh or a transmission limit was violated. There were only 192 such UDS intervals in 2008, a 61 percent decrease from the preceding year. In 2007, there were 86 intervals when the constraint was resolved with a shadow price exceeding $2,000 per

52 ISO-New England Operating Procedure No. 19 – *Transmission Operations* describes how Normal, Long-Term Emergency, Short-Term Emergency, and Drastic Action Limits are used to develop the limits that are used by UDS to determine dispatch instructions.

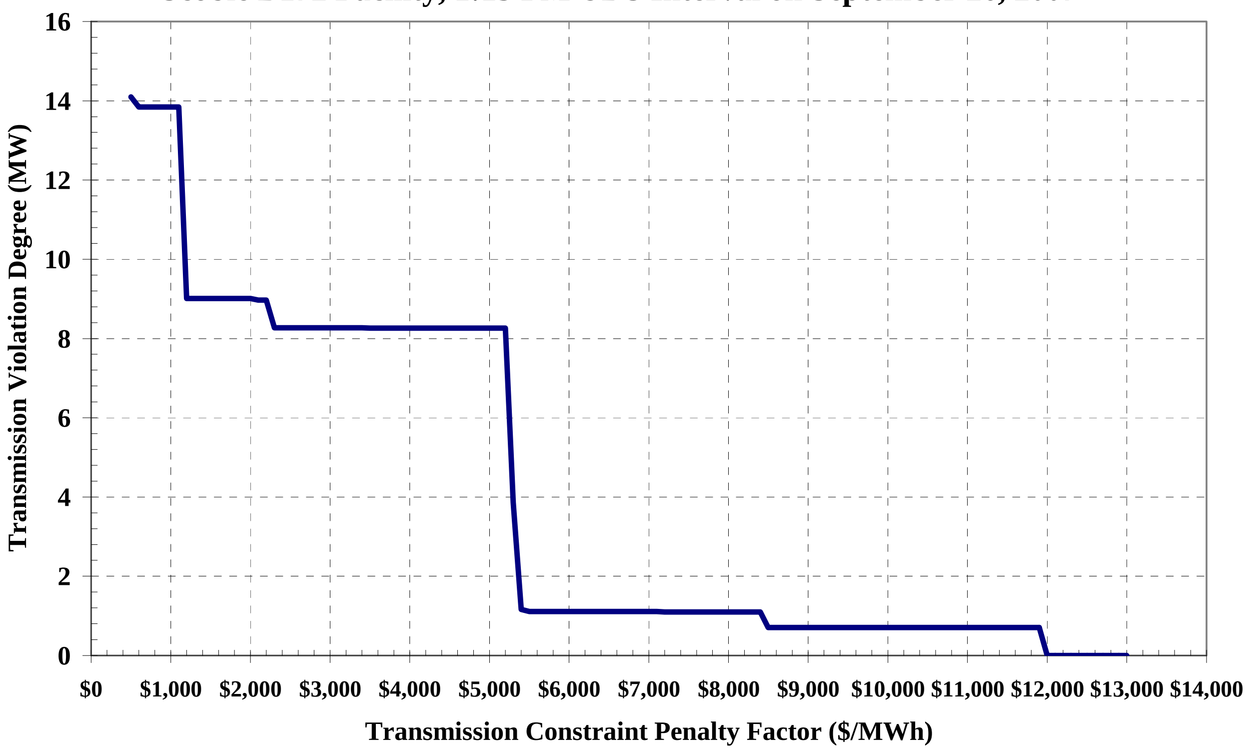
MWh and 25 intervals when the limit was violated. In 2008, there were just 32 intervals when the constraint was resolved with a shadow price exceeding $2,000 per MWh and 4 intervals when the limit was violated.

Although the figure indicates that acute conditions were relatively infrequent, such periods provide important market signals that can influence commitment, dispatch, and investment. During 25 intervals in 2007 and 4 intervals in 2008 when the limit was violated, the shadow price was set by the marginal available offers. In these intervals, the shadow price was below $100 per MWh in 21 intervals, between $100 and $1,000 per MWh in 3 intervals, and between $1,000 and $11,000 per MWh in 5 intervals. These pricing outcomes suggest that the current procedures do not always result in price signals that reflect the shortage of transmission capability. When a constraint is unmanageable, an algorithm is used to "relax" the limit of the constraint for purposes of calculating a shadow price for the constraint and the associated LMPs. Based on our analysis of the same software in the Midwest ISO, we have found that this algorithm tends to produce inefficiently low shadow prices when a constraint is in violation ($0 in many cases). Instead of using the relaxation algorithm to set LMPs, LMPs could be based on the shadow price of the violated constraint, which would equal constraint penalty factor. This would require setting the transmission constraint penalty factors to levels that reasonably reflect the value of satisfying transmission constraints. However, under current market conditions in New England, such a change would have limited market impact because transmission constraints are violated very infrequently.

All of the intervals shown in Figure 23 where the transmission limit was resolved with a shadow price exceeding $10,000 per MWh occurred in 2007, mostly on September 26. In these intervals, LMPs at several nodes exceeded $5,000 per MWh. Such costly redispatch should be limited to situations where it is absolutely necessary for reliability. Depending on the reason for the transmission limit, it may be possible to exceed the limit for a period of time without a significant degradation of reliability. For example, NERC allows some limits to be in violation for up to 30 minutes before it deems a reliability standard to have been violated. In such cases, it would be beneficial to impose a ceiling on the redispatch costs that can be incurred to manage the transmission constraint. A lower penalty factor could be used to impose a reasonable ceiling on redispatch costs.

The following case study provides some indication of how outcomes could be affected if a ceiling, or penalty factor, were imposed on redispatch costs. In the 1:15 PM UDS interval on September 26, 2007, the Scobie B172 facility was binding and the resulting marginal redispatch cost was $11,848 per MWh. In this interval, approximately 150 MW was flowing across the Scobie B172 facility. Figure 24 shows how the flow would have changed if a penalty factor was used to place a ceiling on the costs that could be incurred to manage the constraint. The y-axis shows the violation (i.e., number of megawatts in excess of the flow limit) that would have occurred for a given penalty factor, which is shown on the x-axis.

Figure 24: Effect of Reducing Transmission Constraint Penalty Factors
Scobie B172 Facility, 1:15 PM UDS Interval on September 26, 2007



As expected, if the penalty factor (shown on the x-axis) exceeded $11,848 per MWh, the transmission constraint would be fully resolved. However, if the penalty factor were reduced, some redispatch options would be too costly and UDS would violate the constraint rather than exceed the penalty factor. The figure shows that if the penalty factor were reduced below $11,848 per MWh, UDS would violate the constraint by 0.7 MW. The amount of the violation would increase if the penalty factor were further reduced. For example, if the penalty factor were reduced to $1,000 per MWh, it would result in a violation of approximately 14 MW. Although the figure illustrates just one example of how adjusting a penalty factor would affect operations, it suggests that reducing the penalty factor would lead to only a modest violation that may not significantly undermine reliability in some cases.

Hence, by adjusting the penalty factors for transmission constraints, it may be possible to limit extraordinarily costly redispatch to circumstances when not redispatching would not seriously affect reliability. The penalty factors could also be used to improve the efficiency of price signals during periods of scarce transmission capability. Just as RCPFs are used to set prices when the market is short of operating reserves, Transmission Constraint Penalty Factors could be used to set prices during periods of transmission scarcity.[53] For example, if a Transmission Constraint Penalty Factor of $3,000 per MWh was used to manage the constraint shown in Figure 24, it would result in an 8.2 MW violation of the limit, a $3,000 per MWh shadow price and corresponding LMPs. This approach would also safeguard the market by not allowing the relaxation algorithm to establish inefficiently low shadow prices when a constraint is violated. Hence, we recommend that the ISO evaluate potential enhancements to the current operating procedures that would establish reasonable Transmission Constraint Penalty Factors and allow them to set LMPs when a constraint is in violation.

D. Real-Time Pricing During the Activation of Demand Response

Price-responsive demand has great potential to enhance wholesale market efficiency. Modest reductions in consumption by end-users in high-price periods can significantly reduce the costs of committing and dispatching generation. Furthermore, price-responsive demand reduces the need for new investment in generating capacity. Indeed, the majority of new capacity procured in the first two Forward Capacity Auctions was composed of demand response capability rather

53 The use of RCPFs in the real-time market is described in Section V.B.

than generating capability. As interest increases in demand response programs and time-of-day pricing for end-users, demand will play a progressively larger role in wholesale market outcomes. This part of the section discusses the effects of demand response programs on the efficiency of real-time prices in the wholesale market.

1. Real-Time Demand Response Programs and Participation

Currently, the ISO operates the following four real-time demand response programs:[54]

- Real-Time 30-Minute Demand Response Program – These resources may be interrupted for anticipated capacity deficiencies with 30 minutes notice and receive the higher of the LMP or $500 per MWh for a minimum duration of 2 hours.
- Real-Time 2-Hour Demand Response Program – These resources may be interrupted for anticipated capacity deficiencies with 2 hours notice and receive the higher of the LMP or $350 per MWh for a minimum duration of 2 hours.
- Real-Time Profiled Response Program – These resources may be interrupted for anticipated capacity deficiencies within a specified time period and receive the higher of the LMP or $100 per MWh for a minimum duration of 2 hours.
- Real-Time Price Response Program – These resources may interrupt (but are not required to do so) when they receive notice on the previous day. If they interrupt, they receive the higher of the LMP or $100 per MWh for the eligibility period.

The first three programs are reliability-based programs that activate emergency demand response resources according to the OP-4 protocol during a capacity deficiency.[55] Accordingly, resources participating in these three programs are ICAP resources and receive capacity payments. The fourth program is a price-based program that provides a mechanism for loads to respond when the wholesale price is expected to be greater than or equal to $100 per MWh. Demand resources are only eligible to participate in one Real-Time program at a time. The resources participating under the Real-Time 30-Minute Demand Response Program can be activated with just 30

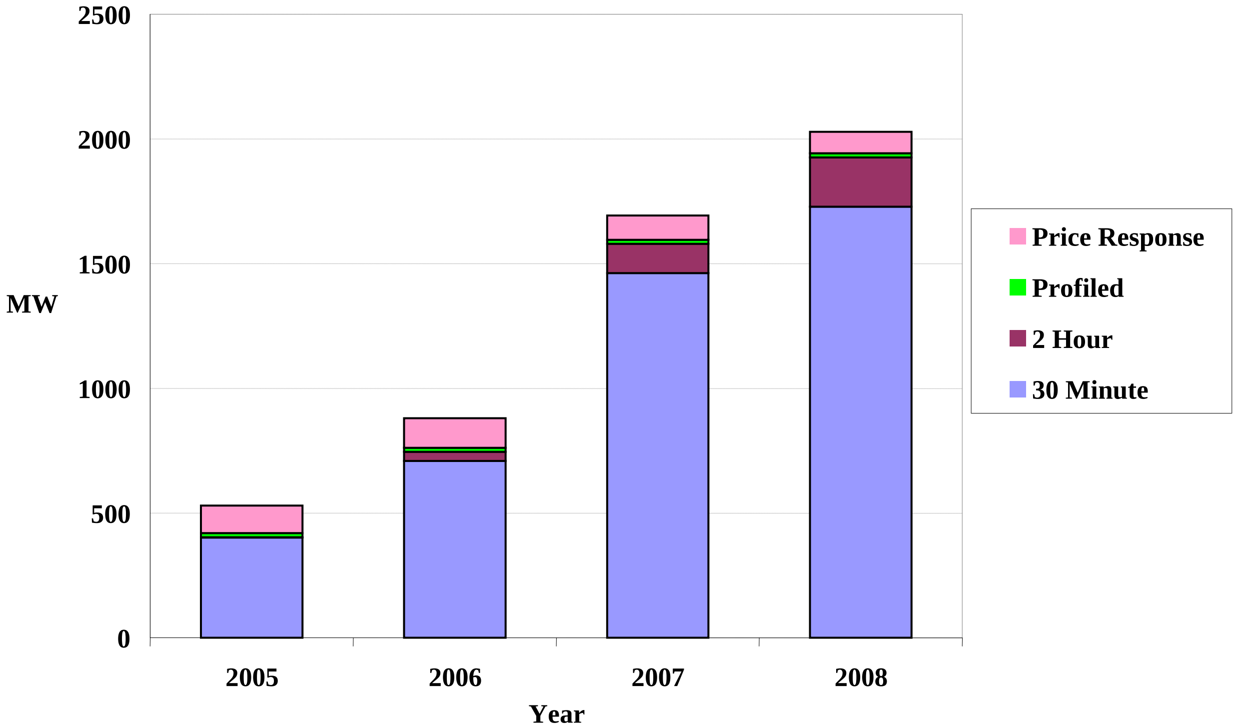
54 The current Demand Response Programs are set to expire on May 31, 2010 as a part of the transition to the Forward Capacity Market.

55 Real-Time 30-Minute Demand Response Program resources are activated under OP-4 Actions 9 and 12. Real-Time 2-Hour Demand Response Program resources and Real-Time Profiled Response Program resources are activated under OP-4 Action 3.

minutes notice. Therefore, they provide a higher degree of reliability benefit than resources participating under the other three programs.

Demand response participation has surged in New England in recent years. Figure 25 shows the quantity of resources enrolled in each of the four demand response programs from 2005 to 2008.[56]

Figure 25: Real-Time Demand Response Program Enrollments 2005 - 2008



The quantity of enrolled resources increased substantially from 530 MW in 2005 to 2,029 MW in 2008. 85 percent of demand response resources were enrolled in the Real-Time 30-Minute Demand Response Program by the end of 2008, which contributed the most to the overall growth of demand response programs. The Real-Time 2-Hour Demand Response Program also experienced notable growth during the past four years, increasing from 2 MW in 2005 to nearly

56 The quantities reported in this figure represent enrollments at the end of each year.

200 MW in 2008. The enrollment in the Real-Time Price Response Program decreased slightly over the period, from 110 MW in 2005 to 86 MW in 2008.

The programs described above are set to expire on June 1, 2010, at the start of the first Capacity Commitment Period under the FCM. However, ISO New England has proposed extension of the Real-Time Price Response Program and the Day-Ahead Load Response Program to May 31, 2011. This will allow time for discussions regarding whether these programs should be terminated or modified, which have already begun in the NEPOOL Markets Committee. The emergency programs are not being extended because the ISO deems them to be unnecessary with the advent of the demand response programs in the FCM. Many resources will transition from one of the current emergency programs to one of the FCM programs.[57] More than 2,700 MW of demand resources sold in the first two FCAs. Hence, FCM is serving to continue the rapid development of demand response resources in New England. These FCM results are discussed in detail in Section X.

2. Real-Time Pricing During Demand Response Activation

The rise in demand response participation will generate substantial market efficiencies, but it also presents significant challenges for efficient real-time pricing. The current real-time demand response resources are not dispatchable and must be activated in advance based on forecasted conditions for a duration of at least two hours.[58, 59] These inflexibilities can lead to two problems for the efficiency of real-time pricing:

- The amount of demand response that is activated may exceed the amount necessary to avoid the conditions that trigger the activation.

---

57 Under FCM, three types of resources sell non-passive demand response that is targeted at emergency conditions, shortage hours, and/or hours when load is expected to exceed 95 percent of the seasonal peak demand.

58 "Dispatchable" refers to resources that are able to modify their consumption or generation in response to remote dispatch instructions from the ISO generated by the real-time market.

59 Loads that are dispatchable in the real-time market are able to participate in the Asset Related Demand ("ARD") programs. ARDs are paid according to day-ahead and real-time LMPs. ARDs are not paid for capacity, however, they are also not charged for capacity obligations.

- The minimum curtailment duration of the demand response resources may be longer than the duration of the conditions that trigger the activation.

Both problems can lead to situations when the activation of demand response depresses real-time prices substantially below the marginal cost of the foregone consumption by the demand response resources. In 2008, these problems affected the efficiency of real-time pricing to a very limited extent because there were no capacity deficiencies that required activating emergency demand response resources. Further, participation in the real-time Price Response Program is still relatively limited. Nevertheless, these problems are likely to be more significant in the future, making it important to address them in the near-term.

Resources were activated relatively frequently under the Real-Time Price Response Program in 2008. The ISO activates these resources when it forecasts that real-time prices will reach $100 per MWh for one or more hours on the following day (not including weekends). Resources are activated for four or six hours, depending on the season, and are paid the higher of $100 per MWh or the real-time zonal clearing price. When resources were activated under this program in 2008, the average real-time clearing price was substantially lower than the average cost of activating these resources. Of the 958 hours when these resources were activated, the clearing price at the New England Hub was less than $100 per MWh in 60 percent of the hours and less than $70 per MWh in 22 percent of the hours.

One reason for the low prices is that the duration of the Real-Time Price Response Program curtailment is usually longer than the forecasted duration of $100 per MWh prices. Another reason is that the demand resources are not dispatchable in the real-time market, and therefore, do not set clearing prices. For example, suppose that the ISO activates demand response resources at a cost of $100 per MWh, allowing the ISO to avoid using a $105 per MWh generator. In this case, the clearing price would be set by the next most expensive generator, which might be at a cost of less than $100 per MWh. In such cases, allowing the demand response resources to set the clearing price could lead to real-time prices that better reflect the cost of deploying resources to meet the demand for energy and operating reserves. Currently, the Real-Time Price Response Program has a relatively small effect on real-time prices because enrollment in the program is limited. However, if participation in price-responsive programs grows, developing mechanisms that enable these resources to set clearing prices will be more important.

Although the current real-time emergency demand response programs will expire on June 1, 2010, similar issues arise with some of the demand response programs under FCM. Under FCM, there are three categories of resources that may be activated when a capacity deficiency is anticipated.[60] Because the demand resources cannot currently set prices in the real-time energy market, the activation of demand response can reduce real-time prices substantially below the marginal value of the foregone consumption of the demand response resources. Hence, under FCM, it will continue to be important to have efficient mechanisms for setting real-time clearing prices when demand response is activated.

The ISO recognizes the challenges that arise from participation by non-dispatchable resources and seeks ways to better integrate demand response in the real-time market. In this regard, the ISO is using the Demand Response Reserves Pilot ("DRR Pilot") Program to develop the capability to allow demand response resources to participate in the reserve markets. Under this program, demand response resources with the capability of responding to a reserve deployment within 30 minutes can qualify to sell TMOR in the real-time market and/or the forward reserve market. Prior to each summer and winter capability period, the ISO will select up to 50 MW of demand response resources in Connecticut to participate in the DRR Pilot. The ISO expects to demonstrate whether demand response resources can reliably provide operating reserves. It will also facilitate the development of control room communications, dispatch, metering, and telemetry to facilitate participation by demand response resources. If substantial quantities of demand response resources are able to participate in the reserve market, it will significantly reduce the cost of meeting the reliability needs of the system in the future.

60 This includes Critical Peak Demand Resources, Real-Time Demand Response Assets, and Real-Time Emergency Generation Assets.

3. Conclusions

The growth of demand response is a positive development that should reduce the cost of operating the system reliably, particularly during peak periods. Demand response provides an alternative to costly new generation investment. However, since the majority of demand response resources are not dispatchable in the real-time market, it can be challenging to set prices that reflect scarcity during periods when demand response resources are activated. Hence, we recommend that the ISO develop rules for allowing the activation of non-dispatchable demand response resources to be reflected in clearing prices when there is a capacity deficiency or when there would have been a deficiency without the activation of demand response resources.

Under FCM, two categories of demand response resources may be activated when the day-ahead forecasted load exceeds 95 percent of seasonal peak demand.[61] In most such hours, generation will be sufficient to satisfy load and reserve requirements and the prevailing real-time prices would not justify the activation of the demand response resources. Because these demand resources cannot currently set prices in the real-time market, their activation may lead to instances where demand response resources are curtailed when the real-time LMP is substantially lower than the marginal value of foregone consumption. It would be more efficient if the demand was not curtailed and additional generation was dispatched. Therefore, we will continue to evaluate market outcomes under FCM to identify whether demand is being curtailed inefficiently.

E. Ex Ante and Ex Post Pricing

The ISO adopted the ex post pricing method when it originally implemented the SMD market design in 2003. The ex post pricing method is also used by PJM and the Midwest ISO, while other ISOs use the ex ante pricing method. In this section, we evaluate the efficiency of the real-time prices produced by the ex post pricing method.

61 This includes Critical Peak Demand Resources and Real-Time Demand Response Assets.

Ex ante prices are produced by the real-time dispatch model (UDS) and are consistent with the cost-minimizing set of dispatch instructions. The ex ante prices are set to levels that give generators an incentive to follow their dispatch instructions.[62]

Ex post prices are produced by the LMP Calculator. At the end of each interval, the LMP Calculator re-calculates dispatch quantities and prices using inputs that are different in several respects from the inputs used by UDS. For each flexible[63] resource, a "real-time offer price" is used in place of its offer curve. For a resource following dispatch instructions, its "real-time offer price" equals the ex ante price at its location or, if it is operating at its maximum output level, the offer price corresponding to its actual production level. For a resource that is under-producing, the "real-time offer price" equals the offer price corresponding to the resource's actual production level. Each flexible resource is treated as having a small dispatchable range around its actual production level, where the upward range is much smaller than the downward range (e.g., approximately 0.1 MW up and 2 MW down). The purpose of the ex post pricing method is to generate a set of prices that is consistent with the actual production levels of generators in the market, rather than their dispatch instructions. This is intended to improve the incentives of generators to follow dispatch instructions.

The evaluation in this section addresses three issues with ex post pricing:

- The current implementation of ex post pricing results in a small but persistent upward bias in real-time prices.
- Ex post pricing does not improve the incentives to follow dispatch instructions.
- Occasional distortions caused by the ex post pricing method lead to inefficient pricing in congested areas.

The end of this section provides a summary of the conclusions and recommendations from the evaluation of ex post pricing.

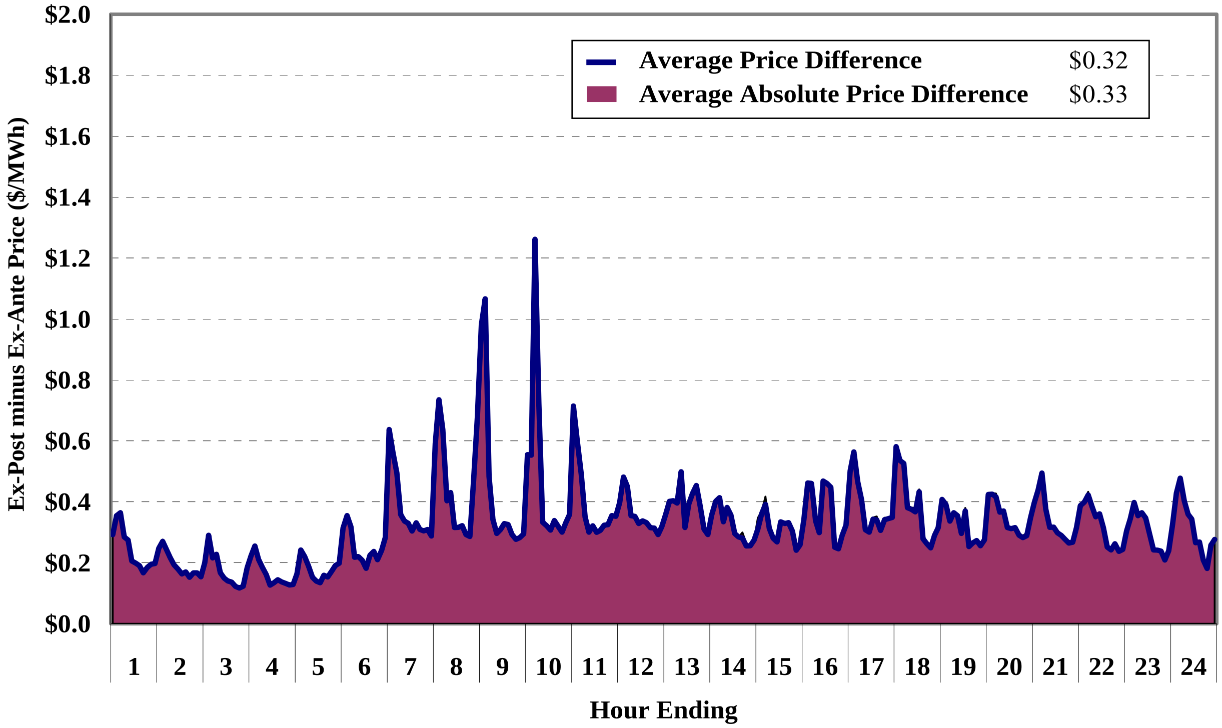
62 This assumes the generators are offered at marginal cost.

63 For most resources, they are treated as flexible if they are producing more than 0 MW and they meet one of the following conditions: (i) being committed for transmission, (ii) being dispatchable and producing less than 110 percent of their dispatch instruction, and (iii) being dispatchable and having a real-time offer price at their actual production level that is less than or equal to the ex ante price.

1. Persistent Differences Between Ex Ante and Ex Post Prices

The first analysis highlights an issue with the current implementation of ex post pricing that leads to a small but persistent upward bias in real-time prices. Figure 26 summarizes differences between ex ante and ex post prices in 2008 at a location close to the New England Hub.[64] This location is relatively uncongested, making it broadly representative of prices throughout New England. The blue line shows average ex post price minus average ex ante price by the time of day. The purple area shows the average absolute price difference by the time day.

Figure 26: Average Difference Between Five-Minute Ex Post and Ex Ante Prices 2008



The average differences between the ex post and ex ante prices were relatively small in 2008. However, the line shows a persistent bias that causes the ex post prices to be slightly higher than ex ante prices in the vast majority of intervals. As a result, average ex post prices were $0.32 per MWh higher than ex ante prices at this location. The persistent bias results from a combination

64 The MillBury station was selected because it is near the New England Hub. The New England Hub was not chosen because UDS does not calculate ex ante prices for load zones or the New England Hub.

of two factors. First, loss factors change slightly between the ex ante price calculation and the ex post price calculation as the pattern of generation and load changes. Even though many units' "real-time offer prices" are equal to the ex ante price (which should make them economically equivalent), these changes in loss factors affect the offer costs of some resources relative to others. The second factor is that the dispatchable range of each resource is generally 20 to 40 times larger in the downward direction than the upward direction.

In a typical interval, there may be 100 or more flexible resources. At locations where the loss factors increase the most from the ex ante model to the ex post model, resources will appear most costly and be ramped downward in the ex post model. Since the downward dispatchable range is much larger than the upward dispatchable range, many resources will be ramped up to their maximum to replace the unit that is ramped down. In one typical interval without congestion, three units were ramped down and more than 70 units were ramped up. As units that are ramped up in the ex post model reach their maximums, increasingly expensive units set ex post prices. Hence, the resource that is marginal in the ex post calculation usually has a loss factor that is higher than in the ex ante calculation, thereby leading to an upward bias in prices.

2. Theoretical Problems with Ex Post Pricing

Ex post pricing has been justified, in part, as a means to provide resources with incentives to follow dispatch instructions. However, ex post pricing does not efficiently provide such an incentive for several reasons. First, suppliers that are primarily scheduled day-ahead will not be substantially harmed by small adjustments in the real-time price. Second, with the exception of the episodic price effects in congested areas, the pricing methodology will not usually result in significant changes in prices when a unit does not follow dispatch instructions. In general, this is the case because many other units will have real-time offer prices in the ex post model that are very close to the offer price of the unit failing to following dispatch. Further, any slight change in the ex post price will not affect the unit failing to follow dispatch in a manner that has any relationship to the cost to the system of its actions. Hence, it is very unlikely that the ex post pricing enhances incentives to follow dispatch instructions.

In fact, because ex post pricing can, on occasion, substantially affect prices in congested areas, it can diminish suppliers' incentives to follow ex ante dispatch instructions when prices in the congested area are volatile. A much more efficient means to send targeted incentives to respond to dispatch instructions is the use of uninstructed deviation penalties.

A final theoretical concern is that ex post prices are theoretically less efficient than ex ante prices. The ex ante dispatch and prices represent the least cost dispatch of the system, given bids, offers, and binding constraints. If a unit is unable to respond to the dispatch instruction, then it implies that less supply is available to the market, and thus, the price should have been set by a more expensive offer. In other words, a higher-cost offer would have been taken if the market had known the unit could not respond. In such a case, however, the ex post pricing method would reduce the energy prices from the ex ante level because the marginal unit loses its eligibility to set prices.

3. Pricing Outcomes in Congested Areas

On occasion, the ex post pricing model substantially alters prices in congested areas. Such occasions arise when the marginal unit for the binding constraint becomes inflexible or flexible but with a reduced offer price[65] in the ex post pricing.

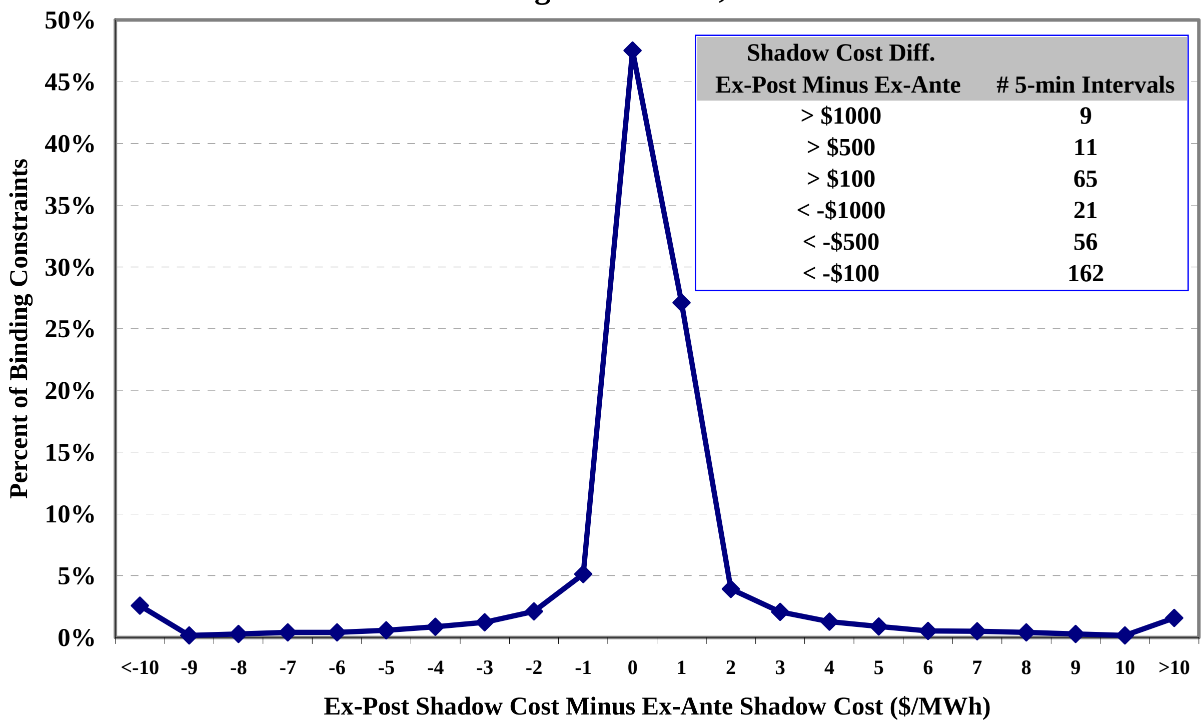
For example, suppose a combustion turbine with an incremental offer of $150 per MWh and an amortized start-up and no-load cost of $100 per MWh is started in order to resolve a load pocket constraint. Suppose that there is also a $50 per MWh unit in the load pocket that is dispatched at its maximum level. The ex ante LMPs in the load pocket will be $250 per MWh. Two pricing inefficiencies can occur in the ex post calculation. First, if the combustion turbine has not started because its start-up time has not elapsed or because it comes on late, the turbine will be deemed inflexible in the ex post calculation. This causes the $50 per MWh unit to set prices because it is the only flexible resource in the load pocket. Second, if the combustion turbine does start-up and is deemed flexible, the amortized start-up and no-load offers are not reflected in the current ex

65 When a fast-start unit is committed by UDS, its combined offer that adds its start-up and no-load offers on top of its incremental energy offer is used. In the ex post pricing, however, when the unit's offer is used, the start-up and no-load offers are not included.

post pricing. As a result, the turbine would set a $150 per MWh ex post price in the load pocket. In either case, the ex post congestion value is substantially reduced, causing significant discrepancy between ex ante and ex post prices in the load pocket. In both cases, the marginal source of supply costs $250 per MWh and the ex ante price is therefore the efficient price.

The significance of this issue depends on the frequency of such instances. Figure 27 summarizes differences in constraint shadow prices between ex post and ex ante calculations in 2008. A positive value indicates a higher shadow cost in the ex post calculation. For example, the value "2" on the horizontal axis means the ex post shadow cost is higher than the ex ante cost by $1 to $2.

Figure 27: Difference in Constraint Shadow Costs Between Ex Post and Ex Ante All Binding Constraints, 2008



Shadow Cost Diff. Ex-Post Minus Ex-Ante	# 5-min Intervals
> $1000	9
> $500	11
> $100	65
< -$1000	21
< -$500	56
< -$100	162

The average difference was not significant in 2008. Nearly 96 percent of all differences were within $10 per MWh and almost 88 percent were within $3 per MWh. However, there were a small number of intervals with substantial differences in congestion costs between the ex ante and ex post calculations. There were only 65 intervals during which ex post shadow prices were at least $100 per MWh higher than ex ante prices, and 162 intervals during which ex post shadow prices were at least $100 per MWh lower than ex ante prices.

4. Conclusions

Our evaluation of the ex post pricing results indicates that it:

- Creates a small upward bias in real-time prices in uncongested areas;
- Introduces small potential inefficiencies by setting prices that are not consistent with dispatch instructions; and
- Sometimes distorts the value of congestion into constrained areas.

The most significant, and perhaps only, benefit of ex post pricing is that it allows the ISO to correct the real-time prices when the ex ante prices are affected by corrupt data or communication failures. Given the theoretical and practical problems with ex post pricing, we recommend that the ISO consider an ex post process that would utilize corrected ex ante prices for settlement, rather than the current ex post prices.

F. Conclusions and Recommendations

Efficient price formation is an important function of real-time market operations. Efficient real-time price signals provide incentives for suppliers to offer competitively, for demand response to participate in the wholesale market, and for investors to build capacity in areas where it is most valuable. Hence, efficient prices provide market participants with incentives that are compatible with the ISO's mandate to maintain the reliability of the system.

This section evaluates several aspects of real-time pricing in the New England market during 2008. Our evaluation leads to the following conclusions and recommendations:

- Price corrections were very infrequent in 2008, which reduces uncertainty for market participants transacting in the New England wholesale market. Furthermore, a large share of the price corrections that did occur affected a very small number of pricing nodes.
- Fast-start generators are routinely deployed in merit order, but the resulting costs are not always fully reflected in real-time prices. This leads to inefficiently low prices, particularly in areas that rely on fast-start generators to manage local congestion.

- ✓ We recommend that the ISO evaluate potential changes in the pricing methodology that would allow the deployment costs of fast-start generators to be more fully reflected in the real-time market prices.

- The current operating procedures use the lowest-cost resources to maintain reliability when acute transmission constraints are binding. However, this process does not always result in efficient prices that accurately reflect operating conditions, and sometimes incurs excessive redispatch costs to manage a constraint.

  - ✓ We recommend that the ISO evaluate potential enhancements to current operating procedures that would establish reasonable Transmission Constraint Penalty Factors and allow them to set LMPs when a constraint is in violation.

- The recent surge in participation in demand response programs is a positive development that will reduce the cost of operating the system reliably, particularly during peak periods. However, the inflexibility of demand response resources creates challenges for setting efficient prices that reflect scarcity during periods when emergency demand response resources are activated.

  - ✓ We recommend that the ISO develop rules for allowing the activation of non-dispatchable demand response resources to be reflected in clearing prices when there is a capacity deficiency or when there would have been a deficiency without the activation of demand response resources.

- Finally, given the theoretical and practical problems with ex post pricing, we recommend that the ISO consider an ex post process that would utilize corrected ex ante prices for settlement, rather than the current ex post prices.

VIII. System Operations

The ISO ensures that sufficient resources will be available in the operating day to satisfy forecasted load and reserve requirements without exceeding the capability of the transmission system. The wholesale market is designed to help the ISO meet these requirements efficiently. In particular, the day-ahead market and the forward reserve market are intended to provide incentives for market participants to make resources available to meet these requirements. The day-ahead market clears physical and virtual load bids and supply offers, and produces a coordinated commitment of resources. The forward reserve market provides suppliers with incentives to make reserve capacity available, particularly from offline fast-start resources.

When the wholesale market does not satisfy all forecasted reliability requirements for the operating day, the ISO performs the Reserve Adequacy Assessment ("RAA") to ensure sufficient resources will be available. The primary way in which the ISO makes sufficient resources available is by committing additional generation. Such commitments generate expenses that are uplifted to the market and increase the amount of supply available in real-time, which depresses real-time market prices and leads to additional uplift. Hence, out-of-market commitment tends to undermine market incentives for meeting reliability requirements.

In this section, we evaluate several aspects of market operations that are related to the ISO's process to ensure that sufficient resources are available to meet the forecasted reliability requirements. In particular, we evaluate the following:

- Accuracy of Load Forecasting – The ISO's load forecasts are used by market participants to inform scheduling in the day-ahead market and by the ISO to determine the forecasted reliability requirements.
- Reliability Commitment – This section summarizes and evaluates reliability commitments, as well as the self-commitment patterns that affect decisions made in the RAA.
- Out-of-Merit Generation and Uplift Expenses – These sections examine the by-products of out-of-market commitment.

A. Accuracy of ISO Load Forecasts

The ISO produces a load forecast seven days into the future and publishes the forecast on its website. This forecast is significant because market participants may use it and other available information to inform their decisions regarding:

- Fuel procurement
- Management of energy limitations
- Formulation of day-ahead bids and offers; and
- Short-term outage scheduling.

In addition, the ISO uses the forecast to estimate the amount of resources that will be needed to satisfy load and reserve requirements without exceeding the capability of the transmission system. The day-ahead forecast is the most important because most scheduling and unit commitment takes place on the day prior to the operating day (either in the day-ahead market or in the RAA).

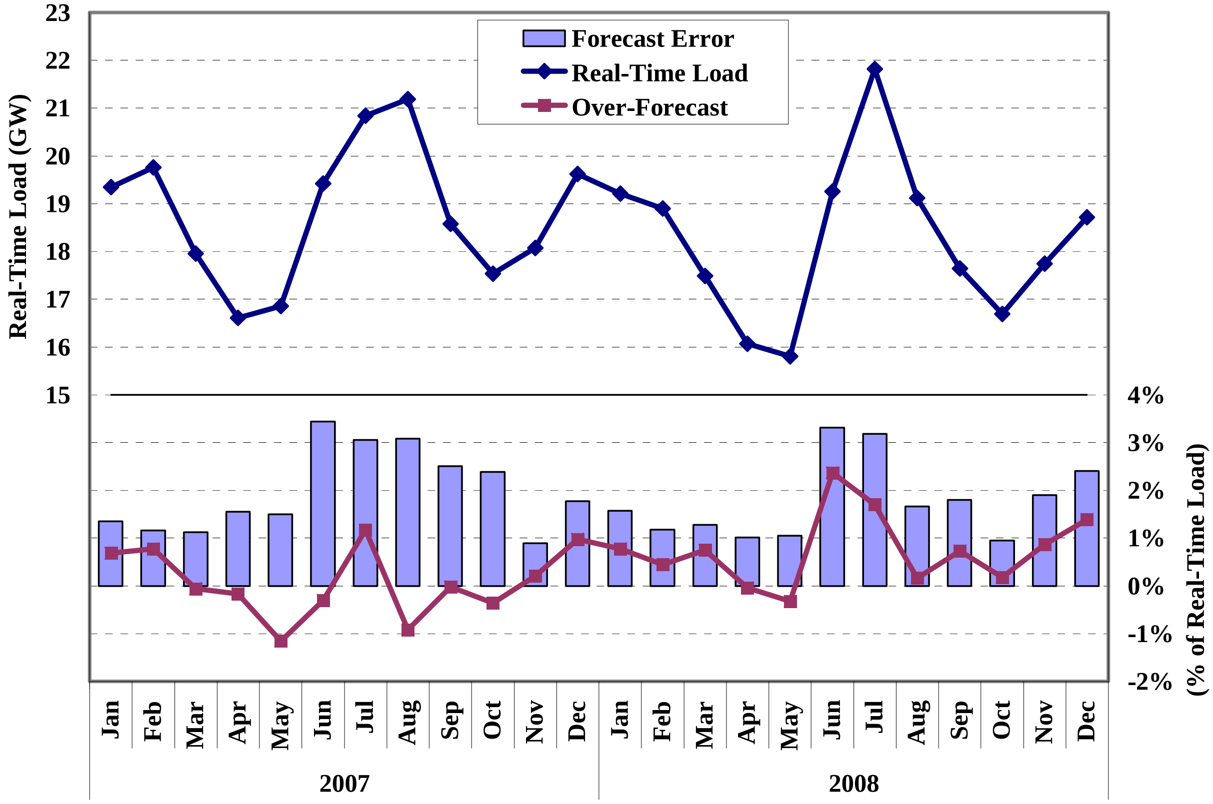
Accurate load forecasts promote efficient scheduling and unit commitment. Inaccurate load forecasts can cause the day-ahead market and/or the ISO to commit too much or too little capacity, which can affect prices and uplift. Therefore, it is desirable for the day-ahead forecast to accurately predict actual load.

Figure 28 summarizes daily peak loads and two measures of forecast error on a monthly basis during 2007 and 2008. The *Over-Forecast* is the percentage by which the average day-ahead forecasted daily peak load exceeded the average real-time daily peak load[66] in each month. Positive values indicate over-forecasting on average and negative values indicate under-forecasting on average. The *Forecast Error* is the average of the absolute difference between the day-ahead forecasted daily peak load and the actual daily peak load, expressed as a percentage of the average actual daily peak load.

---

66 The real-time daily peak load is based on the average load in the peak load hour of each day. Thus, the instantaneous peak load of each day is slightly higher than the values used in Figure 28.

Figure 28: Average Daily Peak Forecast Load and Actual Load Weekdays, 2007 - 2008



The figure shows a characteristic pattern of high loads during the winter and summer and mild loads during the spring and fall. In 2008, annual peak load of 26.1 GW occurred on June 10, although the average daily peak load was highest in July. Overall, loads were milder in 2008 than in the prior year. The average load declined from 15.6 GW in 2007 to 15.2 GW in 2008.

Overall, the ISO's day-ahead load forecasts are highly consistent with actual load. However, the ISO had a tendency to over-forecast load in 2008, particularly in June and July. The average over-forecast increased from 0.1 percent in 2007 to 0.7 percent in 2008. The ISO regularly evaluates the performance of its load forecasting models to ensure there are no factors that bias the forecast unjustifiably.[67]

67 A small bias toward over-forecasting may be justifiable because the costs of under-forecasting (i.e., under-commitment and potential for shortages) are likely larger than the costs of over-forecasting. Furthermore, it may be appropriate when the instantaneous peak load is expected to be substantially higher than the hourly average peak load.

The figure also shows the average forecast error, which is the average of the absolute value of the difference between the daily forecasted peak demand and the daily actual peak demand. For example, a one percent over-forecast on one day and a one percent under-forecast on the next day would result in an average forecast error of one percent, even though the average forecast load would be the same as the average actual load. From 2007 to 2008, the forecast error did not change significantly during the non-summer months, but it improved from 3.2 percent to 2.7 percent during the summer. Overall, we find that the load forecasting performance of the ISO remains good.

B. Commitment for Local Congestion and Reliability

In New England, several load pockets import a significant portion of their total electricity consumption. To ensure these areas can be served reliably, a minimum amount of capacity must be committed in the load pocket. Specifically, sufficient online capacity is required to:

- Meet forecasted load in the load pockets without violating any first contingency transmission limits (i.e., ensure the ISO can manage congestion on all of its transmission interfaces).
- Ensure that reserves are sufficient in local constrained areas to respond to the two largest contingencies;
- Support voltage in specific locations of the transmission system; and
- Manage constraints on the distribution system that are not modeled in the market software (known as Special Constraint Resources ("SCRs")).

In the day-ahead market, generators are scheduled based on the bids and offers submitted by buyers and sellers. A generator is committed when demand bids from load serving entities and virtual traders are high enough for the unit to be economic given its start-up, no-load, and incremental offer components. The willingness of load serving entities and virtual traders to buy (or sell) power in the day-ahead market is partly based on their expectations of LMPs in the real-time market on the following day. Thus, the day-ahead market commitment is strongly affected by expectations of real-time prices.

After the day-ahead market, the ISO may need to commit generation with high commitment costs to meet local reliability requirements. Once the commitment costs have been incurred, these generators may be inexpensive providers of energy and reserves in the local area. Because these commitment costs are not reflected in the market prices, the real-time LMPs frequently do not reflect the full value of online and fast-start capacity in local areas. Like any other forward financial market, the day-ahead market LMPs tend to converge with the real-time LMPs. Hence, day-ahead LMPs also do not reflect the full value of online and fast-start capacity in local areas, which reinforces the tendency of the day-ahead market-based commitment to not satisfy local reliability requirements.

Given the effects of supplemental commitment on market signals, it is important to minimize these commitments while still maintaining reliability. Periodically, the ISO evaluates refinements to the procedures and tools used in the RAA to make the process more efficient. The ISO has also made market enhancements that better reflect local reliability requirements in the real-time market, reducing the need for supplemental commitment. Nonetheless, supplemental commitments are still frequently needed to meet local requirements, so it is important to continue evaluating potential market improvements. This section discusses several initiatives by the ISO to reduce the frequency and effects of supplemental commitment.

In this section, we examine several issues that are related to supplemental commitment for local reliability. This section:

- Summarizes supplemental commitment for local reliability in the past two years;
- Evaluates the consistency of the ISO's operating procedures with decisions to commit resources for local second contingency protection in the RAA;
- Evaluates the accuracy of local second contingency transfer limits forecasted in the RAA which help determine the requirements in local areas; and
- Examines self-scheduling behavior that affects decisions in the RAA.

1. Summary of Commitment for Local Needs

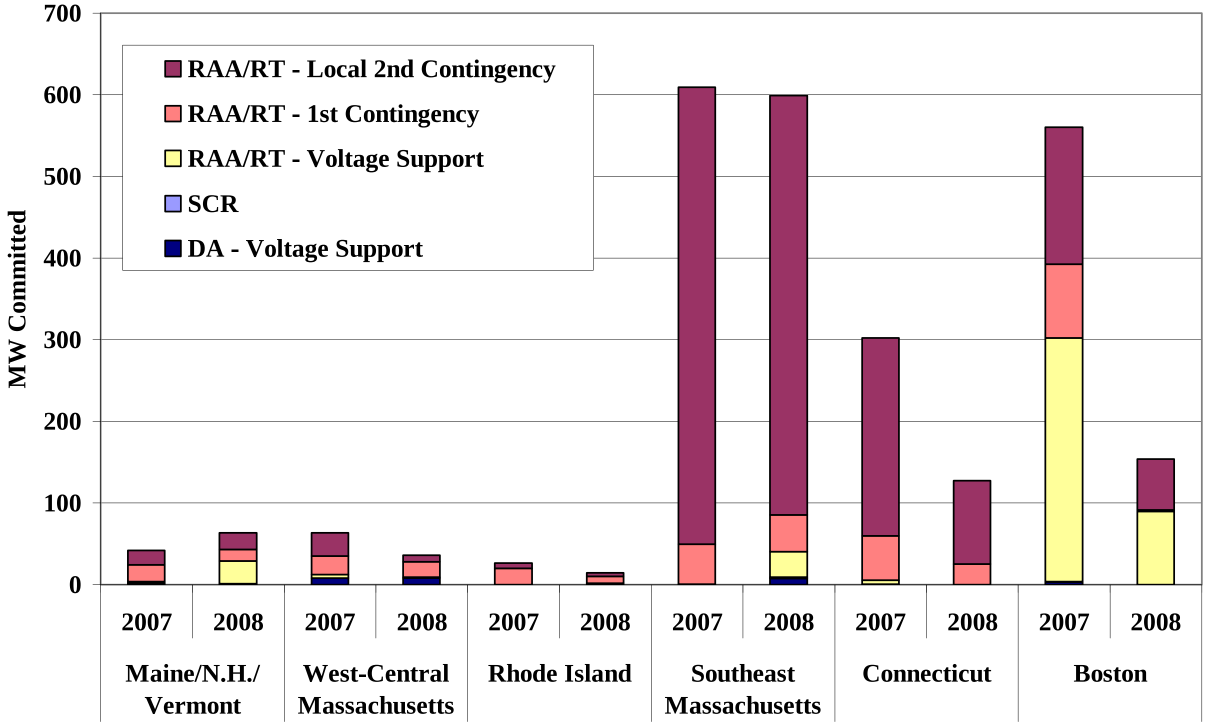
Figure 29 shows the average amount of capacity committed to satisfy local requirements in the daily peak load hour in each zone during 2007 and 2008.[68] The figure shows the entire capacity

---

68 In accordance with its Tariff, the ISO-NE classifies certain day-ahead commitments as Local Second Contingency commitments even though they occur as the result of market-based scheduling activity. Since

of these units, although their impact on prices depends on the amounts of energy and reserves they provide to the real-time market.

Figure 29: Commitment for Local Reliability by Zone
Daily Peak Hour, 2007 - 2008



Note: Capacity committed day-ahead for voltage support that would have been economically committed in the day-ahead market is excluded from the figure.

Supplemental commitment for local reliability declined substantially in 2008 compared to the prior year. The amount of supplemental commitment decreased 38 percent overall, from an average of 1,600 MW in 2007 to 1,000 MW in 2008. Nearly all of the decline occurred in Boston and Connecticut. In Boston, the amount of supplemental commitment fell from 560 MW in 2007 to 150 MW in 2008. In Connecticut, the amount of supplemental commitment fell from 300 MW in 2007 to 130 MW in 2008.

these are not out-of-market commitments, we exclude them from our analyses of supplemental commitment in this section.

In Boston, the need for supplemental commitment was reduced primarily due to revisions that the ISO made in early April 2008 in its operating guide for Boston-area reliability.[69] The revisions recognized the reduced need to commit generation for voltage support following the transmission upgrades into Boston that were made in 2007. As a result of the reduced need to commit generation for reliability, one supplier changed its day-ahead offer behavior, resulting in more market-based commitment in the day-ahead market.[70] The increase in market-based commitment reduced the need to commit generation for local second contingency protection in the Boston area.

In Connecticut, the reduction in supplemental commitment was primarily due to the transmission upgrades made under Phase II of the Southwest Connecticut Reliability Project. Most components of this project were placed in service during 2008. These upgrades significantly increased the transfer capability into Southwest Connecticut, reducing the need to commit local capacity for reliability.

In Southeast Massachusetts, the average amount of supplemental commitment remained constant from 2007 to 2008. Supplemental commitment is frequent there because the units needed to ensure local reliability in the Cape Cod area are usually not economic at day-ahead price levels. The ISO maintains sufficient reserves to respond to the two largest local contingencies without relying upon load shedding, which usually requires at least one of the units at the Canal plant to be online. Prior to 2006, these units were usually committed economically in the day-ahead market, but the increase in residual fuel oil prices relative to natural gas prices and the offers of these units have led them to be economically committed less frequently in the day-ahead market. As a result, these units are frequently committed for local reliability and receive a large amount of NCPC payments. Upgrades to the transmission system in lower Southeast Massachusetts are planned to come in service during 2009. These upgrades should substantially reduce the

69 The operating guides are the sets of procedures used by the ISO's operators to maintain reliability.

70 The behavior change is discussed later in detail in Section VII.B.4.

frequency with which the ISO needs to commit generation for local reliability in this area, thereby reducing the associated uplift costs.

The ISO has also proposed a rule change that should reduce the amount of supplemental commitment for local reliability and the associated uplift costs. The ISO filed with FERC proposed modifications to its mitigation measures that should more effectively address cases where suppliers that are frequently committed for local reliability raise their offers above marginal costs to extract larger NCPC payments. We support the proposed changes and discuss them in greater detail in Part 4 of this section.

To the extent that supplemental commitments occur in import-constrained areas such as Boston, Connecticut, and Lower Southeast Massachusetts, congestion is reduced in these areas. In Section C, we examine how much energy is dispatched out-of-merit as a result of these supplemental commitments.

2. Evaluation of Local Second Contingency Commitments

Supplemental commitment for local reliability can significantly affect market outcomes in constrained areas, so it is important that such commitments be made only when actually needed for reliability. This part of the section summarizes our evaluation of the consistency of the ISO's operating procedures with its decisions to commit generation for local second contingency protection.

The ISO's operating procedures explain the RAA process and the criteria for committing generation for local reliability.[71] Normally, the ISO waits until the close of the Re-Offer Period at 6 pm on the day ahead of the operating day before determining whether additional resources are needed for reliability. However, the assessment of local requirements may occur earlier "when it is recognized that required resources have long notification and start-up times."[72] When the amount of capacity that is required to meet the local second contingency requirements

71 See SOP-RTMKTS.0050.0010 – *Perform Reserve Adequacy Assessment* and SOP-RTMKTS.0050.0005 – *Determine Reliability Commitment for Real-Time*.

72 See Section 5.1.2 in SOP-RTMKTS.0050.0005.

exceeds the amount of capacity that is expected to be online, the ISO commits additional generation. The amount of capacity required to meet the local second contingency requirements is calculated using the RMR Calculation Worksheet. The amount of capacity that is expected to be online is the sum of the generation that was committed in the day-ahead market, self-scheduled during the Re-Offer Period, or committed already for another reliability reason.

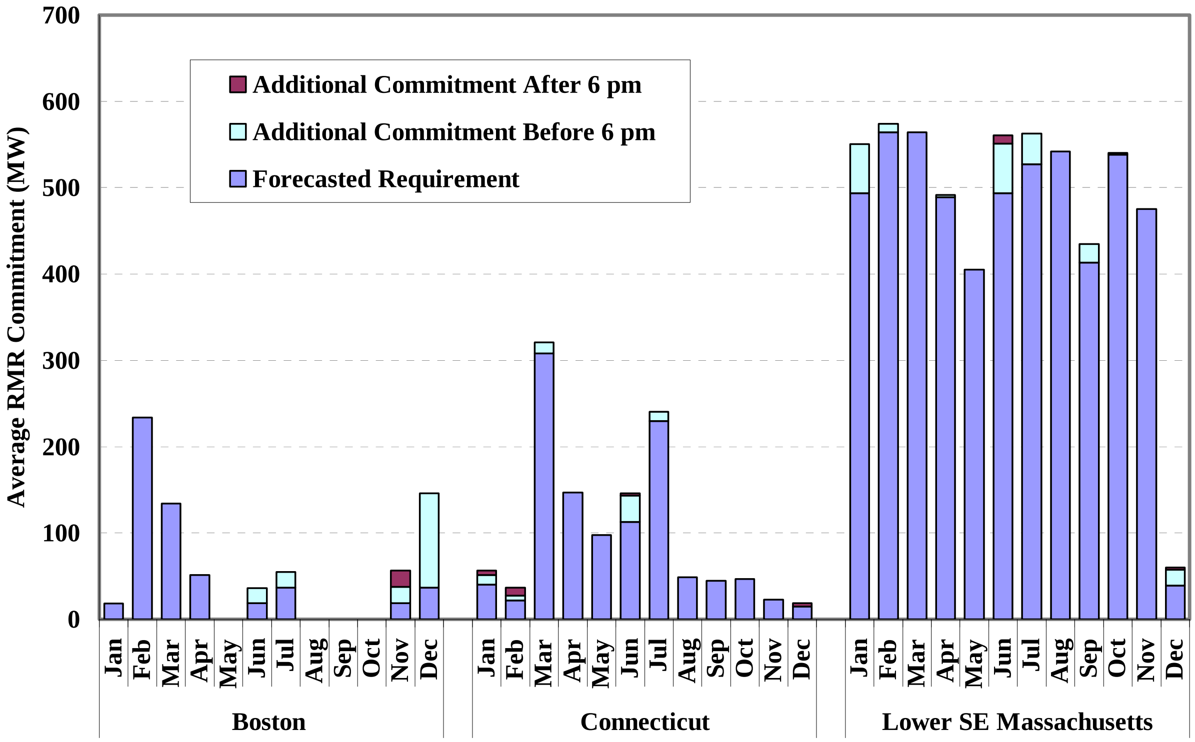
We evaluated local second contingency commitments that occurred on the day or evening prior to the operating day in Connecticut, Boston, and SEMA in 2008. The evaluation separated the second contingency commitments into (i) commitments that were needed to meet the forecasted second contingency requirements for one or more local areas as calculated in the RMR Calculation Worksheet[73] and (ii) additional commitments that were made in excess of those requirements.[74] The additional commitments were further separated according to whether they occurred before or after 6 pm. Figure 30 shows the results of the evaluation. The results for areas in Connecticut are shown together, although Connecticut contains four nested local second contingency protection areas: Norwalk-Stamford, Southwest Connecticut, West Connecticut, and Connecticut.

Local second contingency commitments averaged 60 MW per day in Boston, 100 MW per day in Connecticut, and 480 MW per day in Lower SEMA. The vast majority of capacity committed for local second contingency protection was necessary given the forecasted requirements for the following day. Additional commitments that did not appear necessary to satisfy the local requirements in the RMR Calculation Worksheet represented 25 percent of the commitments in Boston, 8 percent of the commitments in Connecticut, and 4 percent of the commitments in Lower Southeast Massachusetts.

73 The RMR Calculation Worksheet may be revised throughout the RAA process, but this evaluation used the requirements calculated in the final RMR Calculation Worksheet.

74 If only a portion of a second contingency protection resource is needed to meet the forecasted second contingency requirements calculated in the RMR Calculation Worksheet, the entire unit is categorized as satisfying the second contingency requirement.

Figure 30: Evaluation of Local Second Contingency Commitments 2008



Additional commitments are made for a variety of reasons, including concern by the operators regarding the forecasted peak load in the constrained area, uncertainty about the status or availability of a key resource in the area, and doubt regarding fuel supplies to some generation. In some cases, additional commitments are made to provide second contingency coverage to local areas that are not covered by the RMR Calculation Worksheet.[75] Additionally, when some candidates for reliability commitment have long notification times or start-up times, the ISO may need to commit resources prior to the close of the Re-Offer Period or the day-ahead market when there is less certainty about the need for additional resources. Figure 30 shows that almost 90 percent of the additional commitments were made before the close of the Re-Offer Period, indicating that long notification times and start-up times were primarily responsible for the additional commitments.

75 See Section 5.5.6 in SOP-RTMKTS.0050.0005, which describes the use of special studies for such local areas. The procedure for conducting special studies is outlined in SOP-OUTSCH.0050.0020 – *Perform Complex Studies*.

To the extent some commitments are not necessary to maintain reliability, the ISO seeks ways to minimize them because they can inefficiently mute congestion into the constrained areas. Even when local second contingency commitments are necessary based on forecasted operating conditions, there are several other factors that contribute to excess capacity in constrained areas. First, commitment is naturally lumpy because most generators have significant minimum operating levels and minimum run-times. Hence, operators may have to commit substantially more than is actually required to satisfy the local second contingency requirement in a particular hour. Increased reliance on small-scale, fast-start resources in the future should reduce this source of over-commitment. Second, sometimes suppliers self-commit after the ISO commits resources to meet local second contingency requirements, making some of the commitments by the ISO no longer necessary to satisfy the local second contingency requirement. Self-commitments are examined further in Part 4 of this section.

3. Forecasted Local Second Contingency Requirements

In the RAA, the ISO ensures that sufficient resources will be available to meet the forecasted load and reserve requirements during the current and next operating day. The ISO determines the generation and reserve needs in each local area based on forecasted load in the area, the Current Operating Plan, and forecasted transmission capability into the area.[76] Therefore, supplemental commitment decisions partly depend on the forecasted transmission capability. Over-forecasting the transfer capability of an interface can cause under-commitment inside the load pocket. Conversely, under-forecasting the transfer capability leads to over-commitment within the load pocket. Thus, it is important to forecast accurately factors that affect the capability of the transmission system.

In these areas, the required amount of internal generation and reserves depends on the local second contingency operating criteria as follows:

$$\text{Internal Load minus Internal Supply} \leq \text{Second Contingency Limit}$$

76 The Current Operating Plan includes the list of generators that are expected to be online or available to provide reserves while offline.

Internal supply includes generation, online reserves, and offline reserves capable of producing within 30 minutes. The Second Contingency Limit is the lower of (i) the Second Line Contingency Limit, which is the maximum amount of power that could be imported if the two largest line contingencies were to occur, and (ii) the Second Generator Contingency Limit, which is the maximum amount of power that could be imported if the largest line and generator contingencies both occur, minus the amount of the largest generator contingency (because the transmission capability must be available to import the power to replace the generator).

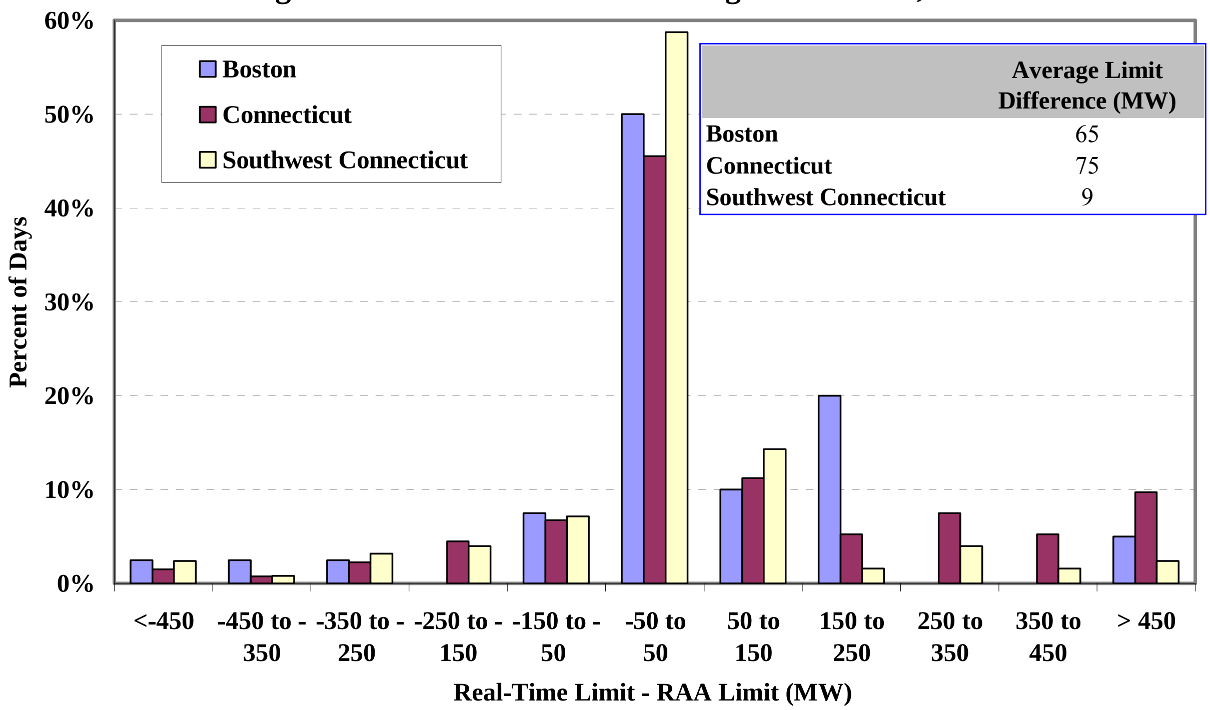
The same methods are used for calculating Second Contingency Limits in the RAA and real-time market. However, real-time capability is estimated based on actual operating conditions, while the capability in the RAA is estimated from forecasts of the next day's operating conditions. Differences between the RAA and the real-time market regarding the following factors can lead to differences in estimated transfer capability:

- The thermal limits of individual elements that make up the interface;
- Outages of key transmission lines;
- The commitment status of generators that influence the distribution of real power and/or reactive power flows across the interface;
- The size of the largest generator contingency; and
- The amount of load that can be shed in the event of a contingency.

There will always be differences between forecasted and actual conditions that will give rise to differences in the transfer capability between the RAA and real-time market. In general, these differences should be random and result in relatively small differences in the RAA and real-time transfer capabilities. In some cases, reliability concerns related to unknown factors in the RAA process may justify use of conservative assumptions that would cause RAA transfer capability to be lower than real-time transfer capability.

Figure 31 summarizes the differences between RAA and real-time transfer capability for three key interfaces. For each interface, we evaluated differences in the daily peak load hour on days in 2008 when local second contingency commitments occurred in the local area, excluding commitments that were reflected in the day-ahead market. This included 40 days for Boston, 134 days for Connecticut, and 126 days for Southwest Connecticut. A positive value in the chart indicates that more transfer capability was available in real-time than was estimated to be available in the RAA.

Figure 31: Differences between Real-Time and RAA Limits
Average Real-Time Limit minus Average RAA Limit, 2008



In the majority of cases, the differences between the RAA and real-time limits were close to zero. The differences were within 50 MW in 50 percent of the examined hours in Boston, nearly 60 percent of the examined hours in Southwest Connecticut, and more than 45 percent of the examined hours in Connecticut.

The inset table shows that the average RAA limit was close to the average real-time limit for the Southwest Connecticut interface. However, for the Boston and Connecticut interfaces, the real-time limits were approximately 70 MW higher than the RAA limits on average. This suggests there may have been a tendency in the RAA process to over-estimate the generation needed in Boston and Connecticut. However, it is also possible that the real-time limits were increased as a result of the supplemental commitments for local reliability that occurred after the RAA limits were estimated. The commitment of a particular generator can affect the transfer limit of an interface (e.g., by changing the pattern of real or reactive power flows across the interface).

There were a limited number of instances when the differences between the RAA limit and the real-time limit were significant. Connecticut, which had differences exceeding 250 MW on 27 percent of the days shown, exhibited the widest distribution of differences among the three interfaces. Overall, the differences were small given the uncertainties faced in the RAA. Nevertheless, such differences can lead to over-commitment or under-commitment, so it is important to identify ways to improve the accuracy of the forecasted limits.

In 2008, the ISO upgraded the software tools used to calculate these transmission limits. A new PowerWorld based application replaces older software that required more manual steps, reducing the total calculation time from hours to minutes. The new software provides:

- Greater accuracy by using a full AC model for all stages of analysis;
- Less dependence on subjective factors since more of the critical inputs to the calculation are pre-defined;
- Increased reliability by allowing the user to run many more scenarios to cover possible uncertainties in the system model;
- Reduced need for ISO employees to perform manual calculations; and
- More capability to re-run studies in case of last minute changes in the system conditions.

We support the ongoing efforts of the ISO to improve the tools it uses to establish the transfer limits.

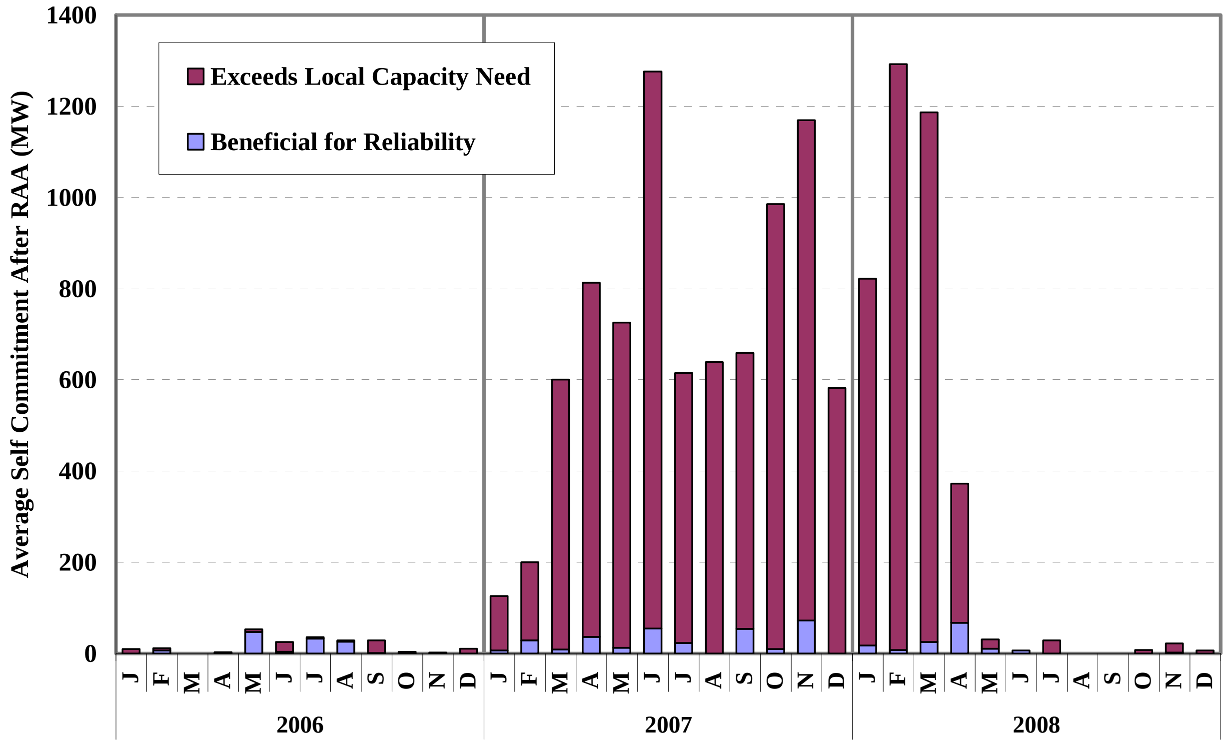
4. Self-Commitment after the RAA

Before committing resources to provide local second contingency protection, the ISO calculates the amount capacity that is expected to be online based on the Current Operating Plan. Depending on the timing of the calculation, this may include capacity committed: (i) through the day-ahead market, (ii) day-ahead for voltage support or other reliability reasons, (iii) in the Re-Offer Period by a self-schedule, and (iv) in the RAA for reliability reasons. If the ISO is short of the local capacity requirement, it will commit additional resources. However, if a supplier self-commits a generator after local second contingency protection resources are committed in the

RAA, it can lead to excess capacity in the load pocket. This excess creates uplift because it leads to local second contingency protection commitments that would not have been necessary if the ISO had been aware of all self-commitments when it conducted the RAA. This section evaluates self-commitments occurring after the RAA.

In recent years, Boston has been the only area where substantial amounts of capacity have been committed by self-schedules after the RAA. Figure 32 summarizes the extent to which self-commitment after the RAA has helped meet remaining local second contingency requirements in Boston versus how often it has led to excess capacity.

Figure 32: Self-Commitment after the Resource Adequacy Assessment
Boston – January 2006 to December 2008



In Boston, the average amount of capacity self-committed after the RAA process has varied dramatically from 2006 to 2008. The quantity of self-commitments was close to zero in 2006, rose dramatically in 2007, and declined to almost zero again in May 2008. This pattern is primarily attributable to a change in behavior regarding several units controlled by one supplier. In 2006, these units were under a reliability agreement with the ISO, which stipulated that the capacity be offered in the day-ahead market at marginal cost. As a result, these units were committed in the day-ahead market on most days because they were usually economic. In 2007, the reliability agreement expired, and the supplier raised its day-ahead offer prices above marginal cost to avoid market-based commitment in the day-ahead market. This frequently required the ISO to commit some of the supplier's capacity for local reliability. The remaining capacity not needed for local reliability was normally self-committed by the supplier after the RAA process. In April 2008, the pattern of self-commitment after the RAA ended because the supplier started to offer its generation closer to marginal cost in the day-ahead market. In the figure, variations in self-commitment after the RAA reflect changes in the frequency of day-ahead market commitment.

Our analysis indicates that only a small quantity of the self-commitments after the RAA was needed to meet the local capacity requirement in 2007 and early 2008. Although it can be efficient to have more than the minimum capacity required in each local area, a large share of these self-commitments occurred after the ISO had already committed units for local second contingency protection. If the ISO had known for certain that these units would be self-scheduled, it would not have needed to commit as many units for local second contingency protection. In some cases, the ISO can de-commit a resource that had been committed in the RAA if a self-schedule occurs later that eliminates the need for that commitment. However, this process has not been fully effective because self-commitments frequently occur after it is too late to de-commit other resources.

Self-commitment after the RAA can lead to excess capacity and inefficient market outcomes in constrained areas. The excess capacity mutes congestion into the load pocket, depressing real-time prices and increasing uplift costs. Additionally, this type of self-commitment undermines convergence between day-ahead and real-time prices, because load serving entities do not always accurately predict the amount of capacity that will be self-scheduled.

While there are some legitimate reasons for self-commitment after the RAA (e.g., suffering a forced outage on a unit that had been scheduled in the day-ahead market), this activity is also consistent with incentive problems that result from frequent supplemental commitment in an area with high market concentration. Local reliability requirements in Boston were frequently satisfied outside the normal market process, and supplementally-committed units were paid their offers when the clearing price was not sufficient for them to recover their as-bid costs. Even in competitive markets, units with pay-as-bid incentives rationally offer above marginal cost. Suppliers with generation that is frequently committed for local reliability often have some degree of local market power and a greater incentive to offer above marginal cost. Generally, if such suppliers submit high-priced offers in the RAA and their generators are not committed, they risk foregoing the profitable opportunity to sell energy in the real-time market. However, this risk does not exist for suppliers that have generators flexible enough to self-commit after the RAA.

The pattern of self-commitment by one supplier in Boston continued until April 2008 when the ISO revised the operating guide for Boston-area reliability. As a result, fewer commitments of these units have been necessary to support voltage in Boston since April 2008. The reduced need to commit for reliability has led the supplier to offer more of its generation closer to marginal cost in the day-ahead market, resulting in more frequent market-based commitment. This has substantially reduced commitments for local reliability and the resulting excess capacity in the Boston area.

Although the pattern of self-commitment in Boston was addressed by a change in the operating guide for Boston-area reliability, the ISO has also proposed a rule change that should reduce the potential for similar occurrences in the future. The ISO filed with FERC proposed modifications to its mitigation measures that should more effectively address cases where suppliers that are frequently committed for local reliability raise their offers above marginal costs to extract larger NCPC payments. The proposed modifications would tighten the conduct and impact thresholds applied to determine when mitigation should be applied to units receiving NCPC payments. Tightening these thresholds would reduce the expected profits from engaging in behavior to increase NCPC payments. We support the proposed changes.

5. Local Reliability Commitment – Conclusions

The analysis in this section highlights changes in the supplemental commitment patterns and points to several conclusions. Overall commitment for local reliability declined in New England by 38 percent from 2007 to 2008 due to reduced commitment in Boston and Connecticut. Commitment for Boston-area reliability has been significantly reduced since April 2008, due to changes in the operating procedures and requirements for Boston. Commitment for Connecticut-area reliability also decreased notably in 2008 because most components of Phase II of the Southwest Connecticut Reliability Project were placed in service during 2008. This project has significantly improved the transmission system, increasing the transfer capability into Southwest Connecticut.

Lower Southeast Massachusetts continues to require one of two large units to be committed almost continuously for local second contingency protection to ensure reliability for the Cape Cod area. In 2008, Lower Southeast Massachusetts accounted for 60 percent of the reliability commitments in New England. Transmission upgrades planned for 2009 are expected to substantially reduce the frequency of these commitments and the resulting uplift charges.

In this section, we found that the majority of the commitments by the ISO for local second contingency protection were necessary to meet forecasted minimum capacity requirements in constrained areas. This is important because unnecessary commitment depresses economic signals in constrained areas. We also found that the forecasted transfer limits used in the RAA were generally consistent with the real-time limits, although the average limits used in the RAA were tighter by a small amount than the average limits used in the real-time for the Boston and Connecticut interfaces.

The last analysis in this section finds that self-commitment after the RAA process in the Boston area was prevalent until April 2008. However, changes to the Boston-area local reliability requirements eliminated the incentive for the supplier to self-commit after the RAA.

At the end of this section, we describe recent changes made by the ISO and recommend additional changes that should improve the local reliability commitment results in the future.

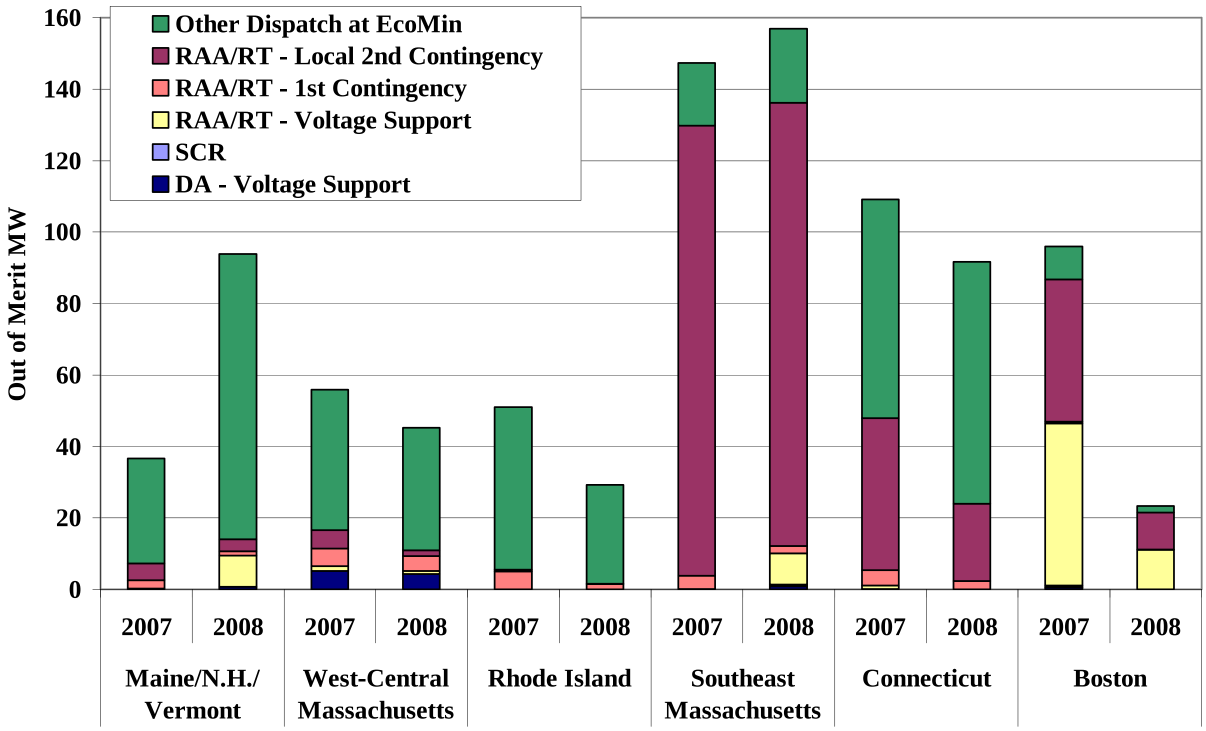
C. Out-of-Merit Dispatch

Out-of-merit dispatch occurs in real time when energy is produced from an output range on a unit whose incremental energy offer is greater than the LMP at its location. Out-of-merit generation tends to reduce energy prices by causing lower-cost resources to set the energy price. In a very simple example, assume the two resources closest to the margin are a $60 per MWh resource and a $65 per MWh resource, with the market clearing price set at $65 per MWh in the absence of congestion and losses. When a $100 per MWh resource is dispatched out-of-merit, it will be treated by the software as a must-take resource with a $0 per MWh offer. Assuming the energy produced by the $100 per MWh resource displaces all of the energy from the $65 per MWh resource, the energy price will decrease to $60 per MWh.

A unit may be dispatched out-of-merit for several reasons. First, a unit may run at its EcoMin to satisfy its minimum run time after having run in-merit in previous hours or in anticipation of running in an upcoming hour. This is efficient because the software is minimizing cost over the total run-time of the unit. Second, a unit committed for reliability reasons during or after the day-ahead market may be out-of-merit at its EcoMin. Units are committed for reliability without regard for their incremental offers above EcoMin and are, therefore, more likely than units committed in the day-ahead market to have incremental offers higher than the LMP.

Third, a unit may be dispatched out-of-merit in real time to satisfy reliability requirements in real time. Similar to the supplemental commitments, operators may request certain units to run at higher levels than their energy offers would justify. This can be necessary for a number of reasons, including: (a) providing voltage support on transmission or distribution facilities; (b) managing congestion on local distribution facilities; or (c) providing local reserves to protect against second contingencies. Figure 33 summarizes the average out-of-merit dispatch by location during peak hours (weekdays 6 AM to 10 PM) in 2007 and 2008.

**Figure 33: Average Hourly Out-of-Merit Dispatch
2007 - 2008 – Weekdays 6 AM to 10 PM**



Note: Capacity committed day-ahead for voltage support that would have been economically committed in the day-ahead market is included in the 'Other Dispatch at EcoMin' category.

In most regions, Figure 33 shows that most of the out-of-merit dispatch is attributable to non-local reliability units being dispatched at EcoMin. However, in Boston and Southeast Massachusetts, most of the out-of-merit dispatch was from units committed in the RAA for local second contingency protection or voltage support.

The average quantity of out-of-merit dispatch from units committed for local reliability (including first contingency, second contingency, voltage support, and SCR) declined 29 percent, from an average of 294 MW in 2007 to 208 MW in 2008. The amount of out-of-merit energy from non-local reliability units (i.e., Other Dispatch at EcoMin) rose modestly from an average of 202 MW in 2007 to 232 MW in 2008.

The changes in out-of-market dispatch that occurred in 2008 tracked the changes in supplemental commitments and were caused by the same underlying factors. The reduced commitment for local second contingency requirements in West-Central Massachusetts, Connecticut, and Boston led to proportionate reductions in out-of-merit energy in those zones. The decreased commitment for voltage support in Boston led to reduced out-of-merit dispatch on generators committed for that reason. The second contingency requirements in Lower Southeast Massachusetts continued to result in a substantial amount of out-of-market dispatch in that area.

Although some resources may need to be dispatched out-of-merit in any system, this should be minimized because it can undermine the efficiency of the locational energy and reserves prices. Furthermore, owners of units that are frequently called out-of-merit order face incentives to offer in excess of marginal costs, which can result in less efficient commitment and dispatch decisions. In addition, when units are offered above marginal costs, it reduces the likelihood that they will be committed economically through the day-ahead market, and increase the need for supplemental commitments. Hence, the pattern can be self-reinforcing.

D. Uplift Costs

To the extent that the wholesale market does not satisfy New England's reliability requirements, the ISO takes additional steps to ensure sufficient supplies are available. The ISO has used reliability agreements and supplemental commitment to ensure reliability, primarily in local import-constrained areas. Reliability agreements give the owners of uneconomic generating facilities supplemental payments in order to keep them in service. Supplemental commitments bring uneconomic capacity online at times when market clearing prices are insufficient. Such generators receive additional payments called NCPC payments, which make up the difference between their accepted offer costs and the market revenue. The costs associated with these payments are recovered from market participants through uplift charges. This section describes the main sources of uplift charges and how they are allocated among market participants.

The following table summarizes several categories of uplift during 2007 and 2008. The main categories of uplift are:

- Reliability Agreements – The uplift from these are allocated to Network Load in the zone where the generator is located.[77] In 2008, 39 percent of the capacity in Connecticut was

77 Network Load includes transmission customers that are served by the Transmission Owner.

covered under reliability agreements. The amount of capacity in Boston covered under reliability agreements declined to 0 percent by the end of 2007 due to transmission upgrades into Boston and the start of capacity transition payments.

- Local Second Contingency Protection Resources – In 2008, 98 percent of the uplift from these units was allocated to Real-Time Load Obligations and Emergency Sales in the zone where the generator is located.[78] The remaining uplift associated with day-ahead rather than real-time commitments was allocated to day-ahead load schedules in the local zone.
- Special Constraint Resources – The uplift paid to these resources is allocated to the Transmission Owner that requests the commitment.
- Voltage Support Resources – The uplift paid to these resources is allocated to Network Load throughout New England and Through-and-Out transactions.
- Other supplemental commitment (including local first contingency resources) – In 2008, 75 percent of this uplift was allocated to Real-Time Deviations throughout New England.[79] The remaining uplift associated with units committed in the day-ahead market is allocated to day-ahead scheduled load throughout New England.

The vast majority of uplift in each of these categories is incurred to address local supply inadequacies. For this reason, it is generally appropriate to allocate these charges to the local customers who derive benefit from their service. The first three of these categories are allocated to local customers, while the uplift charges for Voltage Support Resources and other supplemental commitment are allocated to customers throughout New England.

The following table summarizes the total costs of uplift associated with reliability agreements and supplemental commitment.

---

78 Real-Time Load Obligations include load customers that are served by the Load Serving Entity.

79 Real-Time Deviations include Real-Time Load Obligation Deviations, which are positive or negative differences between day-ahead scheduled load and actual real-time load; generation deviations from day-ahead schedules, which include virtual supply schedules; and generation deviations from the larger of the unit's Economic Minimum and desired dispatch point in each hour.

Table 2: Allocation of Uplift for Out-of-Market Energy and Reserves Costs
2007 & 2008

	Millions of Dollars	
Category of Uplift	**2007**	**2008**
Reliability Agreement		
Connecticut	**$115**	**$110**
Boston	**$3**	**$0**
Other Areas	**$25**	**$19**
Local Second Contingencies		
Connecticut	**$35**	**$24**
Boston	**$22**	**$11**
Southeast Massachusetts	**$108**	**$143**
Other Areas	**$4**	**$4**
Special Case Resources	**$2**	**$2**
Voltage Support	**$46**	**$29**
Other (Mostly Local First Contingencies)	**$29**	**$44**
Total	**$390**	**$387**

Note: Since information is not publicly available on the breakdown of payments under reliability agreements by load zone, this analysis assumes that the ratio of payments to fixed cost guarantees is the same for Boston, Connecticut, and other areas.

Overall, uplift charges did not change significantly from 2007 to 2008. The reductions in commitment for local reliability from 2007 to 2008 led to substantial reductions in uplift, although these were offset by the effects of increased fuel prices. In Connecticut and Boston, decreased commitment for local second contingency protection led to a $22 million reduction in uplift. The decreased commitment of generation for voltage support in Boston accounted for a $17 million reduction in uplift for voltage support. The effects of higher fuel prices in 2008 are demonstrated by the increase in uplift charges in Southeast Massachusetts, where the uplift from local second contingency protection increased $35 million from the prior year.

The reliability agreement costs decreased from $143 million in 2007 to $129 million in 2008, primarily because reliability agreements expired for several units during this period. The transition to a forward capacity market with locational price signals will help New England shift away from relying on reliability agreements to meet resource adequacy criteria in the future.

Reliability agreements provide additional payments to the least economic resources in the market and do not provide incentives for efficient investment. By compensating all resources in a particular area consistently, capacity markets provide more efficient signals for investment.

Section B discusses several initiatives, either recently implemented or under development, that should help reduce the uplift that arises from supplemental commitment in the next few years. Transmission upgrades have allowed better operating procedures for Boston, reducing the need to commit generation for voltage support in 2008. Transmission upgrades planned for 2009 should greatly reduce the frequency of commitment for local second contingency protection in SEMA. The ISO has proposed rule changes related to mitigation of suppliers that are frequently committed for local reliability, which should reduce the ISO's uplift costs.

E. Conclusions and Recommendations

In general, we conclude that the ISO's operations to maintain adequate reserve levels in 2008 were reasonably accurate and consistent with the ISO's procedures. The amount of capacity committed for reliability decreased significantly in Connecticut and Boston, although substantial quantities of supplemental commitment continue to occur, particularly in Lower Southeast Massachusetts. These commitments are necessary because these areas do not have a large quantity of fast-start resources that can help meet the capacity requirements of the local area while offline and the energy and reserves prices in these areas are usually not high enough to support the running costs of the non-fast-start units that are committed for local reliability.

Supplemental commitments and the resulting out-of-merit energy create four issues in the New England market:

- Inefficiencies created because supplemental commitments are made with the objective of minimizing commitment costs (i.e., start-up, no-load, and energy costs at EcoMin), rather than minimizing the overall production costs.
- Dampening of economic signals to invest in areas that would benefit the most from additional investment in generation, transmission and demand response resources.
- Larger and more volatile uplift costs that are difficult for participants to hedge.

- Incentives for generators frequently committed for reliability to avoid market-based commitment to seek additional payments through the reliability commitment process.

The ISO has implemented or is pursuing several additional measures to minimize reliance on supplemental commitments in load pockets including:

- Transmission upgrades into Boston (completed in Spring 2007) and associated changes in the area's local reliability requirements (implemented in early 2008).
- Transmission upgrades into Southwest Connecticut that allow less internal online capacity to maintain reliability (mostly in service in 2008, projected completion in 2009).
- Transmission upgrades into Southeast Massachusetts that enable the ISO to maintain reliability in these areas with less internal capacity (planned for 2009).
- Upgrades to the software tools used to calculate transmission capability into local areas. The new PowerWorld based application improves the accuracy, reliability, and efficiency of the calculations (implemented in 2008).
- Additional mitigation measures that better address the incentives of suppliers that persistently raise their offers above marginal costs to extract larger NCPC payments in areas where generation is frequently committed for local reliability (filed in 2009).

In addition, we recommend the ISO consider providing generators with additional flexibility to modify their offers closer to real-time to reflect changes in marginal costs.

Furthermore, we recommend several changes in Sections V and VII that would help the real-time prices of energy and reserves better reflect the costs of maintaining reliability in the local areas. Since expectations of real-time prices are the primary driver of day-ahead prices, these changes should increase the day-ahead market commitment of generators that satisfy the local reliability criteria.

IX. Competitive Assessment

This section evaluates the competitive performance of the New England wholesale markets in 2008. This type of assessment is particularly important for LMP markets. While LMP markets increase overall system efficiency, they can also provide incentives for the localized exercise of market power in areas with inadequate generation resources or insufficient transmission capability. We identify geographic areas and market conditions that present the greatest potential for market power abuse. We use a methodology for measuring and analyzing potential withholding that was developed in prior assessments of the competitive performance in the New England markets.[80] In this section we address five main areas:

- Mechanisms by which sellers exercise market power in LMP markets;
- Structural market power indicators to assess competitive market conditions;
- Potential economic withholding;
- Potential physical withholding; and
- Conduct that increases NCPC payments.

A summary of our conclusions regarding the competitiveness of the wholesale market is included at the end of this section.

A. Market Power and Withholding

Supplier market power can be defined as the ability to profitably raise prices above competitive levels. In electricity markets, this is generally done by economically or physically withholding generating resources. Economic withholding occurs when a resource is offered at prices above competitive levels to reduce its output or otherwise raise the market price. Physical withholding occurs when all or part of the output of a resource is not offered to the market when it is available and economic to operate. Physical withholding can be accomplished by "derating" a generating unit (i.e., reducing the unit's high operating limit).

80 See, *e.g.*, *2007 Assessment of the Electricity Markets in New England.*

While many suppliers can cause prices to increase by withholding, not every supplier can profit from doing so. The benefit from withholding is that the supplier will be able to sell into the market at a clearing price above the competitive level. However, the cost of this strategy is that the supplier will lose profits from the withheld output. Thus, a withholding strategy is only profitable when the price impact overwhelms the opportunity cost of lost sales for the supplier. The larger a supplier is relative to the market, the more likely it is that the supplier will have the incentive to withhold resources to raise prices.

Other than the size of the market participant, there are several additional factors that affect whether a market participant has market power. First, if a supplier has already sold power in a forward market, then it will not be able to sell that power at an inflated clearing price in the spot market. Thus, forward power sales by large suppliers reduce their incentive to raise prices in the spot market.[81] Second, the incentive to withhold partly depends on the impact the withholding is expected to have on clearing prices. The nature of electricity markets is that when demand levels are high, a given quantity of withholding has a larger price impact than when demand levels are lower. Thus, large suppliers are more likely to possess market power during high demand periods than at other times.

Third, in order to exercise market power, a large supplier must have sufficient information about the physical conditions of the power system and actions of other suppliers to know that the market may be vulnerable to withholding. Since no supplier has perfect information, the conditions that give rise to market power (e.g., transmission constraints and high demand) must be reasonably predictable. The next section defines market conditions where certain suppliers possess market power.

B. Structural Market Power Indicators

The first step in a market power analysis is to define the relevant market, which includes the definition of a relevant product and the relevant geographic market where the product is traded.

81 When a supplier's forward power sales exceed the supplier's real-time production level, the supplier is a net buyer in the real-time spot market, thus, benefits from low rather than high prices.

Once the market definition is established, it is possible to assess conditions where one or more large suppliers could profitably raise prices. This sub-section of the report examines structural aspects of supply and demand affecting market power. We examine the behavior of market participants in later sections.

1. Defining the Relevant Market

Electricity is physically homogeneous, so each megawatt of electricity is interchangeable even though the characteristics of the generating units that produce the electricity vary substantially (*i.e.*, electricity from a coal-fired plant may be substituted with electricity from a nuclear power plant). Despite this physical homogeneity, the definition of the relevant product market is affected by the unique characteristics of electricity. For example, it is not generally economic to store electricity, so the market operator must continuously adjust suppliers' output to meet demand in real time. The lack of economic storage options limits inter-temporal substitution in spot electricity markets.

In defining the relevant product market, we must identify the generating capacity that can produce the relevant product. In this regard, we consider two categories of capacity: (i) online and fast-start capacity available for deployment in the real-time spot market, and (ii) offline non-fast-start capacity available for commitment in the next 24-hour timeframe. While only the former category is available to compete in the real-time spot market, both of these categories compete in the day-ahead market, making the day-ahead market less susceptible to market power. In general, forward markets are also less vulnerable to market power because buyers can defer purchases if they expect prices to be lower in the spot market. The market is most vulnerable to the exercise of market power in the real-time spot market, when only online and fast-start capacity is available for deployment. The value of energy in all other forward markets, including the day-ahead market, is derived from the value of energy in the real-time market. Hence, we define the relevant product for our analysis as energy produced in real time.

The second dimension of the market to be defined is the geographic area in which suppliers compete to sell the relevant product. In electricity markets, the relevant geographic market is generally defined by the transmission network constraints. Binding transmission constraints limit the extent to which power can flow between areas. When constraints are binding, a supplier within the constrained geographic area faces competition from fewer suppliers. There are a small number of geographic areas in New England that are recognized as being persistently constrained and, therefore, restricted at times from importing power from the rest of New England. When these areas are transmission-constrained, they constitute distinct geographic markets that must be analyzed separately. The following geographic markets are evaluated in this section:

- All of New England;
- All of Connecticut;
- West Connecticut;
- Southwest Connecticut;
- Norwalk-Stamford (located in Southwest Connecticut);
- Boston; and
- Lower SEMA.

This sub-section analyzes the seven geographic areas listed above using the following structural market power indicators:

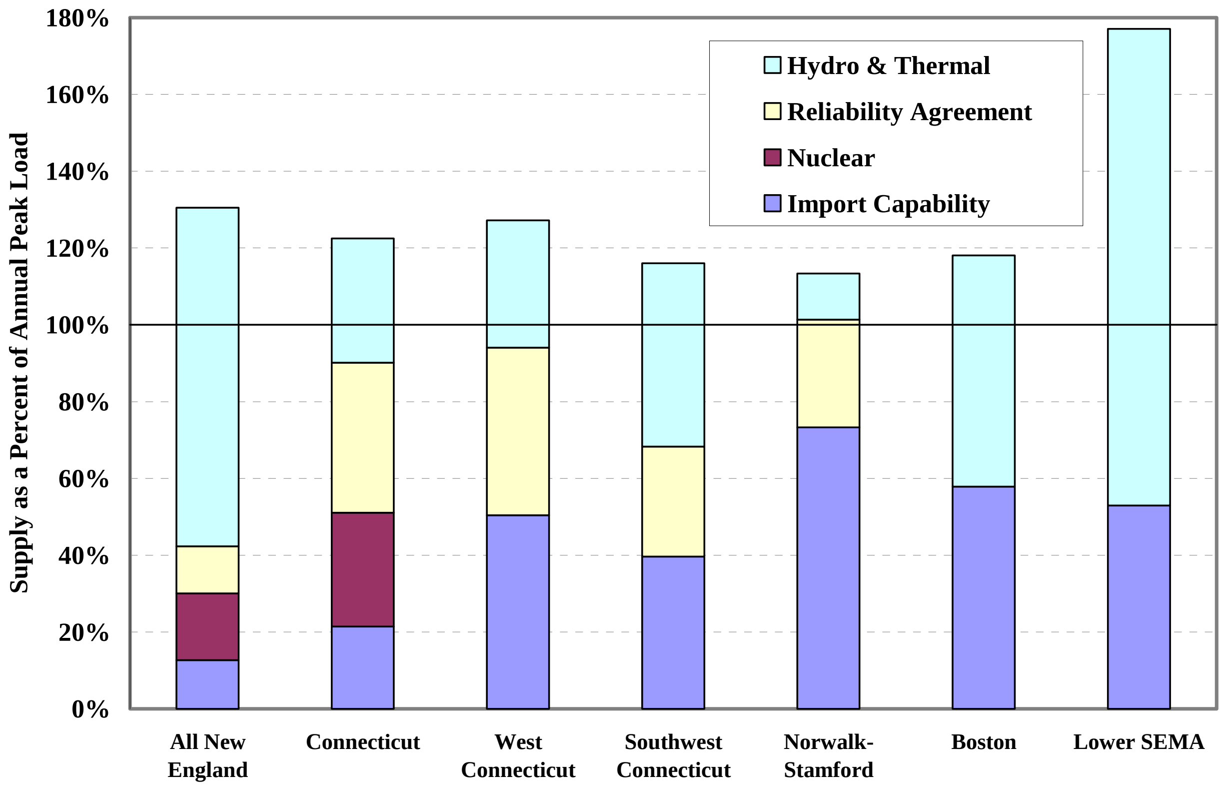
- Supplier market shares;
- Herfindahl-Hirschman indices; and
- Pivotal supplier indices.

The findings from the structural market power analysis in this section are used to focus the analyses of potential economic and physical withholding in Sections C and D.

2. Installed Capacity in Geographic Markets

This section provides a summary of supply resources and market shares in the geographic submarkets identified above. Each market can be served by a combination of native resources and imports. Native resources are limited by the physical characteristics of the generators in the area, while imports are limited by the capability of the transmission grid. Figure 34 shows several categories of supply relative to the load in each of the seven regions of interest.

Figure 34: Supply Resources versus Summer Peak Load in Each Region 2008



For each region, Figure 34 shows import capability[82] and three categories of installed summer capability: (i) nuclear units, (ii) units under reliability agreements, and (iii) all other generators. These resources are shown as a percentage of 2008 peak load, although a substantial quantity of additional capacity (typically more than 1,900 MW) is also necessary to maintain operating reserves in New England. The figure shows that while imports can be used to satisfy 13 percent of the load in the New England area under peak conditions, the six load pockets can serve larger shares of their peak load with imports. Norwalk-Stamford, which has the largest import capability relative to its size, was able to rely on imports to serve more than 70 percent of its load under peak conditions.

82 The import capability shown for each load pocket is the transfer capability during peak load hour, reduced to account for local reserve requirements.

In each region shown in Figure 34, the relative shares for categories of internal supply did not change significantly from 2007 to 2008. This is because the summer peak load levels in 2007 and 2008 were comparable, and because there were very few changes to the supply of internal resources in each region from 2007 to 2008. However, the amount of import capability into several regions changed substantially from 2007 to 2008. The import capability into Southwest Connecticut increased from 28 percent of peak load in 2007 to 40 percent in 2008, primarily due to transmission upgrades under Phase II of the Southwest Connecticut Reliability Project. Most components of this upgrade were placed in service during 2008, which significantly improved the transmission system infrastructure and increased the transfer capability into the Southwest Connecticut area. For the other regions shown in Figure 34, changes in import capability from 2007 to 2008 were primarily attributable to the differences in network topology, generation patterns, and load patterns during peak load hours in the two years.

Figure 34 also shows the margin between peak load and the total available supply from imports and native resources. For each region, the total supply exceeded peak load by at least 10 percent, ranging from 13 percent in Norwalk-Stamford to 77 percent in Lower SEMA. Areas with lower margins may be more susceptible to withholding than other areas.

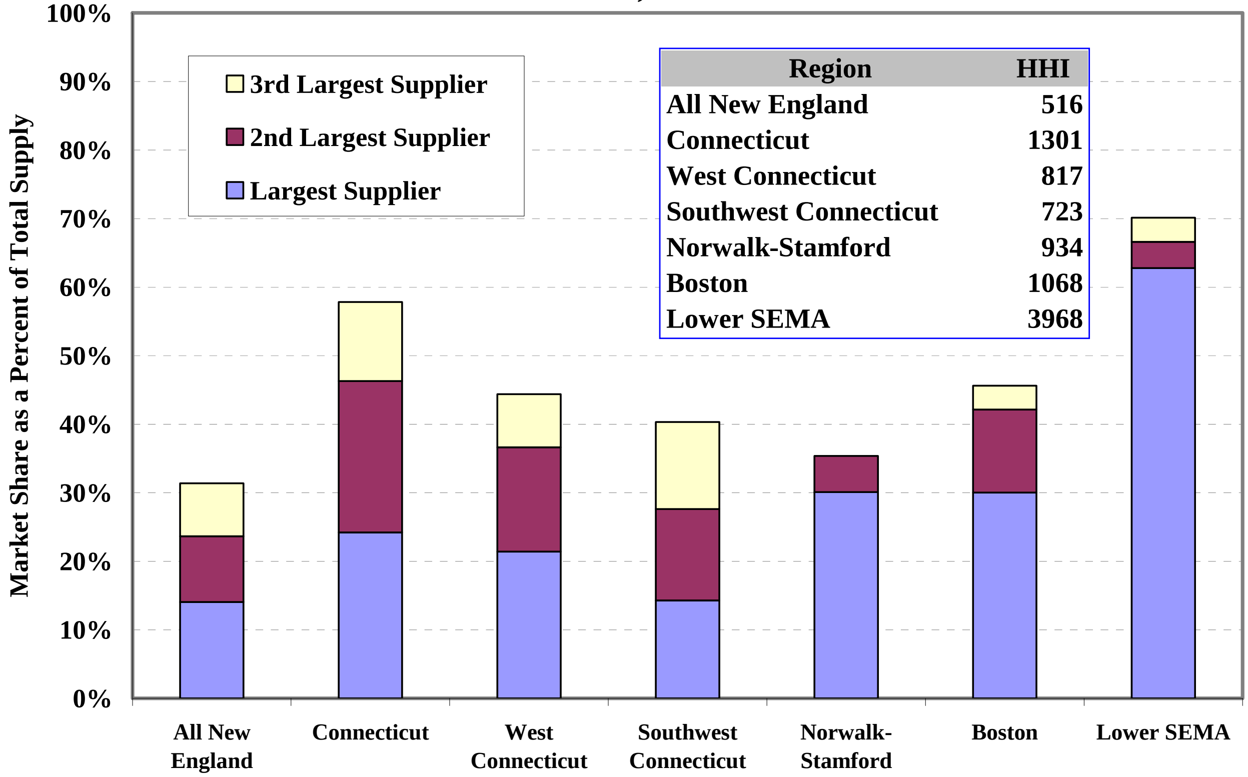
Nuclear capacity and capacity under reliability agreements are shown separately from other internal generation because these resources are likely to pose fewer market power concerns. In order to exercise market power successfully in an electricity market, it is important to be able to withhold capacity only at times when it will be profitable. This is the case because the lost revenue on withheld units can be very costly. Nuclear generators typically cannot be dispatched up and down in a manner that would allow the owner of the unit to profitably withhold. Thus, the owner of nuclear generation would have to also own significant amounts of non-nuclear capacity that could be withheld from the market. The owner of a unit under a reliability agreement cannot economically withhold the unit because the owner is obligated to offer the unit at short-run marginal cost. The short-run marginal cost is reviewed by the ISO's internal Market Monitoring department on an on-going basis. The owner of a unit under a reliability agreement has a strong disincentive to physically withhold because the fixed cost payments are reduced if the unit fails to meet its target available hours as specified in the reliability agreement.

Connecticut continued to rely heavily on nuclear capacity and units under reliability agreements in 2008. Reliability agreements reduce the quantity of capacity that may be withheld to exercise market power. In the Norwalk-Stamford load pocket, approximately 70 percent of internal generation was under a reliability agreement in 2008, reducing the potential to exercise market power in this area.

Market power is generally of greater concern in areas where capacity margins are small. However, the extent of market power also depends on the market shares of the largest suppliers. For each region, Figure 35 shows the market shares of the largest three suppliers coinciding with the annual peak load hour on June 10, 2008. The remainder of supply to each region comes from smaller suppliers and import capability. We also show the Herfindahl-Hirschman Index ("HHI") for each region. The HHI is a standard measure of market concentration calculated by summing the square of each participant's market share. In our analysis, we assume imports are highly competitive by treating the market share of imports as zero in the HHI calculation. For example, in a market with two suppliers and import capability, all of equal size, the HHI would be close to $2200 = [(33\%)^2 + (33\%)^2 + (0\%)^2]$. This assumption tends to understate the true level of concentration, because, in reality, the market outside of the area is not perfectly competitive, and because suppliers inside the area may be affiliated with resources in the market outside of the area.

Figure 35 indicates a substantial variation in market structure across New England. In all New England, the largest supplier had a 14 percent market share in 2008. In the load pockets, the largest suppliers had market shares ranging from 14 percent in Southwest Connecticut to 63 percent in Lower SEMA. Likewise, there is variation in the number of suppliers that have significant market shares. For instance, Norwalk-Stamford had only two native suppliers with unequal market shares in 2008, while Southwest Connecticut had three native suppliers with comparable market shares.

Figure 35: Installed Capacity Market Shares for Three Largest Suppliers
June 10, 2008



The HHI figures suggest that only Lower SEMA is highly concentrated.[83] The HHI for Norwalk-Stamford is 934, which is relatively low for most product markets. This is counter-intuitive since there are only two suppliers in the area. However, because 73 percent of the load can be served by imports, the local suppliers only need to serve 27 percent of the market. Of the remaining areas, Connecticut and Boston have the highest HHI statistics with 1,301 and 1,068, respectively.

While HHI statistics can be instructive in generally indicating the concentration of the market, they do not allow one to draw reliable conclusions regarding potential market power in wholesale electricity markets due to the special nature of the electricity markets. In particular, HHI does not consider demand conditions, load obligations, or the heterogeneous effects of generation on transmission constraints based on their location. In the next sub-section, we

83 The antitrust agencies and the FERC consider markets with HHI levels above 1800 as highly concentrated for purposes of evaluating the competitive effects of mergers.

evaluate the potential for market power using a pivotal supplier analysis, which addresses the shortcomings of concentration analyses.

3. Pivotal Supplier Analysis

While HHI statistics can provide reliable competitive inferences for many types of products, this is not generally the case in electricity spot markets.[84, 85] The HHI's usefulness is limited by the fact that it reflects only the supply-side, ignoring demand-side factors that affect the competitiveness of the market. The most important demand-side factor is the level of load relative to available supply-side resources. Since electricity cannot be stored economically in large volumes, production needs to match demand in real time. When demand rises, an increasing quantity of generation is utilized to satisfy the demand, leaving less supply that can respond by increasing output if a large supplier withholds resources. Hence, markets with higher resource margins tend to be more competitive, which is not recognized by the HHI statistics.

A more reliable means to evaluate the competitiveness of spot electricity markets and recognize the dynamic nature of market power in these markets is to identify when one or more suppliers are "pivotal". A supplier is pivotal when the output of some of its resources is needed to meet demand in the market. A pivotal supplier has the ability to unilaterally raise the spot market prices to arbitrarily high levels by offering its energy at a very high price level. Hence, the market may be subject to substantial market power abuse when one or more suppliers are pivotal and have the incentive to take advantage of their position to raise prices. The Federal Energy Regulatory Commission has adopted a form of pivotal supplier test as an initial screen for market

84 It is true that the DOJ and FTC evaluate the *change* in HHI as part of its merger analysis. However, this is only a preliminary analysis that would typically be followed by a more rigorous simulation of the likely price effects of the merger. It is also important to note the HHI analysis employed by the antitrust agencies is not intended to determine whether a supplier has market power.

85 For example, see Severin Borenstein, James B. Bushnell, and Christopher R. Knittel, "Market Power in Electricity Markets: Beyond Concentration Measures," *Energy Journal* 20(4), 1999, pp. 65-88.

power in granting market-based rates.[86] This section of the report identifies the frequency with which one or more suppliers were pivotal in areas within New England during the study period.

Even small suppliers can be pivotal for brief periods. For example, all suppliers are pivotal during periods of shortage. This does not mean that all suppliers should be deemed to have market power. As described above, suppliers must have both the *ability* and *incentive* to raise prices to have market power. For a supplier to have the ability to substantially raise real-time energy prices, it must be able to foresee that it will likely be pivotal. In general, the more frequently a supplier is pivotal, the easier it will be for it to foresee circumstances when it can raise the clearing price.

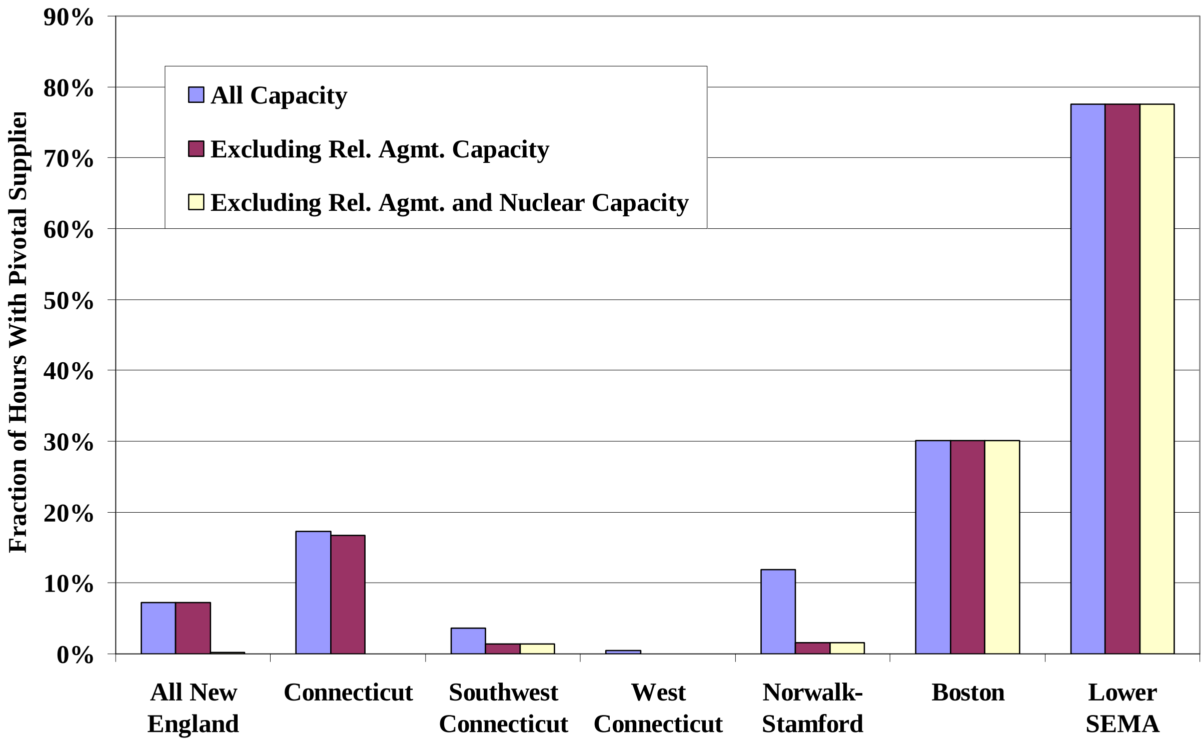
To identify the areas where market power is a potential concern most frequently, Figure 36 shows the portion of hours where at least one supplier was pivotal in each region during 2008. The figure also shows the impact of excluding nuclear units and units under reliability agreements from the analysis. As discussed above, such units are unlikely to be engaged in economic or physical withholding.

Including all capacity, the pivotal supplier analysis raises potential concerns regarding five of the seven areas shown in Figure 36. The areas that do not raise potential concerns are Southwest Connecticut and West Connecticut, where imports typically serve a large share of load and the ownership of internal capacity is much less concentrated than the other load pockets. Potential local market power concerns are most acute in Lower SEMA, where one supplier owns nearly 90 percent of the internal capacity.

Boston exhibited the second-highest pivotal supplier frequency because the largest supplier there owns nearly 60 percent of the internal capacity. In Boston, none of the largest supplier's capacity was nuclear capacity or under a reliability agreement during 2008.

86 The FERC test is called the "Supply Margin Assessment". For a description, see: Order On Rehearing And Modifying Interim Generation Market Power Analysis And Mitigation Policy, 107 FERC ¶ 61,018, April 14, 2004.

Figure 36: Frequency of One or More Pivotal Suppliers by Type of Withheld Capacity 2008



In Norwalk-Stamford, more than 80 percent of the largest supplier's capacity was under a reliability agreement in 2008. As a result, the fraction of hours in which a supplier was pivotal decreases significantly (from 12 percent to only 2 percent) if reliability agreement capacity is excluded.

Although Connecticut had a pivotal supplier in 17 percent of the hours in 2008, the largest supplier in Connecticut owns only nuclear capacity. In order to exercise market power, the largest supplier would need to withhold from non-nuclear resources.[87] Therefore, it is appropriate to exclude the nuclear capacity from the pivotal supplier frequency for Connecticut. This leaves no hours when a supplier was pivotal in Connecticut.

For the entirety of New England, because none of the largest three suppliers had resources under a reliability agreement in 2008, the market power conclusions depend primarily on how nuclear

87 This assumes that the supplier cannot reduce its nuclear output substantially without taking a unit out of service.

capacity affects the incentives of large suppliers. Excluding nuclear capacity from the pivotal supplier analysis for all of New England substantially reduces the pivotal frequency (from 7 percent to less than 1 percent of hours). However, the rationale for excluding nuclear capacity from the analysis does not apply to the largest suppliers in New England. These suppliers have large portfolios with a combination of nuclear and non-nuclear capacity, and while they are not likely to physically withhold their nuclear capacity from the market, their nuclear capacity would earn more revenue if they withheld their non-nuclear capacity. Accordingly, New England as a whole warrants further review.

The pivotal frequency changed significantly in several areas from 2007 to 2008.[88] In Southwest Connecticut, the pivotal frequency declined from 12 percent in 2007 to 4 percent in 2008, primarily due to the transmission upgrades made under Phase II of the Southwest Connecticut Reliability Project. In Connecticut and Boston, the pivotal frequency rose modestly from the prior year, from 14 percent to 17 percent in Connecticut, and from 25 percent to 30 percent in Boston. These increases were partly driven by the declines in reliability commitment, which reduced the supply margin in these areas compared to 2007.[89]

The pivotal supplier summary in Figure 36 indicates the greatest potential for market power in Lower SEMA and Boston. A close examination is also warranted for all of New England, while Connecticut raises lesser concerns. Each area had a single supplier that was most likely to have market power. Accordingly, Sections C and D closely examine the behavior of the largest single supplier in each geographic market.

As described previously, market power tends to be more prevalent as the level of demand grows. In order to strategically withhold, a dominant supplier must be able to reasonably foresee its

---

88 For all of New England, the pivotal frequency remained consistent with the previous year (8 percent in 2007 and 7 percent in 2008). However, in "2007 Assessment of the Electricity Markets in New England," we reported a pivotal frequency of 17 percent in 2007 for all of New England because the analysis did not consider unused import capability as potential supply.

89 The changes in reliability commitment pattern are discussed in detail in Section VII.B.

opportunities to raise prices. Since load levels are relatively predictable, a supplier with market power could focus its withholding strategy on periods of high demand.

To assess when withholding is most likely to be profitable, Figure 37 shows the fraction of hours when a supplier is pivotal at various load levels. The bars in each load range show the fraction of hours when a supplier was pivotal in All New England, Lower SEMA, and Boston. The bars are arranged according to the frequency with which a supplier is pivotal, from lowest to highest. For example, Lower SEMA had the highest frequency of a supplier being pivotal and is, therefore, shown on the far right. West Connecticut, Southwest Connecticut, and Norwalk-Stamford are not shown because in those areas there were very few instances of a supplier being pivotal during 2008. Connecticut is not shown because the only pivotal supplier had exclusively nuclear capacity, which is not expected to provide that supplier with an incentive to withhold.

Figure 37: Frequency of One or More Pivotal Suppliers by Load Level Excluding Capacity under Reliability Agreements, 2008

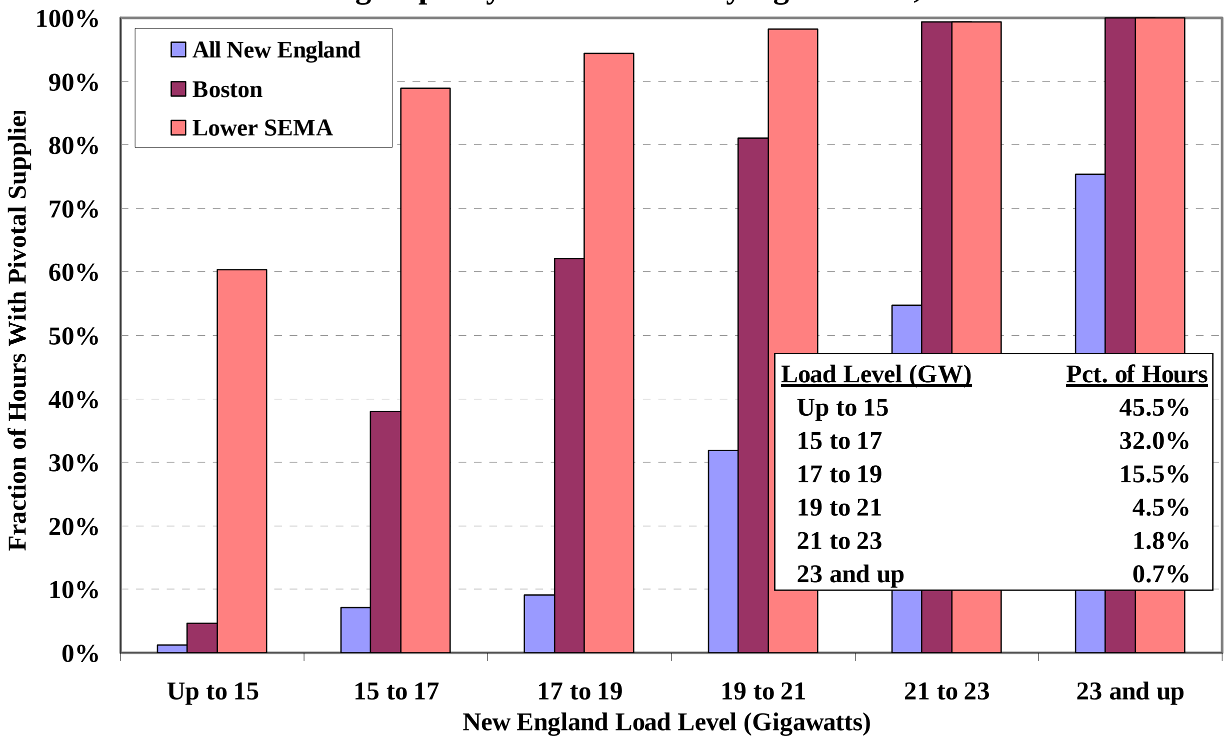


Figure 37 indicates that the largest supplier in Lower SEMA was pivotal in more than 90 percent of hours when the load exceeded 15 GW in New England, and in 60 percent of all hours when load was *less* than 15 GW in New England. The analysis indicates less potential for market power in the other areas. The supplier in Boston was pivotal in at least 60 percent of hours when the load exceeded 17 GW in New England. In All of New England, the largest supplier was pivotal in more than half of the hours when load exceeded 21 GW. The pivotal frequency fell below 5 percent in Boston and All of New England during hours when load was below 15 GW in New England.

Based on the pivotal supplier analysis in this sub-section, market power is most likely to be a concern in Lower SEMA at all load levels, in Boston when load exceeds 17 GW, and in All of New England when load exceeds 19 GW. The pivotal supplier results are conservative for "All of New England" because the analysis assumed that imports would not change if the largest supplier were to withhold. In reality, there would be some increase in imports. The following sections examine the behavior of pivotal suppliers under various load conditions to assess whether the behavior has been consistent with competitive expectations.

C. Economic Withholding

Economic withholding occurs when a supplier raises its offer prices substantially above competitive levels to raise the market price. Therefore, an analysis of economic withholding requires a comparison of actual offers to competitive offers.

Suppliers lacking market power maximize profits by offering resources at marginal costs. A generator's marginal cost is the incremental cost of producing additional output, including inter-temporal opportunity costs, incremental risks associated with unit outages, fuel, additional operating and maintenance ("O&M"), and other incremental costs attributable to the incremental output. For most fossil-fuel resources, marginal costs are closely approximated by their variable production costs (primarily fuel inputs, labor, and variable O&M). However, at high output levels or after having run long periods without routine maintenance, outage risks and expected increases in O&M costs can create substantial additional incremental costs. Generating resources with energy limitations, such as hydroelectric units or fossil-fuel units with output restrictions as a result of environmental considerations, must forego revenue in a future period when they produce in the current period. These units incur an inter-temporal opportunity cost associated with producing that can cause their marginal costs to be much larger than their variable production costs.

Establishing a proxy for units' marginal costs as a competitive benchmark is a key component of this analysis. This is necessary to determine the quantity of output that is potentially economically withheld. The ISO's Internal Market Monitoring Unit calculates generator cost reference levels pursuant to Appendix A of Section III of the ISO's Tariff. These reference levels are used as part of the market power mitigation measures and are intended to reflect the competitive offer price for a resource. The Internal Market Monitoring Unit has provided us with cost reference levels, which we can use as a competitive benchmark in our analysis of economic withholding.

1. Measuring Economic Withholding

We measure economic withholding by estimating an output gap for units that fail a conduct test for their start-up, no-load, and incremental energy offer parameters, indicating that they are submitting offers in excess of competitive levels. The output gap is the difference between the unit's capacity that is economic at the prevailing clearing price and the amount that is actually produced by the unit. In essence, the output gap shows the quantity of generation that is withheld from the market as a result of having submitted offers above competitive levels. Therefore, the output gap for any unit would generally equal:

$Q_i^{econ} - Q_i^{prod}$ when greater than zero, where:

Q_i^{econ} = Economic level of output for unit i; and

Q_i^{prod} = Actual production of unit i.

To estimate Q_i^{econ}, the economic level of output for a particular unit, it is necessary to evaluate all parts of the unit's three-part reference level: start-up cost reference, no-load cost reference, and incremental energy cost reference. These costs jointly determine whether a unit would have been economic at the clearing price for at least the unit's minimum run time. We employ a three-stage process to determine the economic output level for a unit in a particular hour. In the first step, we examine whether the unit would have been economic *for commitment* on that day if it had offered at its marginal costs (i.e., whether the unit would have recovered its actual start-up, no-load, and incremental costs running at the dispatch point dictated by the prevailing LMP,constrained by its EcoMin and EcoMax, for its minimum run time). If a unit was economic for commitment, we then identify the set of contiguous hours during which the unit was economic to have online. Finally, we determine the economic level of incremental output in hours when the unit was economic to run. In all three steps, the marginal costs assumed for the generator are the reference levels for the unit used in the ISO's mitigation measures plus a threshold.

In hours when the unit was not economic to run and on days when the unit was not economic for commitment, the economic level of output was considered to be zero. To reflect the timeframe in which commitment decisions are actually made, this assessment is based on day-ahead market outcomes for non-fast-start units, and based on real-time market outcomes for fast-start units.

Q_i^{prod} is the actual observed production of the unit. The difference between Q_i^{econ} and Q_i^{prod} represents how much the unit fell short of its economic production level. However, some adjustments are necessary to estimate the actual output gap because some units are dispatched at levels lower than their three-part offers would indicate. This can be due either to transmission constraints, reserve considerations, or changes in market conditions between the time when unit commitment is performed and real-time. Therefore, we adjust Q_i^{prod} upward to reflect three-part offers that would have made a unit economic to run, even though the unit may not have been fully dispatched. For example, if the ISO manually reduces the dispatch of an economic unit, the reduction in output is excluded from the output gap. Hence the output gap formula used for this report is:

$$Q_i^{econ} - \max(Q_i^{prod}, Q_i^{offer}) \text{ when greater than zero, where:}$$

$$Q_i^{offer} = \text{ offer output level of i.}$$

By using the greater of actual production or the output level offered at the clearing price, portions of units that are constrained by ramp limitations are excluded from the output gap. In addition, portions of resources that are offered above marginal costs due to a forward reserve market obligation are not included in the output gap.

It is important to recognize that the output gap tends to overstate the amount of potential economic withholding because some of the offers that are included in the output gap reflect legitimate responses by the unit's owner to operating conditions, risks, or uncertainties. For example, some hydro units are able to produce energy for a limited number of hours before running out of water. Under competitive conditions, the owners of such units have incentives to produce energy during the highest priced periods of the day, and they attempt to do this by raising their offer prices so that their units will be dispatched only during the highest priced periods of the day. However, the owners of such units submit offers prior to 6 pm on the previous day based on their expectations of market conditions, so if real-time prices are lower than expected, to the unit may have an output gap. Hence, output gap is not necessarily evidence of withholding, but it is a useful indicator of potential withholding. We generally seek to identify trends in the output gap results that would indicate significant attempts to exercise market power.

We have observed that some units that expect to be committed for local reliability and receive NCPC payments also produce above average output gap. One explanation is that these units raise their offers in expectation of receiving higher NCPC payments and are not dispatched as a result. Such instances are flagged as output gap, even though the suppliers are not withholding in an effort to raise LMPs.

In this section we evaluate the output gap results relative to various market conditions and participant characteristics. The objective is to determine whether the output gap increases when those factors prevail that can create the ability and incentive for a pivotal supplier to exercise market power. This allows us to test whether the output gap varies in a manner consistent with attempts to exercise market power. Based on the pivotal supplier analysis from the previous sub-section, the level of market demand is a key factor in determining when a dominant supplier is most likely to possess market power in some geographic market.

In this section, we examine output gap results by load level in the following areas:

- Boston;
- Lower SEMA;
- Connecticut; and
- All of New England.

2. Output Gap in Boston

Boston is a large net-importing region, making it particularly important to evaluate the conduct of its suppliers. Furthermore, the pivotal supplier analysis raises concerns regarding the potential exercise of market power in Boston where one supplier owns the majority of capacity.

Figure 38 shows output gap results for Boston by load level. Output gap statistics are shown for the largest supplier compared with all other suppliers in the area. Based on the pivotal supplier analysis in the previous sub-section, the largest supplier can expect that its capacity will be pivotal in most hours when load exceeds 17 GW.

**Figure 38: Average Output Gap by Load Level and Type of Supplier
Boston, 2008**

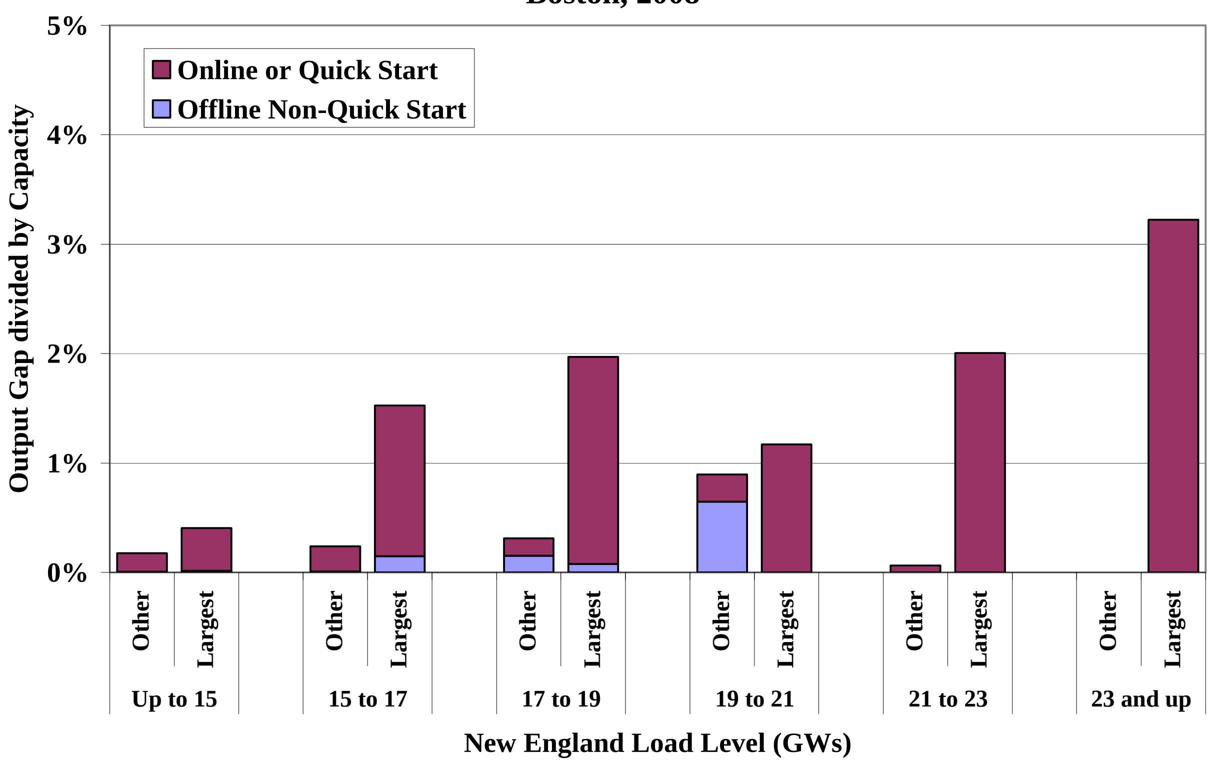


Figure 38 shows that the overall amount of output gap in Boston was modest in 2008, although the amount of output gap increased at higher load levels. The output gap for the largest supplier was less than one half percent of its total capacity when the New England load was below 15 GW, and ranged from approximately 1 to 3 percent when the load exceeded 15 GW. The output gap for other suppliers in Boston was less than 1 percent of their total capacity on all load levels.

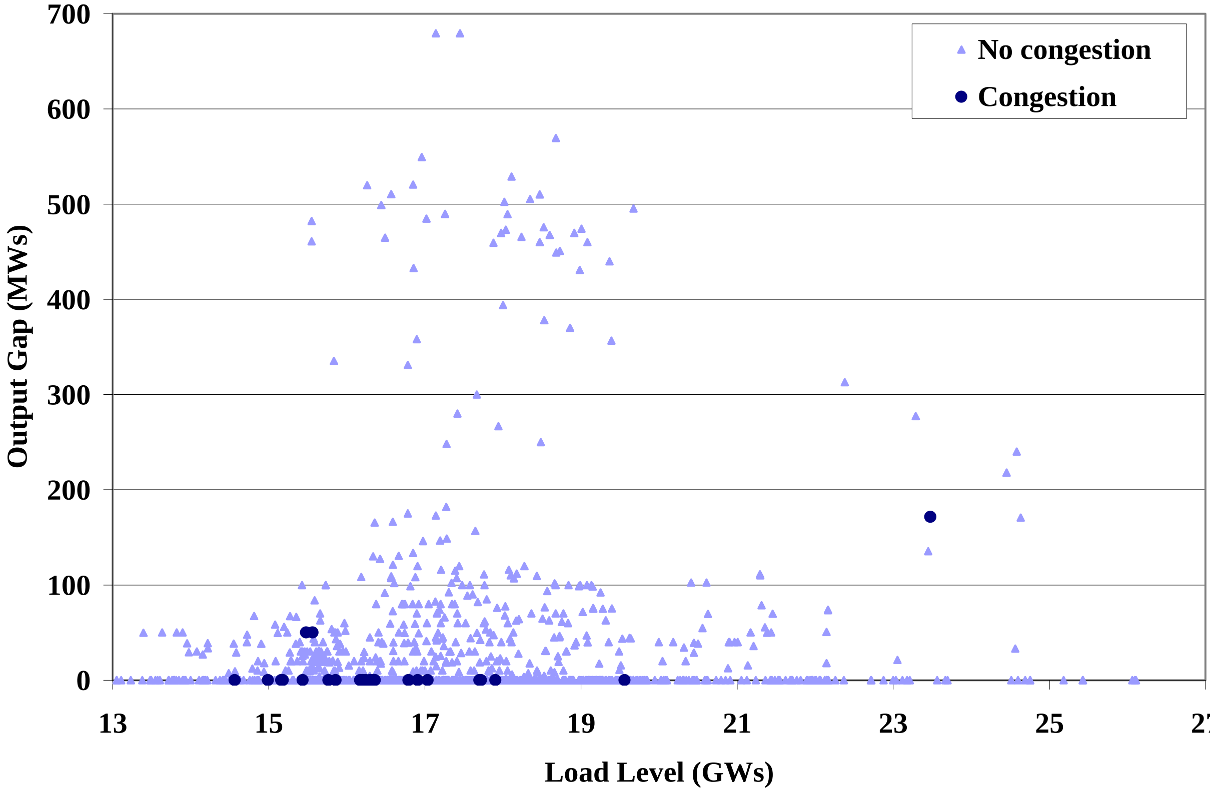
The output gap for the largest supplier was substantially higher than the levels for the other suppliers, and rose slightly during high load periods in absolute terms. These observations raise concerns regarding potential economic withholding by the largest supplier to raise prices in Boston, which warrants additional evaluation.

The next analysis evaluates whether the output gap of the largest supplier in Boston had a significant effect on LMPs in the real-time market by analyzing the conditions under which the conduct occurred. A strategy to raise LMPs would be most effective during periods of transmission congestion into Boston, particularly under high load conditions.

The following figure examines the correlation between the load levels in New England and the output gap of the largest supplier, while separately indicating hours with congestion into Boston. Figure 39 shows the output gap associated with online and fast-start units of the largest supplier on an hourly basis for the highest three load hours of each day. Hours with real-time congestion into Boston are denoted with dark blue circles, while the remaining days are shown with light blue triangles.

There was congestion into Boston in a relatively small number of the hours shown in Figure 39. There was just 1 hour with congestion into Boston when the largest supplier's online units had an output gap of more than 100 MW, and another 2 congested hours in which the largest supplier's online units had an output gap greater than 0 MW. Based on these results, it is unlikely that the largest supplier's conduct had a substantial effect on LMPs in Boston in 2008.

Figure 39: Hourly Output Gap of Largest Supplier in Boston versus Load Online and Fast-Start Units, Top Three Load Hours per Day, 2008



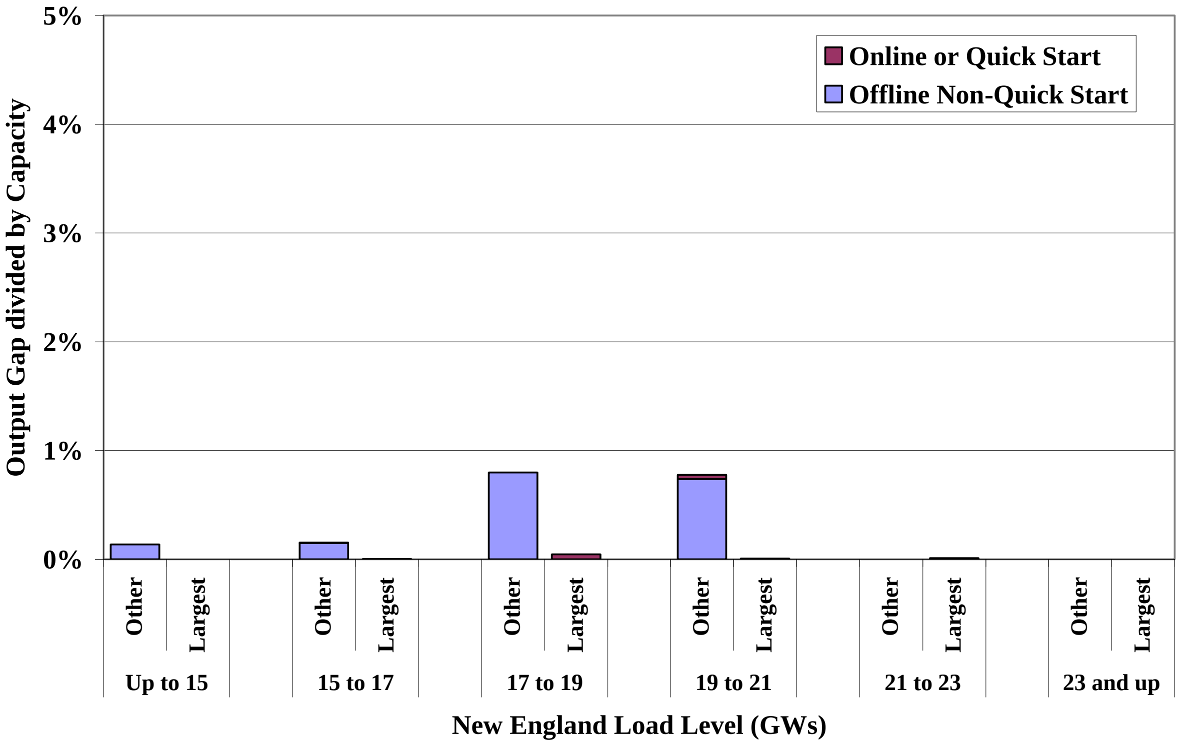
The output gap of the largest supplier was associated with one generator that was frequently committed for local reliability reasons. The supplier offered the generator at elevated prices, which is consistent with the "pay-as-bid" incentives that a supplier faces when it receives compensation based primarily on its offer price instead of the market clearing price. Even in pay-as-bid markets that are perfectly competitive, firms with no market power rationally raise their offer above marginal costs. Although offers by competitive suppliers in pay-as-bid markets rise above marginal costs, one cannot conclude that all increases in offer prices by suppliers that face pay-as-bid incentives are competitive. To the extent suppliers hold resources needed to meet local reliability requirements, they may have local market power that can be exercised by inflating the NCPC payments that they receive when committed for reliability. These concerns are discussed further in Section E.

3. Output Gap in Lower SEMA

The pivotal supplier analysis in the previous section indicates that the largest supplier (who owns 89 percent of the capacity) in Lower SEMA was pivotal in the majority of hours in 2008. For this reason, we closely evaluate the competitiveness of the conduct of suppliers in Lower SEMA.

The following analysis examines output gap patterns in Lower SEMA to determine whether there is evidence of economic withholding. Figure 40 shows the output gap identified in Lower SEMA in 2008 by load level. The output gap is shown separately for the largest supplier and for other suppliers.

Figure 40: Average Output Gap by Load Level and Type of Supplier
Lower SEMA, 2008

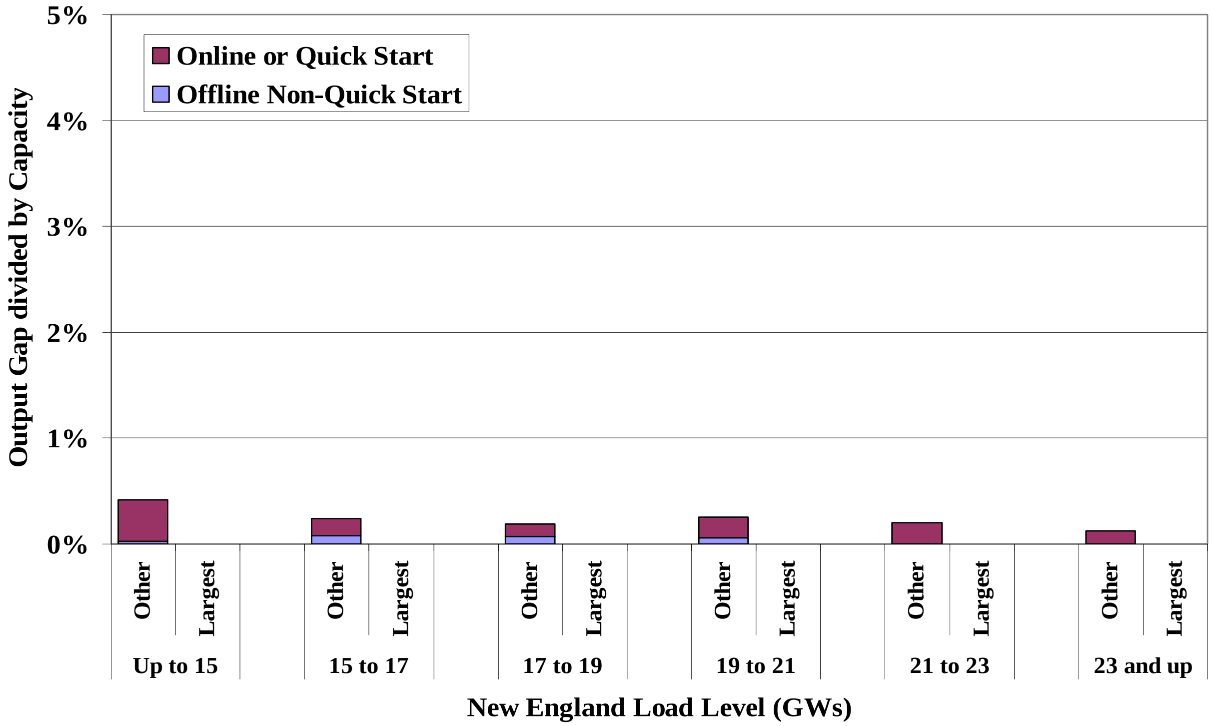


Although the pivotal supplier analysis identified Lower SEMA as the region with the highest potential for market power, the figure shows that the output gap of the largest supplier's online and quick-start units was not significant in 2008. These results clearly indicate that the largest supplier's offers have consistently been priced below the Conduct Threshold for Economic Withholding, which is identified in Appendix A of Market Rule 1.

4. Output Gap in Connecticut

In this sub-section, we examine potential economic withholding in Connecticut. Historically, Connecticut has been import-constrained. However, the pivotal supplier analysis does not raise significant concerns about the potential exercise of market power in 2008 in Connecticut. Figure 41 shows output gap results for Connecticut by load level. Output gap statistics are shown for the largest supplier compared with all other suppliers in the area.

Figure 41: Average Output Gap by Load Level and Type of Supplier Connecticut, 2008

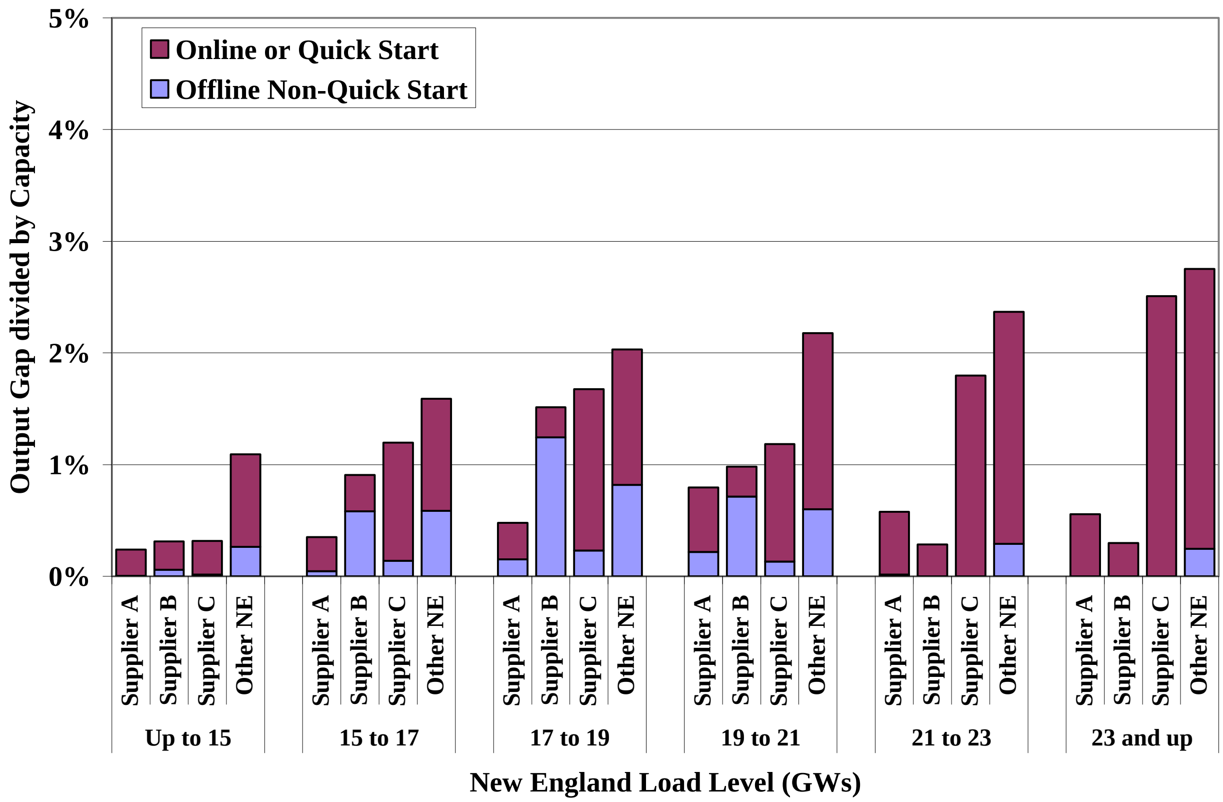


The pivotal supplier analysis indicates that the largest supplier in Connecticut was pivotal in only 1 percent of all hours. Figure 41 also shows that the total output gap of all suppliers was very low relative to the total capacity in Connecticut. The largest supplier owns exclusively nuclear capacity and had no output gap in 2008. The other suppliers also produced very little output gap. These results do not raise concerns regarding economic withholding in Connecticut.

5. Output Gap in All New England

Figure 42 summarizes output gap results for all of New England by load level for four categories of supply. Supplier A had the largest portfolio in New England and was pivotal in approximately 7 percent of the hours during 2008 (excluding capacity under reliability agreements). Suppliers B and C are the second and third largest suppliers in New England and were pivotal during less than 1 percent of the hours. All other suppliers are shown as a group for reference.

Figure 42: Average Output Gap by Load Level and Type of Supplier
All New England, 2008



The figure shows that the region-wide output gap was generally low for each of the four categories of supply, although some categories exhibited higher output gap quantities at higher load levels. Supplier A exhibited a small output gap under all load conditions. Supplier B exhibited a small output gap under all load conditions, although it was somewhat higher in the load range from 15 to 21 GW. Supplier C, which exhibited a higher output gap than Suppliers A and B, is also the largest supplier in Boston. As described above, its output gap can be attributed to conduct designed to raise its NCPC payments rather than to raise real-time LMPs. It is notable that the output gap levels for the three largest suppliers were lower than the output gap levels of all other suppliers. A close review of the underlying data indicates that hydro units tend to exhibit larger output gap quantities. Because these output gap levels are relatively low and the largest suppliers' output gap amounts are lower than the levels for other suppliers (which are not likely to have market power), economic withholding was not a significant concern in New England in 2008.

D. Physical Withholding

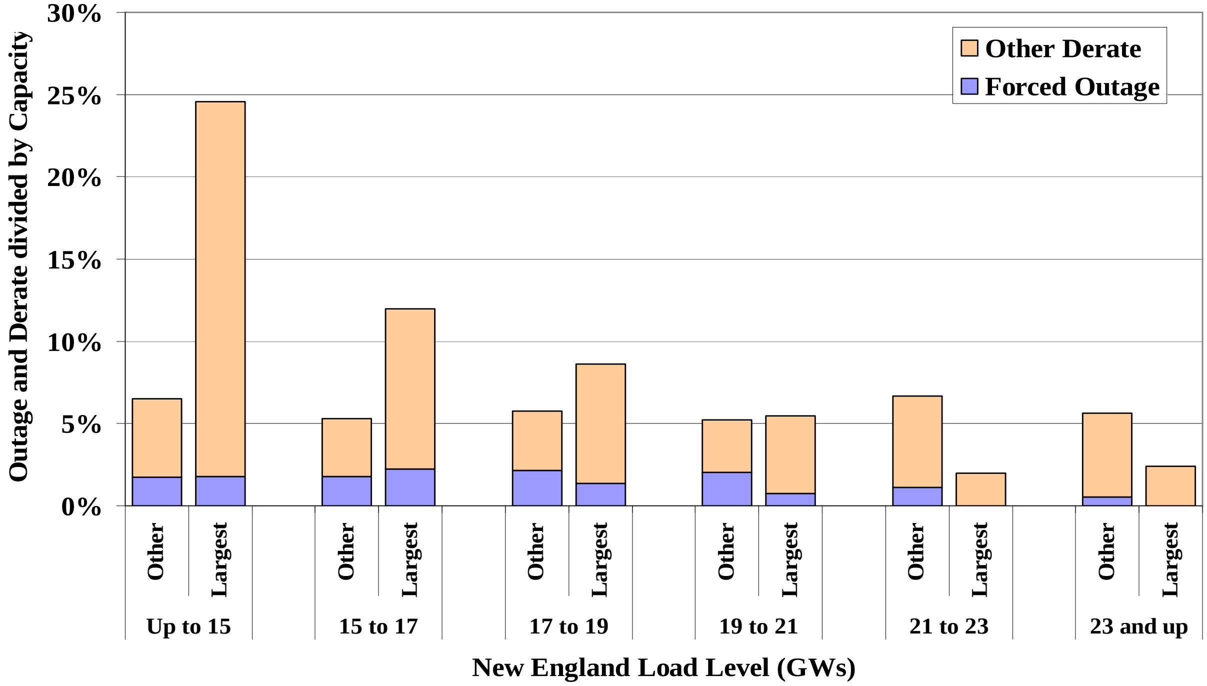
This section of the report examines declarations of forced outages and other non-planned deratings to assess whether they have occurred in a manner that is consistent with the exercise of market power. In this analysis, we evaluate the four geographic markets examined in the output gap analysis above.

In each market, we examine forced outages and other deratings by load level. The "Other Derate" category includes reductions in the hourly capability of a unit from its maximum seasonal capability that are not logged as forced outages or planned outages. These deratings can be the result of ambient temperature changes or other factors that affect the maximum capability of a unit.

1. Potential Physical Withholding in Boston

Figure 43 shows declarations of forced outages and other non-planned deratings in Boston by load level. Based on the pivotal supplier analysis, the capacity of the largest supplier can be expected to be pivotal in most hours when New England load exceeds 17 GW. We compare these statistics for the largest supplier to all other suppliers in the area.

Figure 43: Forced Outages and Deratings by Load Level and Supplier Boston, 2008



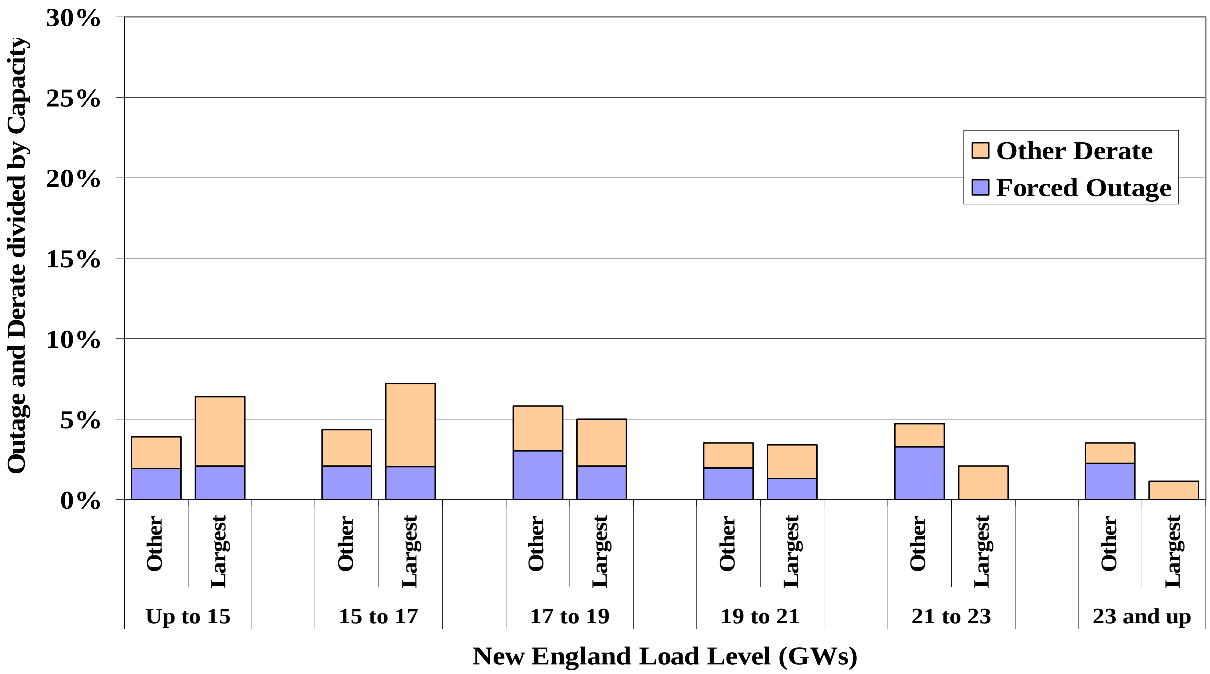
The figure shows the largest supplier's physical deratings as a percentage of its portfolio. The rate of other non-planned outages ('Other Derate' Category) was high at low load levels in 2008, especially when load was less than 15 GW. This was primarily driven by two units that were frequently online in special operating modes (where a portion of the capacity is not available) in early morning hours. This operating practice does not raise competitive concerns because it does not substantially affect prices in these hours. Additionally, this operating practice is economic during low load conditions and, therefore, consistent with competitive conduct.

Figure 43 shows a pattern of deratings and outages consistent with expectations in a competitive market. Although levels of outages and deratings for the largest supplier were high at low load levels, they were comparable or even lower than those of other suppliers when load exceeded 19 GW (when withholding is most likely to be profitable). Furthermore, the largest supplier showed a relatively low level of outages and deratings as load increased to the highest load levels. Even though running units more intensely under peak demand conditions increases the probability of an outage, the results shown in the figure suggest that the largest supplier increased the availability of its capacity during periods of high load when capacity was most valuable to the market. Overall, the outage and deratings results for Boston do not raise concerns of strategic withholding.

2. Potential Physical Withholding in Lower SEMA

Figure 44 summarizes declarations of forced outages and other deratings in the Lower SEMA area by load level in 2008. These statistics are shown for the largest supplier compared with the other suppliers in the area. Based on the pivotal supplier analysis for Lower SEMA, the largest supplier can be expected to be pivotal in 78 percent of the hours. This result indicates a potential market power concern in Lower SEMA even under moderate load conditions.

Figure 44: Forced Outages and Deratings by Load Level and Supplier Lower SEMA, 2008



The largest supplier had a modestly higher rate of forced outages and other deratings than the other suppliers in Lower SEMA under low load levels. When the load exceeded 17 GW, the rates of forced outages and other deratings for the largest supplier fell below the levels for the other suppliers. Furthermore, as load levels rose, forced outages and other deratings declined. These patterns suggest efforts to increase unit availability as load rises, rather than any attempt at physical withholding. Overall, the outage and deratings results for Lower SEMA do not raise concerns of strategic withholding.

3. Potential Physical Withholding in Connecticut

Figure 45 summarizes declarations of forced outages and other deratings in Connecticut by load level. The figure shows these statistics for the largest supplier of capacity in the area and for other suppliers.

Figure 45: Forced Outages and Deratings of the Largest Supplier by Load Level Connecticut, 2008

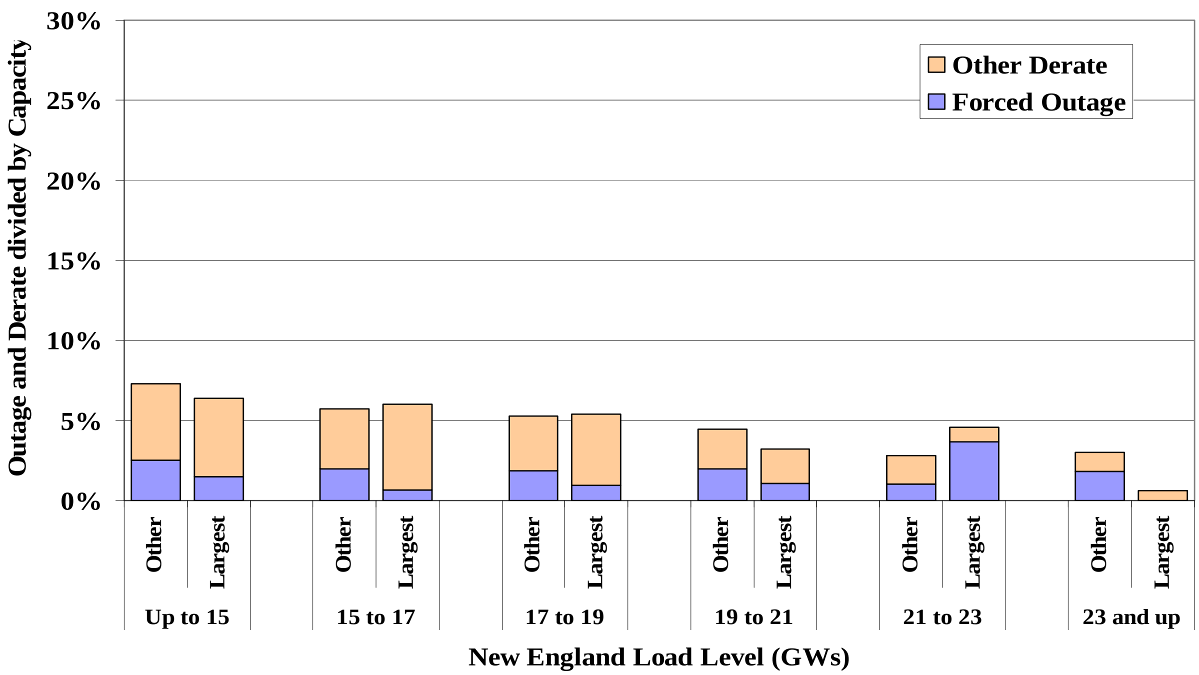
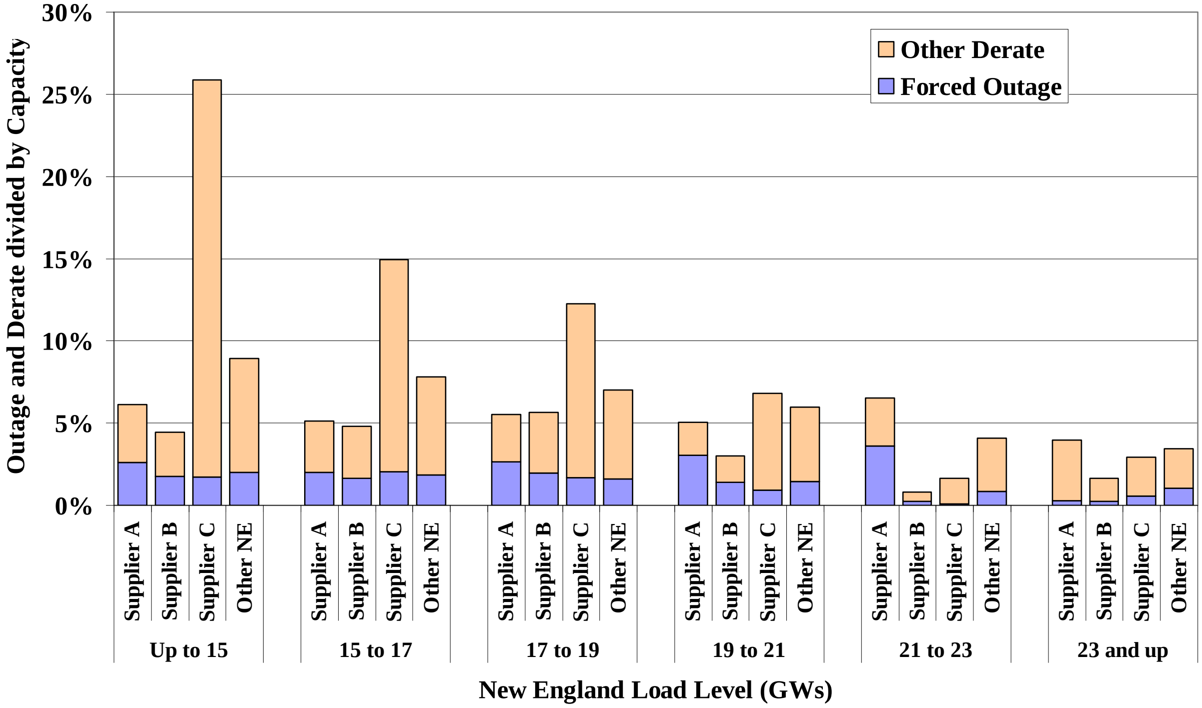


Figure 45 shows that the physical derating and forced outage quantities for the largest supplier and all other suppliers in Connecticut were moderate under all load conditions in 2008. Hence, these deratings and outages do not raise concerns about physical withholding in Connecticut.

4. Potential Physical Withholding in All New England

Having analyzed each of the major constrained areas in New England, Figure 46 summarizes the physical withholding analysis for all of New England by load level. The results of this analysis are shown for four groups of supply. Supplier A had the largest portfolio in New England and was pivotal in approximately 7 percent of the hours during 2008 (excluding capacity under reliability agreements). Suppliers B and C are the second and third largest suppliers in New England and were pivotal during less than 1 percent of the hours, respectively. All other suppliers are shown as a group for comparison purposes.

Figure 46: Forced Outages and Deratings by Load Level and Supplier
All New England, 2008



Supplier A exhibited rates of forced outages and other non-planned deratings that were moderate under all load conditions. Supplier B had relatively low levels of forced outages and other non-planned deratings, particularly under higher load conditions. Supplier C exhibited rates of forced outages and other non-planned deratings that were comparable to other New England suppliers when loads exceeded 19 GW, but were substantially higher at lower load levels, especially when load was less than 15 GW. The pattern for Supplier C, which is also the largest supplier in Boston, is explained by factors that do not raise competitive concerns.

As a group, the other New England suppliers showed higher derating levels under low load conditions, although their derating levels decreased as load levels increased. These patterns generally suggest that New England suppliers increased the availability of their resources under peak demand conditions. The increased availability is particularly notable when we consider the effects of high ambient temperatures on thermal generators. Naturally, ambient temperature restrictions on thermal units vary along with load and are difficult to distinguish from physical withholding through a review of market data. It is beyond the scope of this report to determine whether individual outages and other deratings were warranted. However, the overall quantity of capacity subject to the deratings was consistent with expectations for a workably competitive market, so we do not find evidence to suggest that these deratings constituted an exercise of market power.

E. Conduct Raising NCPC Payments

In addition to withholding capacity from the market to raise clearing prices, suppliers can also exercise market power by inflating the NCPC payments they receive when they are needed for local reliability. In this regard, the report finds that the largest supplier in the Boston area engaged in conduct designed to increase its NCPC payments in the early months of 2008. Specifically, the supplier:

- Increased its day-ahead offer prices for its large, economic units, which:
  - Reduced the day-ahead commitment of the supplier's large units; and
  - Required the ISO to supplementally commit some of the supplier's other capacity;
- Self-committed the large economic units on most days in which they were not committed by the ISO.

This conduct led to substantial excess capacity in the Boston area and rendered the supplemental commitments by the ISO unnecessary in retrospect. This conduct is discussed further in Section VIII.B.4. One by-product of this conduct was to increase the amount of output gap in Boston in 2008, which is evaluated in Section C.2. However, the evaluation concludes that the output gap did not substantially affect LMPs in 2008.

This pattern of conduct changed in early April 2008, when the ISO revised the local reliability requirements for Boston, reducing the need to commit generation for voltage support in the area. As a result, the supplier started to offer more of its generation closer to marginal cost in the day-ahead market. The revisions to the local reliability requirements were made possible by the addition of new transmission capability into Boston in 2007.

Although the incentive for this pattern of conduct has been removed, we recently coordinated with the Internal Market Monitoring Unit on an evaluation of the criteria used to mitigate offers that increase NCPC payments. We agree with the Internal MMU's conclusion that the mitigation criteria for conduct that affects NCPC payments in constrained areas should be modified, particularly given the introduction of locational forward reserve markets and forward capacity markets that provide substantial amounts of revenue toward fixed cost recovery. Furthermore, we support the recent proposed Tariff modifications filed by the Internal MMU to enhance the mitigation measures when generation is committed for local reliability.

F. Conclusions

Based on the analyses of potential economic and physical withholding in this section, we find that the markets performed competitively with little evidence of market power abuses or manipulation in 2008. The pivotal supplier analysis suggests that market power concerns exist in several of areas in New England. However, the abuse of this market power is limited by the ISO-NE's market power mitigation measures and the large amount of capacity under reliability agreements. Nonetheless, ISO-NE should continue to monitor closely for potential economic and physical withholding, particularly in constrained areas.

While there is no substantial evidence that suppliers withheld capacity from the market in order to raise clearing prices, there is evidence that at least one supplier engaged in conduct to inflate its NCPC payments when they were needed for local reliability. However, the revision in local reliability requirements for the Boston area largely alleviated this problem. To more effectively address this type of conduct in the future, the Internal Market Monitoring Unit recently filed proposed modifications to the mitigation measures, which we support.

X. Forward Capacity Market

ISO New England has had an installed capacity market since it began operations in 1998. However, the original capacity market design lacked several features now recognized as important to the success of capacity markets. In particular, the original capacity market did not reflect the locational value of capacity resources, nor did it provide stable capacity price signals that potential investors could use to accurately predict investment returns for new resources. The Forward Capacity Market ("FCM"), which was filed and approved in 2006, was designed to address these deficiencies by establishing a new market mechanism to attract and maintain sufficient resources to satisfy the New England's long-term resource planning requirements efficiently.

This section of the report provides an overview of the FCM design and evaluates the outcomes of the first two auctions, which were conducted in February and December 2008. The final part of this section provides a summary of our conclusions and recommendations.

A. Forward Capacity Market Design

Capacity markets are generally designed to provide incentives for efficient investment in new resources. A prospective investor estimates the cost of investment over the life of the project minus the expected variable profits from providing energy and ancillary services (after netting the associated variable costs). This difference between investment costs and variable profits, which is known as *Net CONE*, is the estimated capacity revenues that would be necessary for the investment to be profitable.

In an efficient market, the investment projects that have the lowest Net CONE occur. The capacity price should clear at a level that is higher than the Net CONE of the most economic investments that are needed and lower than the Net CONE of investments that are not needed. In this manner, the market facilitates investment in efficient capacity resources (i.e., the projects that have the lowest Net CONE) to meet system planning requirements. The resulting clearing price provides a signal to the market of the value of capacity.

FCM was designed to satisfy the resource adequacy requirements in New England efficiently by using competitive price signals to retain existing resources and attract new supply. FCM has several elements that are intended to work together to accomplish these goals. Some of the key elements are:

- *Installed Capacity Requirement* – The FCM procures the Net Installed Capacity Requirement ("NICR")[90] of the New England Control Area and the capacity judged necessary to achieve regional reliability standards.
- *Local Sourcing Requirement* – Before each auction, the existing[91] installed capacity in each zone, less retirement and export bids, are compared to the zone's Local Sourcing Requirement ("LSR").[92] If the amount of capacity is greater than the LSR, the zone will not be modeled as a separate import-constrained zone in the auction. Export-constrained zones are always modeled in the auction.
- *Forward Procurement* – The FCM procures the forecasted capacity requirement three years in advance of the Capacity Commitment Period.[93] The three-year lead time[94] is intended to encourage participation by new resources and allow the market to adapt to resources seeking to leave the market.
- *Locational Pricing* – The FCM produces locational price signals, which reflect the locational value of capacity given the LSRs. This provides incentives for investment in zones where new resources are needed most.
- *Descending Clock Auction* – The first stage of a Forward Capacity Auction (primary "FCA"), which is held three years prior to the Capacity Commitment Period, is structured

90 The NICR is equal to the Installed Capacity Requirement minus the HQICC. This treats a portion of the capacity from Hydro Quebec as a load reduction rather than as supply.

91 This includes capacity that was sold in previous FCAs but that is not yet in operation.

92 The LSR is the minimum amount of capacity that is needed in the load zone to reduce the probability per year of firm load shedding below 10 percent.

93 The Capacity Commitment Period is defined as a one-year period from June 1 to May 31 of the following year, starting with the period of 2010/2011.

94 The first two auctions were held in February and December 2008 for the commitment periods of 2010/2011 and 2011/2012, in order to limit the length of the transition period. Subsequent auctions will eventually take place approximately 3½ years before the commitment period begins.

as a descending clock auction.[95] The auction produces uniform capacity clearing prices that are used as the basis for payments to capacity suppliers.

- *Reconfiguration Auction* – As a supplement to the primary FCA, Reconfiguration Auctions ("RAs") will be conducted to allow the ISO and qualified resources to adjust their holdings. The ISO shall conduct two types of RAs: (1) annual RAs prior to the relevant commitment period; and (2) monthly RAs prior to each month of the commitment period.
- *New Capacity Treatment* – Existing capacity participates in the FCM each year and has only a one-year commitment, while new capacity resources can choose an extended commitment period from one to five years at the time of qualification. Both new and existing capacities are paid the same market clearing price in the first year, provided there is sufficient competition and sufficient supply. The price paid to new capacity after the first year is indexed for inflation.
- *Resource Qualification Process* – Qualification occurs during the 13 months prior to the auction, with specific requirements set for each resource type. The main purpose of qualification is to ensure that participating resources are credibly available when needed.

The FCM design also includes several provisions to guard against the abuse of market power. Demand resources and intermittent generation resources compete with traditional generation to provide capacity, limiting supply-side market power in the capacity and energy market and enhancing economic efficiency. Certain delist and export bids are subject to review by the market monitor prior to the FCA to address economic withholding. New capacity qualification rules and the three-year advance procurement feature allow new capacity projects to compete in the FCA.

B. Analysis of Forward Capacity Auction Results

Two FCAs have been held to date: the first in February 2008 for the commitment period of 2010/2011 ("FCA1"), and the second in December 2008 for the commitment period of

95 The descending clock auction is a multiple-round dynamic auction procedure. In each round, the auction manager announces prices, one for each of the zonal products being procured, then the bidders indicate the quantities of each product they wish to supply at the current prices. In the following round, prices for products with excess supply decrease, and the bidders make new offers with equal or lower quantities (compared to the prior offer) at the new prices. This iterative procedure stops when supply no longer exceeds demand in each zone.

2011/2012 ("FCA2"). In each auction, there was a substantial surplus of capacity over the NICR. Accordingly, each auction cleared at the floor price: $4.50 per kW-month in FCA1 and $3.60 per kW-month in FCA2.

No import-constrained zones were deemed necessary because the amount of existing capacity exceeded the LSR in each area. Maine was modeled as an export-constrained zone in both auctions, but there was no price separation between Maine and the rest of New England. This section evaluates the following three aspects of the outcomes from the first two FCAs:

- Summary of Capacity Procurement – This part of the section summarizes the procurement of capacity resources by type and by location compared to the capacity requirements.
- Delist Capacity – Existing capacity resources can leave the market by delisting. This part of the section evaluates the delist bids of existing resources.
- New Capacity Procurement – This part of the section examines the recent investments in new capacity.

1. Summary of Capacity Procurement

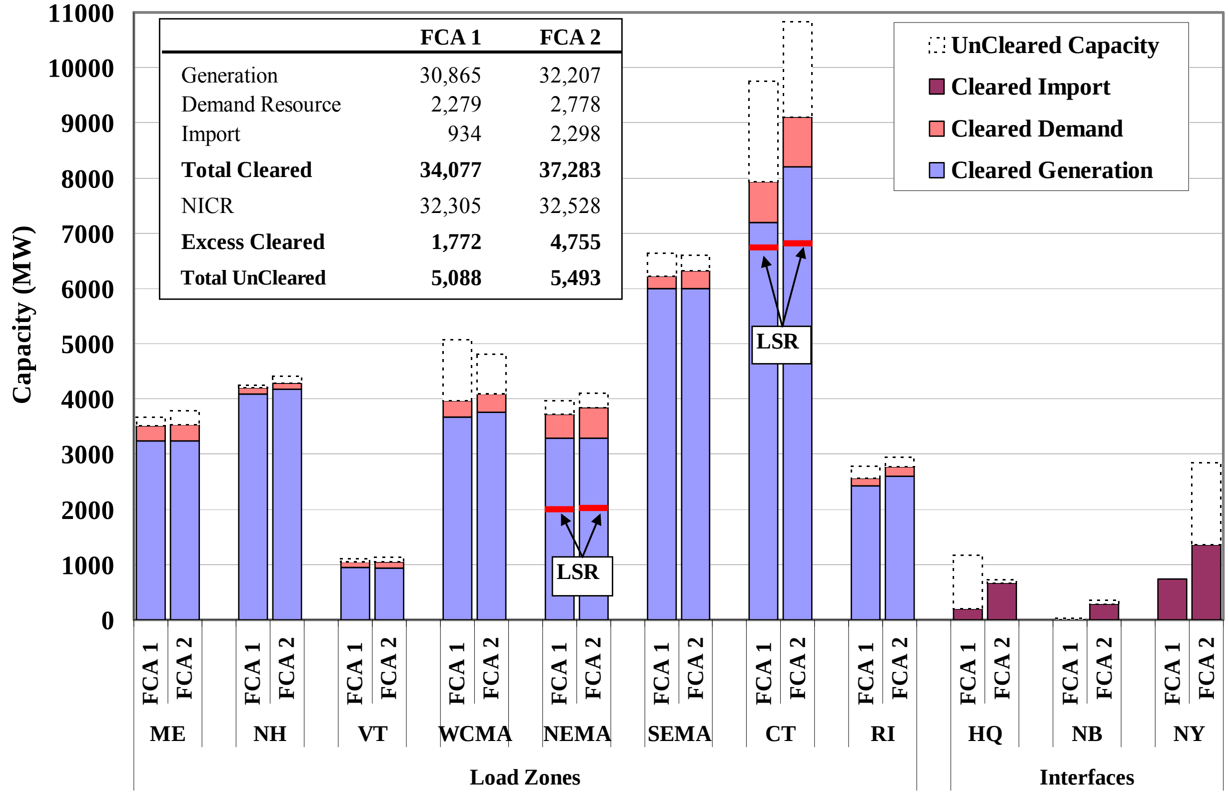
Figure 47 summarizes the procurements in the first two FCAs, showing the distribution of cleared and uncleared capacity by location. Cleared resources are divided into generating resources, demand response resources, and imports from external areas.[96] The amounts of cleared resources are shown relative to the LSRs for Connecticut and NEMA and the NICR for all of New England.

In each auction, the amount of capacity procured was more than sufficient to satisfy the reliability requirements. In FCA1, over 34 GW of resources were procured, exceeding the NICR by nearly 1,800 MW. In FCA2, over 37 GW of resources were procured, exceeding the NICR by approximately 4,800 MW.

---

96 The amount of cleared demand response resources shown in the figure has been adjusted to exclude Real-Time Emergency Generation resources in excess of 600 MW.

Figure 47: FCM Auction Clearing Summary by Location
FCAs 1 & 2



Prior to the auctions, it was determined that the existing capacity was sufficient to satisfy the local requirements, so no import-constrained zones were modeled. Accordingly, the amount of procured capacity exceeded the LSRs in Connecticut and Boston by approximately:

- 1,200 MW in FCA1 and 2,300 MW in FCA2 for Connecticut; and
- 1,700 MW in FCA1 and 1,800 MW in FCA2 for NEMA.

In each auction, there was a substantial amount of qualified resources that did not clear. New proposed resources accounted for more than 80 percent of the uncleared capacity, while the uncleared capacity from existing resources is evaluated in Part 2 of this section.

Generating resources provided the vast majority of capacity in each auction, satisfying 96 percent of the NICR in FCA1 and 99 percent in FCA2. In the two historically import-constrained areas (Connecticut and NEMA), the amount of procured generation resources was sufficient to satisfy the LSR. Demand response resources satisfied approximately 7 percent of the NICR in FCA1 and 9 percent of the NICR in FCA2. Approximately 70 percent of the cleared demand response resources were *active* demand resources, which reduce load in response to real-time system conditions or ISO instructions. The rest were *passive* resources that also reduce load, but not in response to real-time conditions or instructions (e.g., energy efficiency). Imports from Hydro Quebec, New Brunswick, and NYISO also accounted for a significant portion of the procured capacity, increasing from 934 MW in FCA1 to 2,298 MW in FCA2.[97]

Substantial excess capacity cleared in the first two auctions as a result of the price floor that was stipulated in the Settlement Agreement. The price floor will be used for the last time in FCA3. Hence, the surplus capacity purchases will be eliminated in FCA4 and future auctions.

2. Delist Capacity

FCM provides a mechanism to retain existing resources in New England. Stable price signals encourage existing resources to stay in-service, reducing the need to satisfy reliability requirements using out-of-market payments (e.g., payments from reliability agreements). Relying on out-of-market payments is undesirable because doing so provides the most compensation to the least efficient resources in the market. Hence, the use of out-of-market payments tends to reduce the efficiency of investment in the wholesale market.

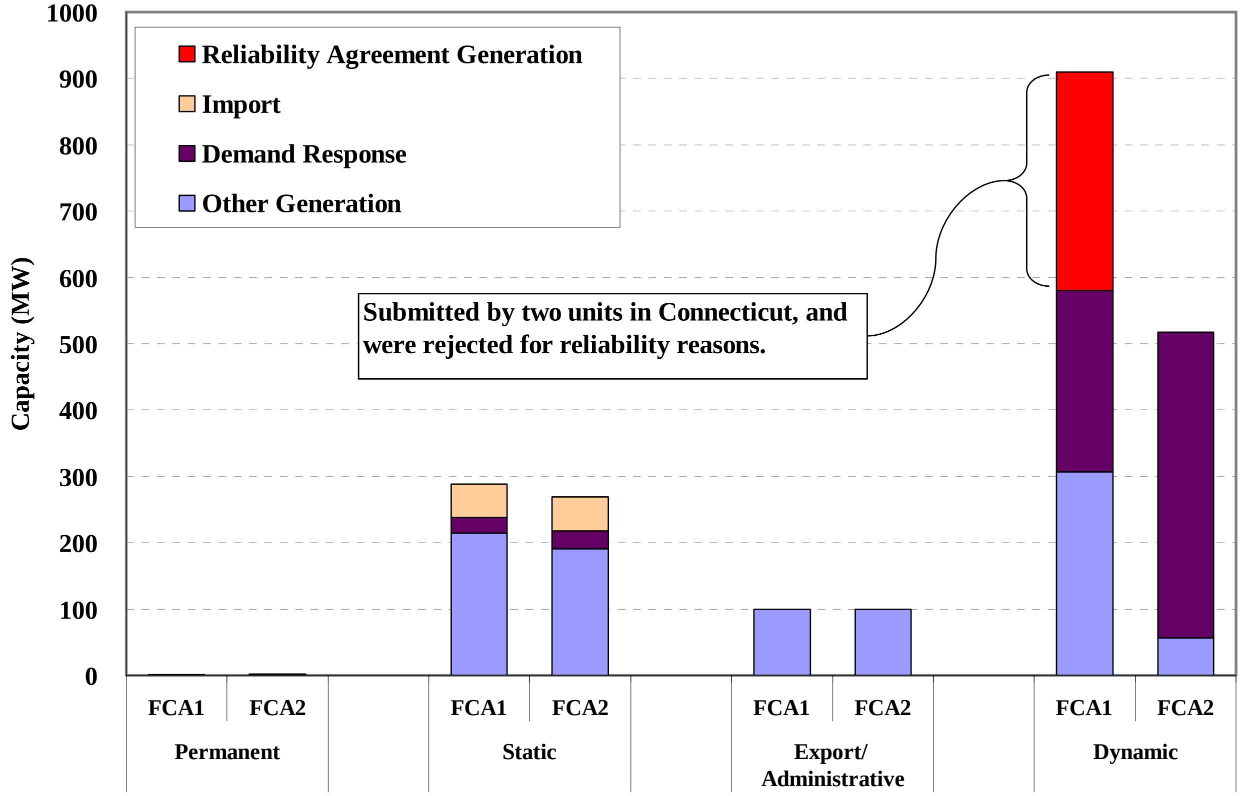
Under FCM, existing resources have the option to submit delist bids to indicate they intend to delist (i.e., make unavailable) all or part of their capacity during the commitment period if the capacity price is below the level specified by their delist bid. The ISO reviews delist bids and may reject them for reliability needs or in accordance with the mitigation rules. The following figure evaluates several categories of delist bids in the first two FCAs. The figure shows four categories of delist bids: permanent (retirement), static, export or administrative, and dynamic.[98] Delist bids are also separated according to the type of resource: generation not currently under a

97 A large portion of the import capability from Hydro Quebec is included in the HQICC, which is treated as a load reduction in the NICR rather than as supply.

98 Each category of delist bid is defined in *Market Rule 1, Section 13.2.5.2*.

reliability agreement, generation currently under a reliability agreement, demand response resources, and imports.

Figure 48: Summary of Delist Bidding by Type FCAs 1 & 2



Approximately 970 MW and 890 MW of existing resources delisted in the first two auctions. Dynamic delist bids accounted for the majority of accepted delist bids in each auction. Dynamic delist bids are the only type of delist bid that can be submitted during an auction. Other types of delist bids must be submitted prior to the auction. Delist bids that are less than 80 percent of CONE[99] are not subject to mitigation, while bids above 80 percent must be approved by the market monitor as consistent with the net going forward costs of the resource.[100]

99 CONE is an acronym for *Cost of New Entry*, although it has a very specific meaning in the context of FCM, which is defined in *Market Rule 1, Section 13.2.4*.

100 This is the estimated cost of keeping a resource in service minus any estimated energy and ancillary services revenues. The method of estimating the net going forward cost is defined in *Market Rule 1, Section 13.1.2.3.2.1.2*.

Generation accounted for a large share of the delisted capacity, approximately 64 percent in FCA1 and 39 percent in FCA2. This includes approximately 100 MW that delisted in each auction in order to export capacity to New York across the Cross-Sound Cable. Otherwise, most of the delisted generation capacity was associated with small output ranges on individual units.

All delist bids were accepted except for two bids submitted in FCA1 that were associated with 330 MW of generating resources in Connecticut. These two generators were under a reliability agreement with the ISO during 2008, and their delist bids were rejected during the auction when the ISOs determined in the Transmission Security Analysis that both generators were needed for Connecticut reliability. Since the rejected delist bids were substantially smaller than the excess cleared capacity for all of New England, the auction would have cleared at the price floor with or without the rejected bids. Hence, the decision to reject did not affect the auction clearing price in FCA1. However, the delist bids were rejected for Connecticut reliability even though the Connecticut LSR was satisfied by nearly 1,200 MW in the auction, indicating that the LSR did not adequately reflect the capacity needs of Connecticut.

Other than the two generators in Connecticut whose delist bids were rejected, none of the generation capacity that is currently under reliability agreements attempted to delist in either auction. Nineteen units (approximately 3,200 MW in total) were under reliability agreements during 2008. Therefore, approximately 90 percent of the capacity under reliability agreements was retained in the first two capacity commitment periods without the use of out-of-market payments. The two units whose delist bids were rejected will receive out-of-market payments in the first commitment period to make up the difference between the auction clearing price and their delist bid prices.

A limited amount of capacity delisted from the first two commitment periods, and the delisted capacity did not affect market clearing results due to the amount of excess capacity that was cleared in both auctions. Most of the existing resources were retained through the market, and only 330 MW of capacity required out-of-market payments to be retained in service for the first commitment period. Accordingly, the amount of out-of-market payments from the ISO to existing suppliers will decline dramatically from approximately $129 million in 2008 to approximately $7 million in the first Capacity Commitment Period. Additionally, a 70 MW generation resource (that was existing prior to the FCM) in Connecticut is also receiving out-of-market payments under an RFP of the Connecticut PUC.[101] Although the use of out-of-market payments to retain existing resources has been reduced, their use still raises efficiency concerns because the resources receiving out-of-market payments receive more compensation than competing resources in the same local area that provide comparable reliability benefits. The need for out-of-market payments implies that the clearing price in the local area is understated, producing inefficient incentives for new investment.

Finally, the rejection of the delist bids of 330 MW of capacity in Connecticut in FCA1 highlights an issue with the method of determining the LSRs. The Connecticut LSR was much lower than the capacity requirement that was implied by the Transmission Security Analysis that was the basis for rejecting the delist bids. As a result, the delist bids were rejected to protect Connecticut area reliability even though an excess of 1,200 MW was procured in Connecticut relative to the LSR. In principle, markets would always be designed to satisfy the reliability needs of the system, which allows the market prices to accurately reflect these needs. The Internal MMU has also indentified this issue and recommended that the LSR criteria be made consistent with the criteria that is used in the Transmission Security Analysis to determine whether a delist bid should be rejected for zone-level reliability.[102] We support this recommendation.

3. New Capacity Procurement

A key objective of the FCM is to provide efficient market incentives for investment in new resources. The FCA provides a mechanism for prospective investors to build new resources when it will be profitable based on the auction clearing price. As a result of competition between prospective investors, the investment projects that have the lowest Net CONE should clear in the auction and result in the most efficient investment over time.

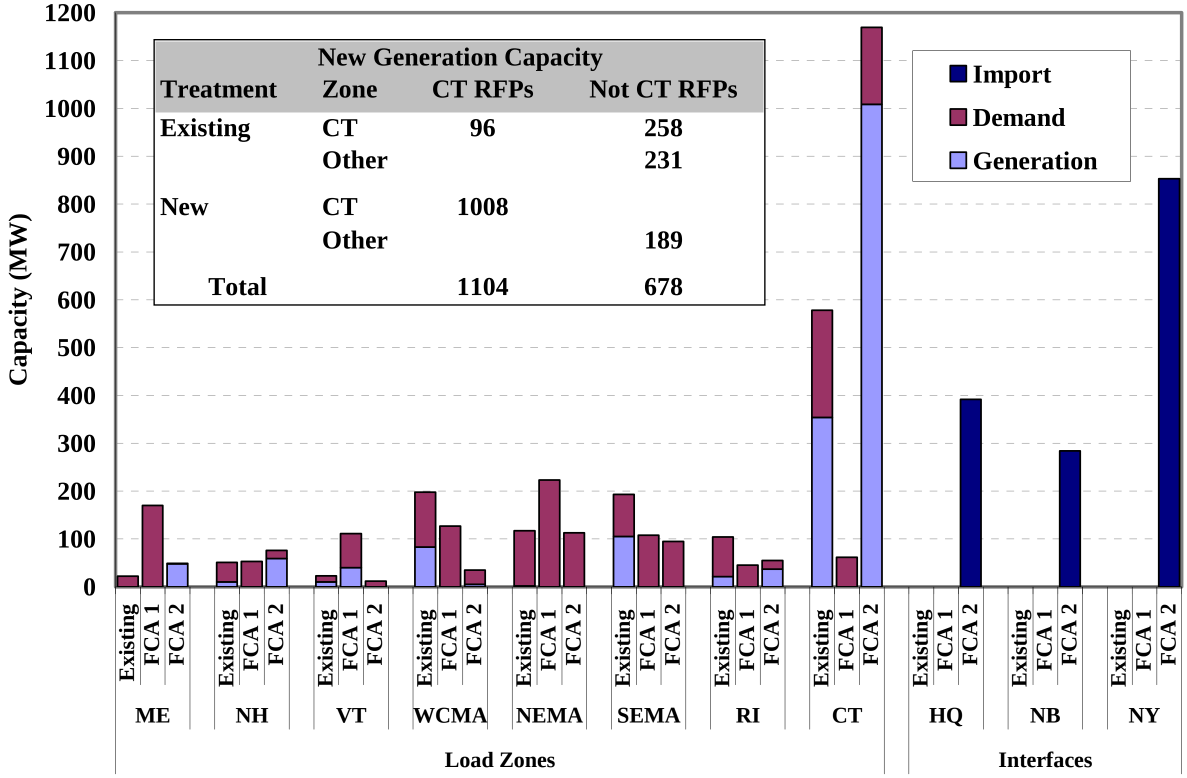
---

101 See *State of Connecticut DPUC Investigation of Measures to Reduce Federally Mandated Congestion Charges (Long-Term Measures)*, May 3, 2007, Docket No. 05-07-14PH02, page 2.

102 See *Review of the Forward Capacity Market Auction Results and Design Elements* by Dave LaPlante, Hung-Po Chao, Sam Newell, Metin Celebi, and Attila Hajos.

Figure 49 shows the amounts of new capacity that were procured in the first two FCAs by load zone or external interface. Capacity is divided by resource type: generation, demand response, and import capacity. Capacity is also shown according to whether it received existing treatment in FCA1, cleared in FCA1, or cleared in FCA2.[103]

Figure 49: New Capacity Procurement by Location
FCAs 1 & 2



To determine whether new capacity entered due to the FCM revenue, the table in the figure identifies the quantity of capacity contracted under Connecticut PUC Request for Proposals (RFP) that may receive additional capacity payments beyond those from the FCM.[104]

103 The owners of resources that were expected to be in-service prior to the first Capacity Commitment Period could elect to be treated as existing resources in FCA1. Accordingly, they are able to submit delist bids rather than supply offers.

104 See *State of Connecticut DPUC Investigation of Measures to Reduce Federally Mandated Congestion Charges (Long-Term Measures)*, May 3, 2007, Docket No. 05-07-14PH02, page 2. See also *State of Connecticut, DPUC Review of Peaking Generation Projects*, June 25, 2008, Docket No. 08-01-01, page 64.

In the first two FCAs, more than 4 GW of new capacity was procured from generation, demand response resources, and imports.[105] The following discussion reviews and evaluates the procurements of new capacity by resource type that are shown in Figure 49.

Import Capacity

The large quantity of new capacity sold by importers indicates that they expected that the revenues from providing capacity to New England during the Capacity Commitment Period to be greater than the revenues from providing capacity to another market during the same period. Many of the capacity importers to New England have the option to sell capacity into New York in future periods. Hence, the amount of capacity imports may decrease in FCA4 when the floor price is no longer used. Similarly, the amount of capacity that delists in order to export may increase in the future when the floor price is removed.

Demand Response Capacity

Demand response resources have sold substantial amounts of capacity under FCM, indicating that the Net CONE of many demand response resources is lower than the capacity clearing prices. However, if demand response activation becomes more frequent in the future, the Net CONE of many demand response resources should increase. This increase would happen if the heavier reliance on demand response results in much more frequent emergency load curtailments that are costly for demand response providers to satisfy. If this happens, it will put upward pressure on capacity clearing prices and/or reduce the amount of capacity provided by demand response resources.

The other issue raised by the large share of new capacity provided by demand response resources is whether the capacity obligations they receive are comparable to the obligations borne by other types of resources that clear in the FCM. One notable difference is the fact that demand response resources are not currently obligated to pay the Peak Energy Rents (PER) to the ISO. The fact that other types of new resources must bear different obligations can inefficiently bias the

105 This excludes new resources treated as existing resources because they are already committed to enter.

investment incentives in favor of demand response resources. This issue is discussed more fully in the conclusion to this section.

Generation Capacity

A substantial amount of new generation capacity (1,782 MW) has entered the market under FCM. Entry of generation resources would generally not be expected when the price clears at the price floor as it did in FCA1 and FCA2. The floor price is generally believed to be substantially lower than the Net CONE for new investment in most types of generation. However, the table in the figure above shows that 1104 MW of the new investment in generation received additional payments under the RFPs of the Connecticut PUC. Almost another 500 MW are resources that received existing treatment, which indicates that their entry decisions were not contingent on the outcome of an FCA. We distinguish these two types of new investment because the FCA did not directly facilitate the entry, although the existence of the FCM may have motivated the processes that resulted in the entry.

Alternatively, entry that occurs only because its offer is accepted in the FCA (not because the supplier was awarded a contract under a state RFP or was already building the unit) is entry that will ultimately allow the FCM market structure to efficiently govern investment over the long-run. For this reason, we seek to determine how the FCM market has affected this class of capacity investment. The table in the figure shows that only 189 MW of new generating resources cleared in the FCAs that were not under the CT RFP or treated as existing resources. Most of these resources are facilities powered by renewable fuels, designed to up-rate existing resources, or made to re-power existing power plants. Such projects may have a lower Net CONE than most of the potential investments in new generation, which explains why they would clear at the floor prices in the first two FCAs. In fact, given the prevailing surplus in New England, it would have been surprising if a substantial amount of new generating resources had cleared in the FCM.

Based on the new capacity that has been or is planned to be placed in service, the amount of capacity committed to New England in the second Capacity Commitment Period exceeds the New England capacity requirement by 15 percent. FCM has provided strong incentives for the sale of new capacity by demand response resources and importers. However, once the price floor is no longer used (starting in FCA4), the amount of excess capacity purchased in the auction will be eliminated, which will likely reduce the amount of new capacity that is purchased in the FCAs over the next few years.

It is too early to determine whether the FCM will efficiently facilitate investment in new generation when it is needed. Because the price has cleared at the floor price, which is well below most estimates of the Net CONE for new generation, the market has not yet been needed to facilitate investment in new generation resources.

C. Conclusions and Recommendations

The Forward Capacity Market was successfully introduced by ISO New England with no significant procedural problems. The qualification processes and the auctions have occurred on schedule, which is noteworthy for a major wholesale market design initiative. Furthermore, the results of the auctions have been competitive, and sufficient capacity is planned to be in-service to satisfy the needs of New England through May 2012. The use of out-of-market payments by the ISO to retain existing resources has been virtually eliminated. This has significantly improved incentives to capacity suppliers compared with the current reliance on reliability agreements to retain existing capacity.

However, most of the new investment in generation under FCM has been motivated by out-of-market payments related to RFPs of the Connecticut PUC. It is unlikely that substantial amounts of additional generation investment will occur until capacity clearing prices rise significantly. Therefore, it will be difficult to determine whether the FCM facilitates efficient investment in new generation until the current capacity surplus is diminished.

However, the large quantities of demand response resources that have entered at prices well below the Net CONE of new generation is a notable outcome of the first two auctions. This raises potential efficiency concerns to the extent that capacity obligations of different resources vary. One notable difference is that the PER provisions do not apply currently to demand resources. These provisions are essentially a financial call option on the Peak Energy Rents, the value of which should be embedded in the capacity clearing price. The fact that demand response resources would receive the value of this option in the capacity clearing price without having the PER obligations that apply to generating resources distort investment incentives in favor of demand response resources. The Internal MMU identified this issue in its recent FCM report as well. It recommended that the PER provisions be extended to demand response resources along with efficient energy payments when the demand response resources curtail their load, which we support.[106]

The rejection of the delist bids of 330 MW of capacity in Connecticut in FCA1 highlights an issue with the method of determining the LSRs. The Connecticut LSR was much lower than the capacity requirement that was implied by the Transmission Security Analysis that was the basis for rejecting the delist bids. As a result, the delist bids were rejected to protect Connecticut area reliability even though an excess of 1,200 MW was procured in Connecticut relative to the LSR. The Internal MMU has also indentified this issue and recommended that the LSR criteria be made consistent with the criteria that is used in the Transmission Security Analysis to determine whether a delist bid should be rejected for zone-level reliability.We support this recommendation.

106 See *Review of the Forward Capacity Market Auction Results and Design Elements* by Dave LaPlante, Hung-Po Chao, Sam Newell, Metin Celebi, and Attila Hajos.