2009 Assessment of the Electricity Markets in New England

By:

David B. Patton, Ph.D.
Pallas LeeVanSchaick, Ph.D.
Jie Chen, Ph.D.
POTOMAC ECONOMICS, LTD.

External Market Monitor
ISO New England Inc.

June 2010

Table of Contents

List of Figures

List of Figures (cont.)

List of Tables

Guide to Acronyms

APR	Alternative Price Rule
ASM	Ancillary Services Market
CONE	Cost of New Entry
CT DPUC	Connecticut Department of Public Utility Control
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
LSR	Local Sourcing Requirement
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.
RA	Reconfiguration Auction
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 External Market Monitor 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.[1] In this Annual Assessment, we provide our annual evaluation of the ISO's markets for 2009 and our recommendations for future improvements. This report complements the State of the Market Report produced by the Internal Market Monitor, which provides its evaluation of the market outcomes in 2009.

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.

1 The duties of the External Market Monitor are listed in Appendix A.2.2 of "Market Rule 1."

I. Executive Summary

This report assesses the efficiency and competitiveness of New England's wholesale electricity markets in 2009. 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 wholesale price signals that govern investment and retirement decisions in the long-term. The markets achieve the short-term benefits by coordinating the commitment and dispatch of the region's resources, which 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 level of reliability with which it is delivered.

A. Introduction and Summary of Findings

In addition to providing a summary of market outcomes in 2009, 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 2009. Our analysis raised no competitive concerns that suppliers withheld resources to raise prices in the New England markets. Energy prices fell by almost 50 percent from 2008 to 2009, due primarily to the sharp decrease in natural gas prices. Natural gas prices decreased 52 percent in 2009 from the prior year.[2] In a competitive market, suppliers will face strong incentives to offer their supply at prices close to their short-run marginal costs of production.[3,4] Because fuel costs constitute the vast majority of the marginal costs of most generation, lower fuel costs translate to lower offer prices and market clearing prices in a well-functioning, competitive market. The continuing close correspondence of electricity prices and natural gas prices in New England demonstrates of the competitiveness of ISO New England's markets.

In previous years, we found that frequent supplemental commitment encouraged some generators to raise their offers above competitive levels (i.e., above marginal cost).[5] This was because generators committed for local reliability often do not face meaningful competition and may have local market power.[6] However, supplemental commitment for reliability became much less

2 Natural gas prices are based on the day-ahead prices reported by Platts for the Algonquin pipeline for the City Gate Rate.

3 Short-run marginal costs are the incremental costs of producing additional output, including any foregone opportunity costs of producing such output, in a timeframe short enough to preclude expanding, retiring or converting the assets to another use. For convenience, we will refer to these costs as "marginal costs".

4 The incentive to submit offers at prices close to marginal cost is affected by the design of the market. This incentive exists in markets that establish clearing prices paid to all sellers, as is the case in the wholesale electricity markets run by ISO New England. Markets that make payments to individual suppliers that are based on the supplier's offer (i.e., pay-as-offer markets) create incentives for suppliers to raise their offers above their marginal costs.

5 Potomac Economics, Ltd., *2008 Assessment of the Electricity Markets in New England*, June 2009, ("2008 Assessment").

6 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.

frequent in 2009 due to significant transmission upgrades in historically congested areas. As a result, the competitive issues that arise from supplemental commitment were virtually eliminated in 2009. To address competitive issues associated with reliability commitments in the future, the ISO filed changes to the mitigation rules that were accepted by the Federal Energy Regulatory Commission in October 2009.[7]

Operational Efficiency of the Markets

The day-ahead and real-time markets operated efficiently in 2009 as prices reflected underlying market fundamentals. Electricity prices in New England have been strongly correlated with changes in underlying fuel prices, as one would expect in a well-functioning market. Upgrades to the transmission system in Connecticut and Southeast Massachusetts were completed in 2009 that have had several notable effects on market operations. First, they have sharply reduced the need for the ISO to commit generation for local reliability. Such commitments lead to surplus capacity in real time, which depress energy and ancillary services prices in the real-time market. Accordingly, real-time prices were more consistent with the marginal cost of production in 2009.

Second, the transmission upgrades have been key in reducing New England's heavy reliance on reliability agreements, which have been used to ensure that units needed for reliability remained in operation. This is important because reliability agreements are poor substitutes for transparent market prices and do little to facilitate efficient investment. Taken together, the reduction in supplemental commitment and reliability agreements is primarily responsible for a $248 million reduction in uplift charges from 2008 to 2009.

Recommendations

Overall, we conclude that the markets performed competitively in 2009 and were operated well by the ISO. Based on the results of our assessment, however, we offer thirteen recommendations to further improve the performance of the New England markets. Nine of the thirteen were recommended in our 2008 Annual Assessment as well. This is not surprising since many of the

7 *Order Conditionally Accepting Market Rule 1 Revisions*, FERC Docket No. ER09-1546-000 (October 2, 2009).

recommendations require substantial resources and must, therefore, be prioritized with the ISO's other projects and initiatives. However, most of these nine recommendations are either currently being evaluated by the ISO or have been included in the Wholesale Markets Plan for implementation over the next five years. In addition, in the past year the ISO completed implementation of two high-priority recommendations from last year's report.

The following sections summarize our findings, and a table of recommendations can be found at the end of this executive summary.

B. Energy Prices and Congestion

Electricity prices fell 48 percent from 2008 to 2009 primarily due to a 52 percent decrease in natural gas prices.[8] A close correlation between natural gas prices and electricity prices is expected in 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 2009 as peak demand levels were considerably lower than in previous years and the system was operated effectively. The peak load was 25.1 GW in 2009, substantially lower than the summer peak load forecast of 27.9 GW and the all-time peak load of 28.1 GW that occurred in 2006.

Congestion and Financial Transmission Rights

In 2009, New England experienced very little congestion into historically-constrained areas, such as Boston, Connecticut, and Lower Southeast Massachusetts ("Lower SEMA"). In fact, most of the price separation between net exporting regions and net importing regions has been due to transmission losses, rather than transmission congestion. In 2009, Lower SEMA exhibited the most congestion of any area in New England with average congestion of $1 per MWh relative to the New England Hub in the day-ahead market.

8 The average electricity price is weighted by the New England load level in each hour.

Transmission upgrades completed in 2009 significantly reduced congestion into Lower SEMA and Connecticut. In the day-ahead market, the average congestion-related price difference between the New England Hub and Lower SEMA fell from $10.10 per MWh in 2008 to $0.96 per MWh in 2009. Likewise, the average congestion price difference between the New England Hub and Norwalk-Stamford fell from $4.79 per MWh in 2008 to $0.72 per MWh in 2009. In addition to the transmission upgrades, the sharp reduction in natural gas prices also contributed to lower congestion since redispatch costs are generally highly correlated with fuel prices.

These reductions in congestion-related Locational Marginal Price ("LMP") differences translate to associated reductions in overall congestion revenue collected in the day-ahead and real-time markets. Congestion revenue decreased from more than $120 million in 2008 to $25 million in 2009, a decrease of almost 80 percent. In general, these revenues are used to fund the FTRs sold by ISO New England.

The ISO operates annual and monthly markets for FTRs.[9] FTRs allow participants to hedge the congestion and associated basis risk between any two locations 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:

- Annual FTR prices generally over-estimated the congestion that prevailed in the energy market in 2009. This suggests that participants did not fully anticipate the effects of the transmission upgrades and lower fuel prices on congestion-related price differences.
- 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.

The congestion revenue of $25 million collected by the ISO in 2009 was not quite sufficient to pay the full target value of the FTRs. Hence, the holders of FTRs received an average of 97 percent of the target value of their FTRs.

9 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).

Day-Ahead to Real-Time Price Convergence

When prices in the day-ahead market converge well with the real-time market it indicates that the day-ahead market accurately represents expected real-time market conditions. This is important because most supply and demand settlements occur in the day-ahead market and FTRs settle against day-ahead prices. Additionally, most generation is committed in the day-ahead market so good price convergence leads to the most economic commitment of resources to serve load in real time.

We evaluated price convergence at the New England Hub, which is broadly representative of prices outside of transmission-constrained areas. We found that the differences between day-ahead and real-time prices were relatively small in 2009, indicating good overall convergence. However, we found that average real-time prices were six percent higher than average day-ahead prices in the last five months of 2009, which is unusual in both magnitude and direction. This pattern was likely attributable to the reduction in surplus online capacity (following the transmission upgrades), which had previously limited the frequency of high real-time prices.

In general, congested areas exhibit worse convergence than other areas in LMP markets because prices in these areas are more volatile and difficult to predict. However, because congestion decreased markedly in 2009, convergence improved significantly from 2008 to 2009 in New England's most congested areas (Lower SEMA and Connecticut).

C. External Interface Scheduling

Efficient scheduling of the interfaces between New England and its neighbors can have a significant effect on the market outcomes in New England. Hence, we evaluate 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 ranged from 1,670 MW during peak hours to 1,180 MW during off-peak hours in 2009, 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 370 MW during peak hours in 2009), while participants schedule power flows in both directions on the larger interface depending on the relative prices. In 2009, an average of 200 MW was exported to New York during peak hours and 102 MW was imported from New York during off-peak hours.

Market participants should arbitrage the prices in New York and New England 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. In order to utilize the interfaces more efficiently, ISO New England and the New York ISO have agreed to collaborate on a project to coordinate the physical interchange of power between the markets in real time. Such coordination is 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 from 2006 to 2009. The simulations indicated that consumers in New England would have saved $64 million in 2009. These savings are likely to be higher in the future because:

- Fuel prices were relatively low in 2009, which reduces the price effects of inefficient scheduling; and
- Surplus capacity in the real-time market has been falling as fewer supplemental commitments have been necessary for local reliability. Lower surplus capacity levels

lead to more frequent operating reserve shortages and associated price volatility, which would be reduced by fully utilizing the interfaces.

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 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

Reserve clearing prices were relatively low in the real-time market in 2009, with little variation by location. Clearing prices averaged approximately $0.70 per MWh for 10-minute spinning reserve (“TMSR”) and $0.50 per MWh for 10-minute non-spinning reserve (“TMNSR”). Reserve clearing prices were $0 in 98 percent of real-time intervals because sufficient surplus capacity was online to meet the reserve requirements without redispatching generation. However, as supplemental commitments for local reliability decreased sharply in the second half of 2009, surplus online capacity levels fell substantially. As one would expected, this led to higher operating reserve clearing prices due to the tighter supply conditions. Average clearing prices in the last quarter of 2009 increased to approximately $1.80 per MWh for TMSR and $1.50 per MWh for TMNSR. These prices remain well below the prices for comparable products in other markets.

We also found that although the ISO has local reserve zones, which initially were defined to include Boston, Southwest Connecticut, and Connecticut, reserve prices did not vary substantially by location. This is largely due to transmission upgrades that have significantly reduced local requirements in these areas. However, we found in prior years that actions taken by the ISO to maintain local reserves were often more costly than the local Reserve Constraint

Penalty Factor ("RCPF") of $50 per MW-hour, which indicates that the local RCPFs were 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 increased the local RCPFs to $250 per MW-hour on January 1, 2010. Our analysis indicates that this level would likely have been sufficient to maintain reserves in local areas under most conditions.

Finally, we evaluate the designation of the local reserve zones. There are five local areas that require the commitment and dispatch of resources to meet local second contingency protection requirements, but are not local reserve zones.[10] Because they are not designated as local reserve zones, the ISO relies entirely on the transmission interfaces to satisfy local reserve needs in these areas.[11] 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. Although this may change in the the future, it should be noted that the benefits from designating new reserve zones would be limited under current market conditions due to the infrequent need to rely on imported reserves in these areas.

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 then 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 TMNSR for all of

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

11 This is done by activating a "proxy second contingency" transmission limit. We refer to this space on the interface as "imported reserves".

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 held in 2009, 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 requirement. In Boston and the Rest of System areas, TMNSR cleared at $6.30 per kW-month in the Summer 2009 auction and at $6.08 per kW-month in the Winter 2009/10 auction, while TMOR cleared at $0 per kW-month in both auctions.[12] TMNSR and TMOR prices cleared at the same level as Rest of System in Boston because recent transmission upgrades allowed the local requirement to be satisfied with External Reserve Support.[13]

The fact that there is a single cap of $14 for all reserve 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. Our recommendation to modify the price cap to differentiate the payment for higher quality reserves or reserves in more constrained areas would address these incentive issues.

This report identifies some issues with the performance of the forward reserve market. The market has not resulted in lower Net Commitment Period Compensation ("NCPC") as some had

---

12 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 2009, the capacity revenue was from Transition Payments of $4.05 per kW-month. Thus, a seller of TMNSR outside the local areas in the Summer 2009 auction would receive $6.30 per kW-month, but also give up $4.05 per kW-month.

13 External Reserve Support is the portion of the local reserve requirement that is met by additional transmission capability, thereby reducing the amount of reserves procured locally.

expected because it has not succeeded in procuring high-cost units that are frequently committed for reliability. Additionally, the forward reserve market has produced price signals that are not consistent with the prevailing surpluses in the local areas, although this issue may be resolved as the forward reserve requirements adjust to reflect the new transmission investment. In the long-run, an evaluation of whether this market remains necessary may be warranted given the fact that the FCM includes very similar locational requirement.

E. Regulation Market

The regulation market performed competitively in 2008 with an average of approximately 700 MW of available supply competing to serve an average of 89 MW of regulation demand.[14] 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 fell from $51 million in 2008 to $23 million in 2009, largely due to the decrease in natural gas prices. Natural gas-fired combined cycle generators, which provide most of the regulation service, are usually committed more frequently during periods of low gas prices. This increases the availability of low-priced regulation offers and leads to lower regulation expenses.

Given the complex interaction of regulation and energy, co-optimizing the scheduling of resources for regulation, energy, and operating reserves would improve the economic efficiency of the markets.

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. However, these out-

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

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. 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 2009.

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. The significance of this issue has grown recently because the use of fast-start units increased in the second half of 2009 due to the reduction in supplemental commitment and the associated decline in surplus online capacity.
   - This issue is common to most RTO markets because fast-start resources have significant start-up and no-load offers that are difficult to incorporate in the price-setting logic of the real-time market.

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. It is important for markets to set efficient prices that reflect such conditions. This issue became less significant in 2009 as congestion became less frequent. 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 be consistent with 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 almost 2,300 MW in

January 2010. 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.
- ISO New England is working with stakeholders to facilitate participation by demand response in the energy and ancillary services markets. As participation by non-dispatchable demand response resources increases, it will be increasingly important to develop rules that allow them to set price when appropriate.

5. *Ex Ante and Ex Post Pricing:* 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 most 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 lead 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 growth in demand response resources and the increased use of fast-start units. 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 should provide 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 scheduling incentives in the day-ahead market. Hence, such actions should only be made when necessary. In this section, we evaluate several aspects of the ISO's operations and processes for satisfying reliability requirements in 2009.

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.

Commitment for Local Reliability

Overall, supplemental commitment for reliability decreased from a daily average of 1,000 MW in 2008 to 300 MW in 2009. In Connecticut and Lower SEMA, such commitment declined sharply in 2009, primarily due to the effects of recent transmission upgrades into both areas.

Decreased commitment for reliability has also led to a decline in the amount of daily surplus capacity (i.e., the amount of online reserves and fast-start reserves minus the real-time reserve requirement in the peak load hour) from an average of 1,300 MW in prior to the transmission upgrades in 2009 to 750 MW after the SEMA upgrades in mid-2009. This decline in surplus online capacity has affected the market in a number of ways that are discussed throughout the report.

Supplemental Commitment for System-Wide Reliability

Given the effect of surplus capacity on prices, it is important to evaluate the supplemental commitments made by the ISO and self-commitments made by market participants after the day-ahead market. Inefficient commitments in either area will tend to distort real-time prices, while inefficient supplemental commitments by the ISO will also lead to increased uplift costs.

As detailed above, transmission upgrades have substantially reduced the need for the ISO to commit generation to satisfy local reliability requirements. However, the ISO must still periodically make commitments to satisfy New England's system-wide reliability requirements. Our evaluation indicates that when the ISO has made supplemental commitments for system-wide capacity needs, they have generally been needed and resulted in only low levels of surplus capacity. However, there have been a number of days when large quantities of supplemental commitments have resulted in large quantities of surplus capacity.

It is important to recognize that uncertainties tend to have a bigger effect in New England than in other markets due to the limited quantity of fast-start generating resources in New England. This causes the ISO in some cases to have to rely on slower-starting units that must be notified well in advance of the operating hour when uncertainty regarding load, imports, and generator availability is high. Nonetheless, we have identified some areas for the ISO to evaluate regarding the assumptions that are made in the reliability assessment process used to determine when supplemental commitments are necessary.

Out of Merit Generation

The decline in surplus capacity has led to a reduction in the amount of out-of-merit generation that operates in the real-time market. Out-of-merit generation is energy produced from units with energy offers that exceed the LMP at their locations. In general, out-of-merit generation is undesirable because it distorts real-time prices and indicates a lack of correspondence between the market requirements and the system's operating requirements.

During peak hours, the average amount of out-of-merit generation from units committed for reliability fell from 208 MW in 2008 to 60 MW in 2009. Likewise, the average amount of out-of-merit generation from economically committed units running at their minimum output level fell from 232 MW in 2008 to 155 MW in 2009. The reduction in the amounts of surplus capacity and out-of-merit generation are positive because they indicates that the demand for energy and reserves are being satisfied more efficiently and producing more efficient price signals.

Uplift Charges

The transmission upgrades in Connecticut and Southeast Massachusetts have reduced the need for supplemental commitment for local reliability, as well as the use of reliability agreements to retain capacity that supports local reliability. Accordingly, the associated uplift charges decreased sharply in 2009, from $387 million in 2008 to $139 million in 2009.

Conclusions

Our assessment of system operation indicates that the ISO has been very effective, and we found no major concerns. Additionally, the cost of satisfying the system's local reliability and system-wide reliability requirements decreased substantially in 2009.

However, we recommend three changes listed in the table of recommendations below. These changes, together with the pricing improvements we recommend, should further improve the performance of the real-time markets and improve the economic signals that they produce.

H. Competitive Assessment

The report evaluates the market concentration and competitive performance of the markets operated by ISO New England in 2009. The most substantial market power exists in constrained areas that can become separate geographic markets with a limited number of suppliers when congestion arises. However, transmission upgrades completed in 2009 increased import capability to Connecticut and Southeast Massachusetts, which has reduced potential market power in those areas.

This assessment identifies structural market power issues and evaluates the conduct of market participants in several 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:

- Following the transmission upgrades, the largest suppliers in three of the seven areas were pivotal in a large number of hours.
- When we account for the large amounts of nuclear capacity and reliability agreements, we find a pivotal supplier in more than 30 percent of the hours in Boston and All of New England.
- A supplier was pivotal in 22 percent of the hours in Lower SEMA in 2009, but the supplier was not pivotal in any hours after transmission upgrades were completed in July.
- Market power will be a more significant concern in Connecticut once the reliability agreements expire. Hence, it will be important to continue to monitor the area 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. Due to the decline in commitment for local reliability, this was not a significant concern in 2009. However, this was a significant issue in previous years. To better address this issue in the future, we worked with the Internal Market Monitoring Unit to develop revisions to the mitigation rules. These revisions were filed by the ISO and were approved by FERC in October 2009.

I. Forward Capacity Market

The Forward Capacity Market was introduced to provide efficient economic signals that augment those provided by the energy and ancillary services markets in order to govern long-term investment and retirement decisions. The FCM consists of annual Forward Capacity Auctions ("FCA") held three years in advance of the commitment period when the capacity must be delivered. The first FCA was held in February 2008 to procure capacity for the 2010/2011 commitment period.

The FCA results to date have been competitive, and sufficient capacity is planned to be in-service to satisfy the needs of New England through May 2013. The ISO's use of out-of-market payments 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 three FCAs, more than 4 GW of new capacity was procured from generation, demand response resources, and imports. However, most of the new investment in generation under FCM has been motivated by out-of-market payments related to RFPs of the Connecticut Department of Public Utility Control ("DPUC"). 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 request for proposals ("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.

In the first two auctions, large quantities of demand response resources have entered at prices well below the net entry costs for new generation. 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 currently do not apply to demand response resources. This may inefficiently bias investment in favor of demand response resources. Hence, we recommend changes that would make the obligations accepted by demand response resources and generation resources more consistent.

The results of the first three forward capacity auctions highlight an issue with the Local Sourcing Requirements ("LSRs"), which are the capacity requirements modeled in the auction for individual zones. Several de-list bids (911 MW of capacity) were rejected because of reliability concerns in Connecticut and Boston even though the Connecticut and Boston LSRs were satisfied. To address this inconsistency between the auction and the reliability criteria, the ISO made a FERC filing to modify the criteria for determining the LSRs before the next auction.

In this report, we evaluate the recently proposed reforms of the Forward Capacity Market. In particular, we discuss three elements of the market design: the Alternative Price Rule, the modeling of capacity zones, and the market power mitigation rules. Based on our analysis, we recommend the following:

- The ISO should revise the Alternative Price Rule ("APR") so that the full effect of OOM entry is removed from the price-setting mechanism by establishing the clearing price at the level that would prevail if OOM entry did not occur;
- The ISO should revise the APR to account for out-of-market entry when new capacity is not needed, and when the OOM quantity is less than the amount of new capacity needed. In both cases, the OOM capacity can lower prices inefficiently;
- The ISO should not treat rejected de-list bids as OOM capacity;
- The ISO should permanently model the capacity zones in order to allow capacity prices to reflect local capacity requirements; and
- The ISO should develop market power mitigation measures to support changes to the market design.

J. Table of Recommendations

Recommendation	Section	High Benefit	Feasible in ST[15]
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 corrected 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.F		
Ancillary Services Markets			
5. Eliminate the "Rest of System" TMOR forward reserve requirement because it is not necessary.	V.E		✓
6. Consider replacing the forward reserve market's price cap with a tiered cap to recognize higher-value reserves.	V.E		✓
7. 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			
8. Modify demand response resources' obligations to be comparable to those of generation resources and imports.	X.C	✓	✓
9. Revise the APR or any replacement provisions such that they more fully mitigate the price effects of OOM entry and do not treat rejected de-list bids as OOM.	X.C	✓	✓
10. Permanently model the capacity zones in order to allow capacity prices to reflect local capacity requirements.	X.C	✓	✓
11. Modify market power mitigation measures to be effective given the changes in the market design.	X.C	Needed for 10	✓
System Operations			
12. Develop provisions to coordinate the physical interchange between New York and New England in real-time.	III.C	✓	
13. Evaluate assumptions made in its capacity evaluation process to determine when supplemental commitments are needed, particularly import/export assumptions.	VII.E		✓

15 *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 2009. 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 2008 and 2009. 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.[16] The average New England Hub price decreased from about $83 per MWh in 2008 to approximately $43 per MWh in 2009. The 48 percent decrease was primarily due to significant decreases in fuel prices from 2008 to 2009. The average natural gas price fell from $10.16 per MMbtu in 2008 to $4.86 per MMbtu in 2009, a 52 percent decrease. 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 2008 and 2009.[17] The figure also shows the average natural gas price,[18] which is a key driver of electricity prices when the market is operating competitively. In 2009, approximately 41 percent of the installed generating capacity in New England burns natural gas as its primary fuel.[19] Low-cost coal and nuclear resources typically produce at full output, while natural gas-

16 Future contracts are currently listed on the New York Mercantile Exchange and Intercontinental Exchange that settle against day-ahead and real-time LMPs at the Hub.

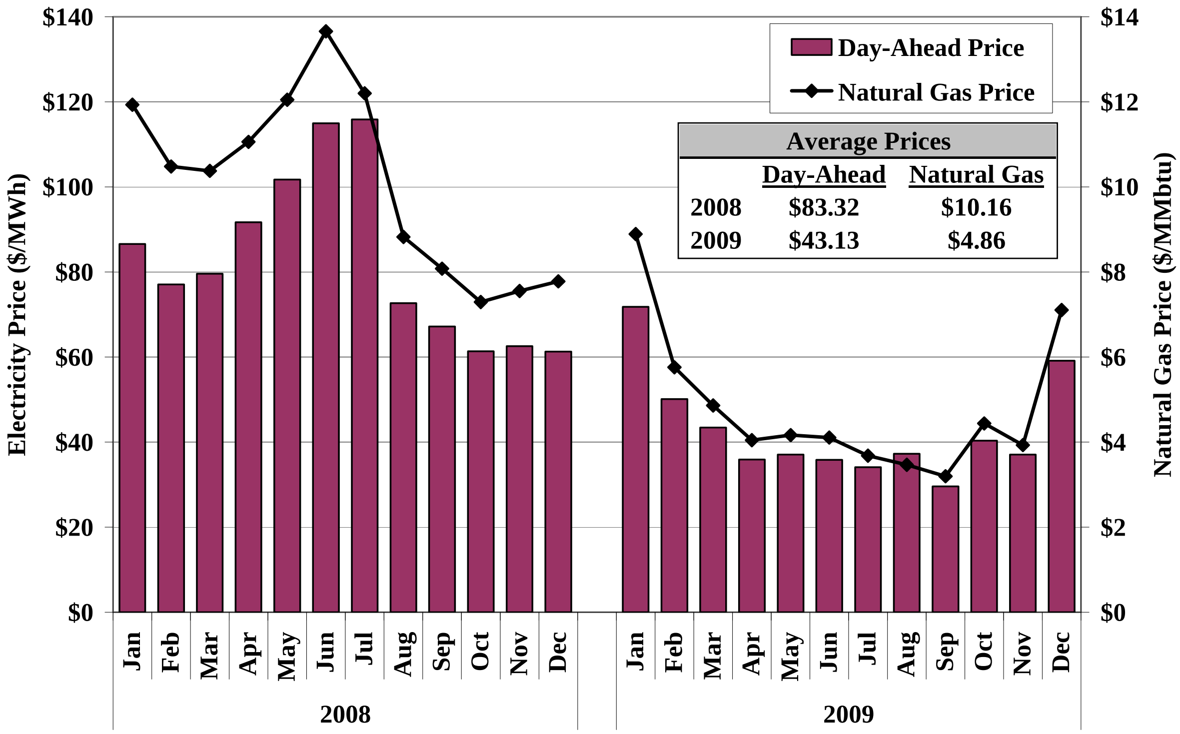
17 This average is weighted by the New England load level in each hour.

18 The figure shows the day-ahead price reported by Platts for the Algonquin pipeline at City Gates.

19 ISO New England, “2009-2018 Forecast Report of Capacity, Energy, Loads, and Transmission (CELT) Report,” April 2009.

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.

Figure 1: Monthly Average Day-Ahead Prices and Natural Gas Prices New England Hub, 2008 – 2009

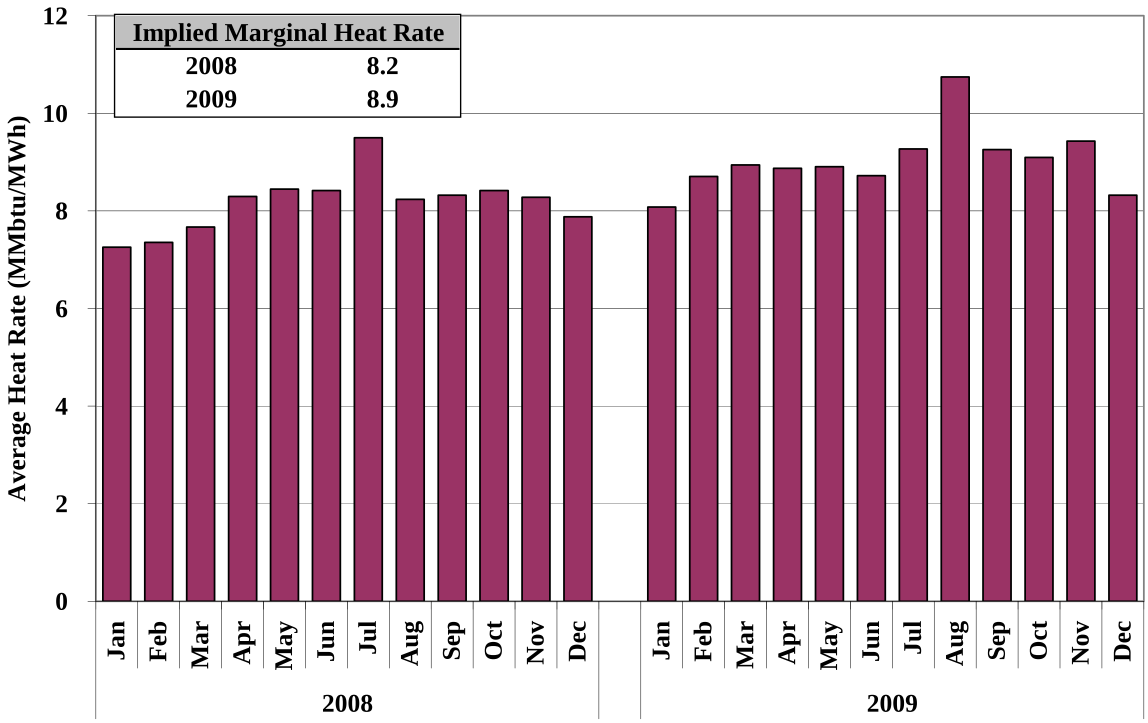


As expected, natural gas price fluctuations were the most significant driver of variations in monthly average electricity prices in 2008 and 2009. Natural gas prices increased sharply from March to June and decreased from July to October in 2008, leading to comparable changes in electricity prices over the same period. Similarly in 2009, the decrease of electricity prices from January to September and subsequent increase through December coincided with the movements in natural gas prices in these months.

Electricity prices usually increase during high summer and winter load periods when the demand for cooling and heating are highest. For example, the monthly average prices in July 2008 and August 2009 both increased from the prior month although the natural gas prices decreased. However, these effects of seasonal changes in demand were smaller than the effects of changes in fuel prices in 2009. The average load was highest during the summer in 2009, but the average electricity prices were the lowest during this period due to the unusually low natural gas prices.

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 2008 and 2009.

Figure 2: Monthly Average Implied Marginal Heat Rate
Based on Day-Ahead Prices at New England Hub, 2008 – 2009



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 peak summer months 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 implied marginal heat rates were July 2008 and August 2009, which were also the months with the highest average loads.

The average implied marginal heat rate rose approximately 10 percent from 2008 to 2009. During the summer months, the implied marginal heat rate averaged 8.6 MMbtu per MWh in 2008 and 9.5 MMbtu per MWh in 2009. Outside the summer, the average implied marginal heat rate rose from an 8.0 MMbtu per MWh in 2008 to 8.8 MMbtu per MWh in 2009. The increase in the implied heat rate was due to several factors:

- Some of the generation costs are not related to fuel prices, leading the implied heat rate to rise as fuel prices fall. For example, variable operating and maintenance ("VOM") expenses typically do not change with fuel prices.
- Generators incurred additional costs to produce power in 2009 due to the Regional Greenhouse Gas Initiative ("RGGI") compliance obligations, which require fossil fuel-fired generators to purchase allowances to cover their emissions since January 2009.
- Real-time price spikes were more frequent in 2009 – real-time prices that correspond to an implied heat rate greater than 20 MMbtu per MWh occurred in 104 hours in 2009 versus 36 hours in 2008. This increase is partly due to a reduction in the average quantity of surplus capacity in 2009 from prior years. Surplus capacity is the amount of generation online in excess of the energy and reserve needs of the system. It decreased because generator commitments for local reliability needs decreased substantially in 2009, which is discussed in greater detail in Section VII.

B. Prices in Transmission Constrained Areas

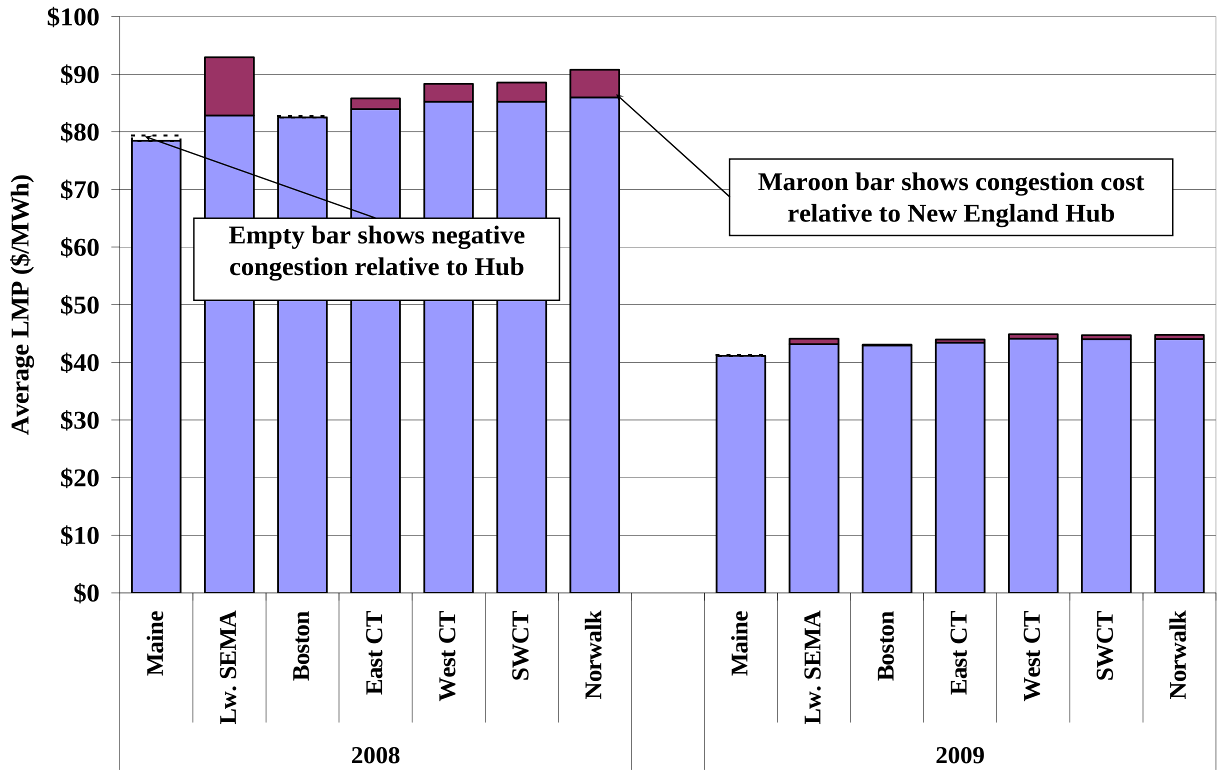
Historically, there have been significant transmission limitations between net-exporting and net-importing regions in New England. In particular, exports from Maine to the rest of New England have frequently been limited by transmission constraints, while Connecticut and Boston were often unable to import enough power to satisfy demand without dispatching expensive local generation in the past. ISO New England uses locational marginal prices ("LMPs") 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 I.

We analyzed the differences in energy prices between several key locations during the study period. Figure 3 shows load-weighted average day-ahead LMPs in 2008 and 2009 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. 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.

Figure 3: Average Day-Ahead Prices by Location 2008 – 2009



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

Congestion declined significantly throughout New England from 2008 to 2009. Lower SEMA remained the area most affected by congestion in 2009, although the average congestion price difference between the Hub and Lower SEMA fell from approximately $10 per MWh in 2008 to less than $1 per MWh in 2009. Congestion into areas in Connecticut also decreased notably from 2008 to 2009. The average congestion price difference between the Hub and the Connecticut area ranged from $2 per MWh (East Connecticut) to $5 per MWh (Norwalk-Stamford) in 2008 compared to a range of 50 cents per MWh (East Connecticut) to 80 cents per MWh (West Connecticut) in 2009. There was virtually no congestion into Boston in 2008 and in 2009 as a result of transmission improvements into the area that were completed in 2007.

Several factors contributed to the substantial decline in congestion from 2008 to 2009. First, average natural gas prices fell more than 50 percent from 2008 to 2009. Redispatch costs incurred to manage congestion are highly correlated with natural gas prices because gas-fired resources are typically the marginal resources redispatched. Hence, congestion costs should decrease as fuel prices decrease. Second, the average load declined approximately 620 MW or 4 percent from 2008 to 2009, which reduced the demand for imports into historically import-constrained areas. [20]

The third and most significant factor was that transmission upgrades increased the transfer capability into the constrained areas, which greatly reduced the frequency of congestion. Short-term transmission upgrades in Lower SEMA were placed in-service in early July that significantly increased the thermal transfer capability of the Lower SEMA transmission interface. This virtually eliminated congestion into Lower SEMA in the last six months of 2009. The transmission upgrades under Phase II of the Southwest Connecticut Reliability Project (Middletown to Norwalk Project) were also completed in 2009. These upgrades significantly improved the transmission system infrastructure and increased the transfer capability into the Norwalk-Stamford and Southwest Connecticut areas. Congestion into Boston has been very limited since the spring of 2007 when the NSTAR 345 kV Transmission Project was placed in-service, which substantially increased the transfer capability into Boston.

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. An outage or failure to secure fuel can force them to purchase replacement high-priced energy from the spot market.

---

20 Load is equal to total generation plus net interchange. This includes load associated with transmission losses and pump storage units.

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 expected to be lower than real-time prices, buyers would increase purchases and sellers would decrease sales in the day-ahead market.

Good convergence between day-ahead and real-time prices is important. Since the day-ahead market facilitates most of the energy settlements and generator commitments in New England, good price convergence with the real-time market helps ensure efficient day-ahead generator commitments and external schedules that reflect actual real-time operating needs.

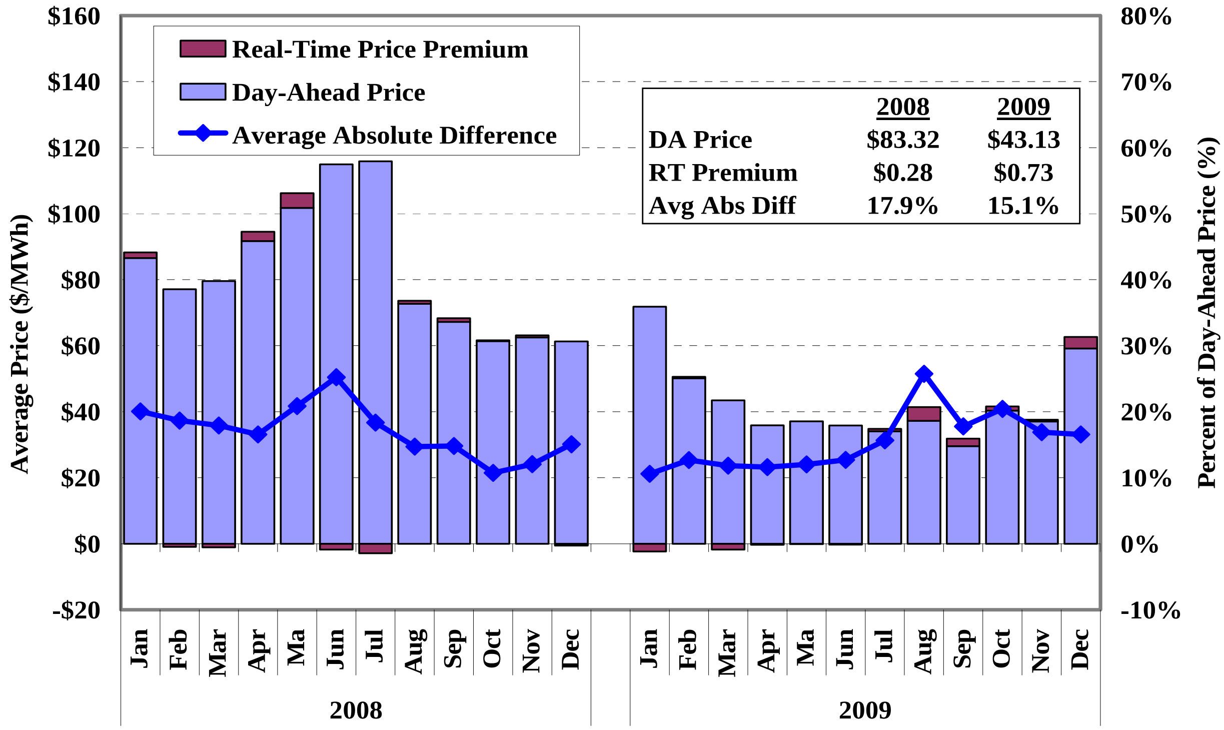
In the remainder of this section, we evaluate the convergence of prices between day-ahead and real-time markets. Section D examines convergence of energy prices at the New England Hub, which is broadly representative of the New England market. Section E examines convergence of energy prices in several 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. This is the most important measure because it indicates whether the day-ahead prices reflect an accurate expectation of real-time prices. 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 2008 and 2009.[21] 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.

Figure 4: Convergence of Day-Ahead and Real-Time Prices at New England Hub 2008 – 2009



In electricity markets, average day-ahead prices tend to be slightly higher than average real-time prices. This may be partly because many buyers are willing to pay a small premium to purchase at day-ahead prices to avoid the more volatile real-time prices. In addition, load settling in the real-time market is usually subject to higher NCPC allocations. While this is consistent with the

21 These are averaged on a load-weighted basis.

day-ahead premium that was evident in some months in 2008 and 2009, more than half of the months exhibited a real-time premium. In fact, each of the last six months of 2009 exhibited real-time premiums. This may be due to the declining surplus capacity (generation online in excess of the energy and reserve needs of the system) that occurred in this timeframe as the ISO committed less generation after the day-ahead market for reliability purposes. These trends and the underlying causes are evaluated in detail in Section VII. For the purposes of the price convergence results above, we would note that declining operating capacity margins in real time will generally lead to increased price volatility and higher average real-time prices. The real-time price premiums that prevailed in the second half of 2009 may indicate that participants have not yet fully adjusted their real-time price expectations to account for the lower operating capacity margins, which are the basis for their day-ahead bids and offers.

Overall, however, both years exhibited a relatively small average real-time premium -- $0.28 per MWh in 2008 and $0.73 per MWh in 2009. These 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 decreased modestly from 17.9 percent in 2008 to 15.1 percent in 2009. 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 can be diminished by fuel price volatility, which increases the participants' uncertainty 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.[22] 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 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 2008 – 2009

	Real-Time Clearing Price ($/MWh)		Day-Ahead - Real-Time Price Difference ($/MWh)		Hourly Absolute Price Difference (percent of RT Price)	
Region	**2008**	**2009**	**2008**	**2009**	**2008**	**2009**
New England Hub	$83.60	$43.86	-$0.28	-$0.73	17%	14%
Maine	$78.00	$41.72	$0.45	-$0.59	17%	14%
Lower Southeast Massachusetts	$88.28	$44.31	$4.65	-$0.20	20%	15%
Boston	$83.35	$43.67	-$0.80	-$0.58	17%	14%
Areas in Connecticut:						
East Connecticut	$85.13	$44.29	$0.67	-$0.34	18%	14%
West CT (excluding SWCT)	$87.11	$45.17	$1.20	-$0.29	18%	15%
SWCT (excluding Norwalk)	$86.97	$45.06	$1.58	-$0.33	18%	15%
Norwalk-Stamford	$88.67	$45.35	$2.12	-$0.59	19%	15%

Price convergence is generally slightly better in the less congested locations because unforeseen market events tend to have larger price effects in isolated areas. However, this was not the case in 2009 as the largest average price difference was at the New England Hub. The difference

22 The average day-ahead and real-time prices are load-weighted average prices.

between the average day-ahead price and average real-time price was generally higher in the historically import-constrained locations than at the Hub. The average absolute difference shows a similar pattern.

For example, Lower SEMA exhibited poor convergence between the day-ahead and the real-time market in 2008. The difference between the average day-ahead and real-time prices was $4.65 per MWh. This was due primarily to changes in the commitment of key generators after the day-ahead market. On many days, the majority of the generation in Lower SEMA 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. In 2009, convergence improved considerably in Lower SEMA as a result of transmission upgrades that nearly eliminated the need for reliability commitment.

In general, price convergence improved in the historically import-constrained areas from 2008 to 2009 as congestion into these areas decreased substantially. The average absolute price difference fell from 20 percent in 2008 to 15 percent in 2009 in Lower SEMA and from 19 percent in 2008 to 15 percent in 2009 in Norwalk-Stamford. These declines were more significant than at unconstrained locations such as the New England Hub.

Overall, we find that the 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 Interconnection. 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 2009. 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.

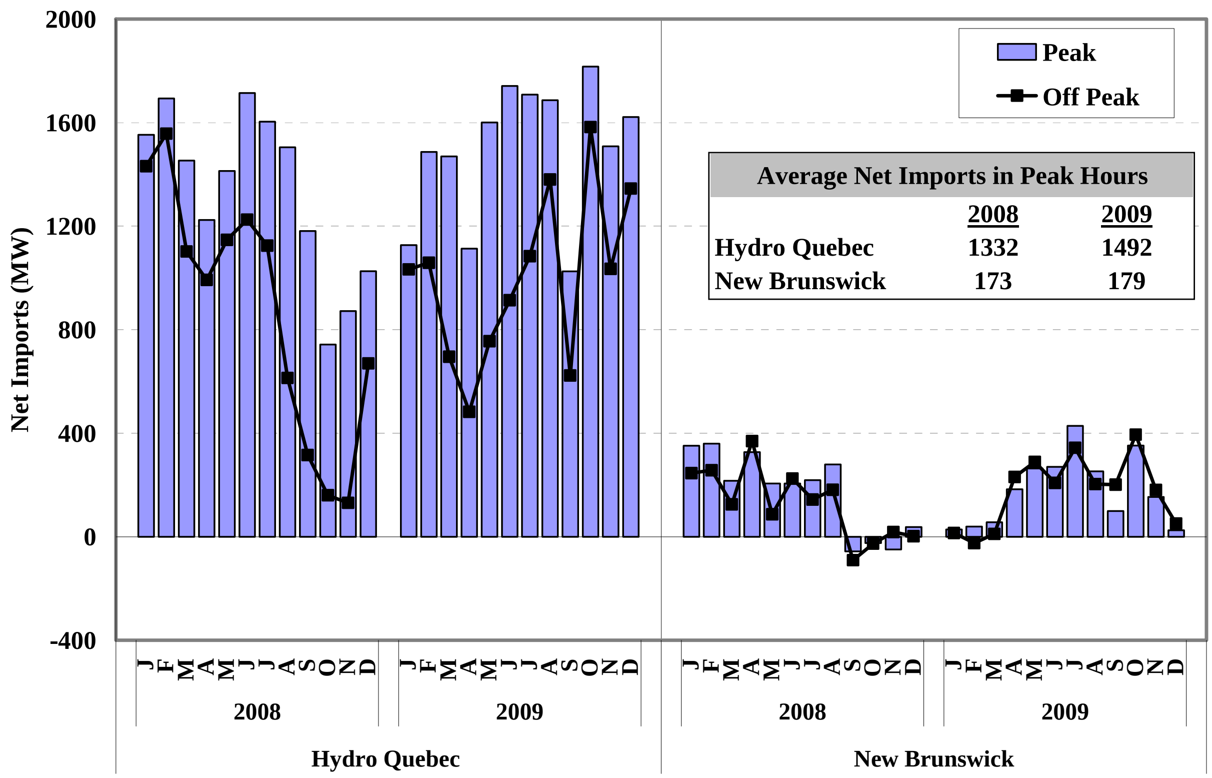
ISO-NE is interconnected with three neighboring control areas: the New York ISO, TransEnergie (Quebec), and the New Brunswick System Operator. New England and New York are interconnected by three interfaces: the Roseton Interface, which includes several AC tie lines connecting upstate New York to Connecticut, 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 2008 and 2009. 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 2008 and 2009. Figure 5 shows the average net imports across the three interfaces with Quebec and New Brunswick by month, for peak and off-peak periods.[23] The net imports across the two interfaces linking Quebec to New England are combined.

Figure 5: Average Net Imports from Canadian Interfaces
2008 – 2009



Average Net Imports in Peak Hours	2008	2009
Hydro Quebec	1332	1492
New Brunswick	173	179

23 Peak hours include hours ending 8 to 23, Monday through Friday (not including NERC holidays), 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 460 MW in 2008 and by 490 MW in 2009. This reflects 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.

Hydro Quebec tends to sell more power to New England during months when electricity prices are high. This was evident in 2008 when electricity prices were elevated in the first half of the year primarily due to high natural gas prices, and then fell considerably in the second half of the year. Although electricity prices fell substantially from 2008 to 2009 due to the decline in fuel prices, net imports from Hydro Quebec increased modestly from 2008 to 2009. This unexpected increase may be partly due to lower demand in Quebec. At the end of 2008, Hydro Quebec reduced its ten-year forecast of demand from native load customers in Quebec, allowing it to export more power to neighboring areas.[24]

Figure 6 shows average net imports across the three interfaces with New York by month in 2008 and 2009 for peak and off-peak periods. The net imports across the Cross-Sound Cable and the 1385 Line are combined. The 1385 Line came back into service at the beginning of January 2009 after being out of service for most of 2008 due to cable replacement work and problems with the phase shifter. Due to the outages of the 1385 Line, the combined net imports in 2008 reported in the figure primarily reflect flows across only the Cross-Sound Cable interface.

Figure 6 shows that the direction and the level of flows varied considerably across the primary interface with New York during the two years. Power usually flowed into New York during peak periods and into New England during off-peak periods.

[24] See the *Hydro Quebec Strategic Plan 2009 – 2013*, page 3.

Figure 6: Average Net Imports from New York Interfaces
2008 – 2009

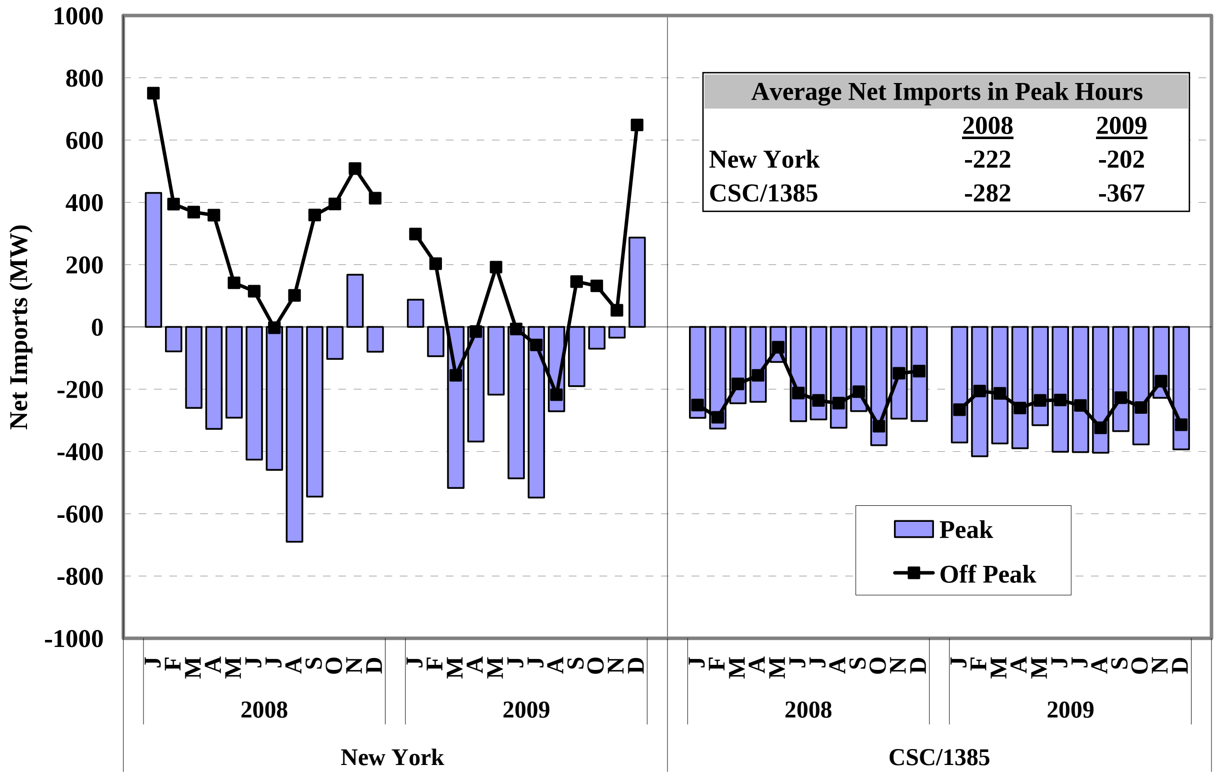


Figure 6 also indicates that New England tends to import more during the winter months when slightly colder temperatures occur in New England than New York. Net imports are higher in off-peak hours as well when New York has lower-cost supply that can displace natural gas-fired generation that often remains on the margin in these hours. However, New England was a net exporter to New York overall in both 2008 and 2009. On average, the net exports to New York were 360 MW in 2009, up from 190 MW in 2008.

The figure shows that flows were relatively consistent from New England to Long Island across the Cross-Sound Cable and the 1385 Line, averaging approximately 250 MW in 2008 and 310 MW in 2009. The average quantity of exports increased from 2008 to 2009 primarily because the 1385 Line returned to service in January 2009. The Cross-Sound Cable and the 1385 Line, which have a transfer capability of 330 MW and 100 MW respectively, were fully utilized to export power to Long Island in most of the hours that they were in service.

B. Interchange with New York

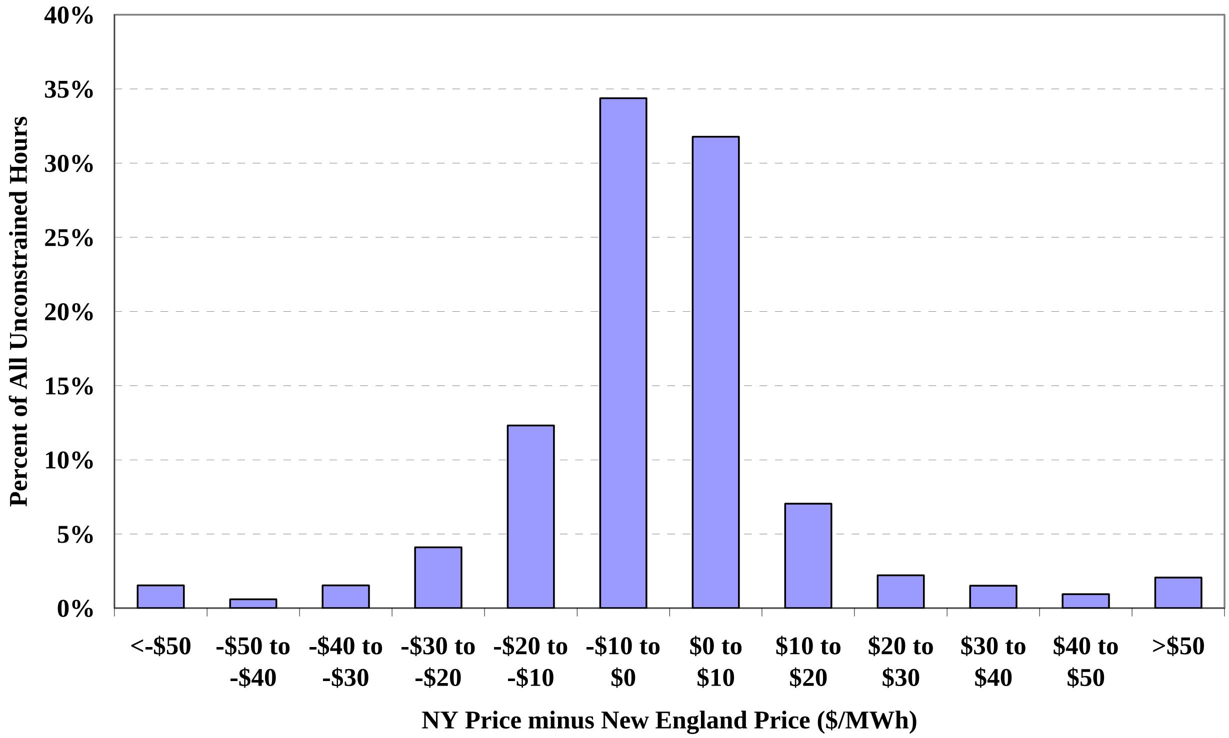
The performance of New England's wholesale electricity markets depends not only on the efficient use of internal resources, but also the efficient use of transmission interfaces with adjacent areas. This section evaluates the efficiency of scheduling between New England and New York. Since both regions 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.[25] The 1385 Line is also not included because it was usually fully scheduled from Connecticut to Long Island in 2009. 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 not constrained.[26]

Figure 7: Real-Time Price Difference Between New England and Upstate New York Unconstrained Hours, 2009



25 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.

26 The prices used in this analysis are the prices at the New England proxy bus in the New York market (i.e., New York price) and the prices at the New York proxy bus in the New England market (i.e., New England price).

While the factors described 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 approximately 34 percent of the unconstrained hours have real-time price differences of greater than $10 per MWh. In more than 5 percent of the hours, the price difference is greater than $40/MWh.

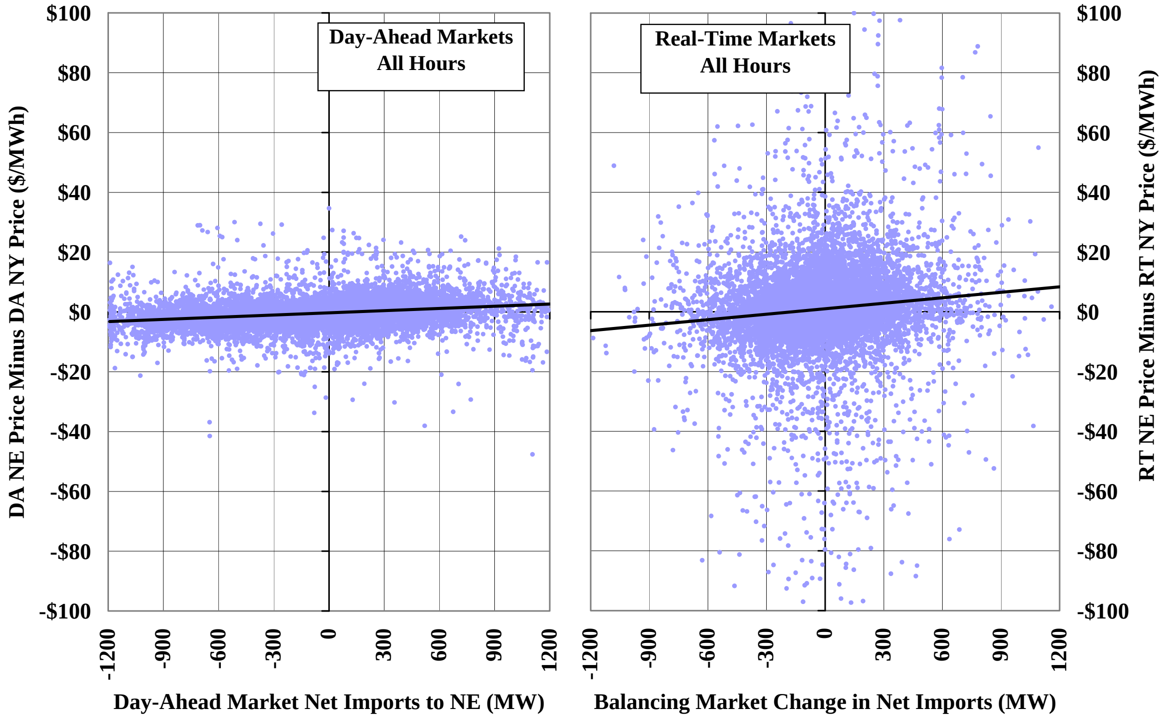
These results indicate that the current process does not fully utilize the interface. Given the pattern of price differences shown, there are many hours when increasing flows from the lower priced region to the higher priced region would have significantly improved the efficiency of clearing prices and production in both regions. This failure to fully arbitrage the interfaces leads to market inefficiencies that could be remedied if the ISOs were to coordinate interchange.

The prior analysis shows that market participants have not fully arbitraged the interfaces between New York and New England. The next analysis evaluates whether the incremental changes in participants' schedules have been consistent with the relative prices in the two regions and have, therefore, improved price convergence and efficiency.

Figure 8 shows, for each hour in 2009, 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 lines presented in each panel of the figure show statistically significant positive correlations between the price difference and the direction of scheduled flows in the day-ahead and real-time markets. The 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, 2009



Although the arbitrage is not complete, the positive correlation between the price differences and the schedule changes indicate that participants respond rationally to the price differences. Additionally, total net revenues from cross-border scheduling in 2009 were $5.5 million in the day-ahead market and $4.4 million in the real-time market (not accounting for transaction costs).[27] 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.

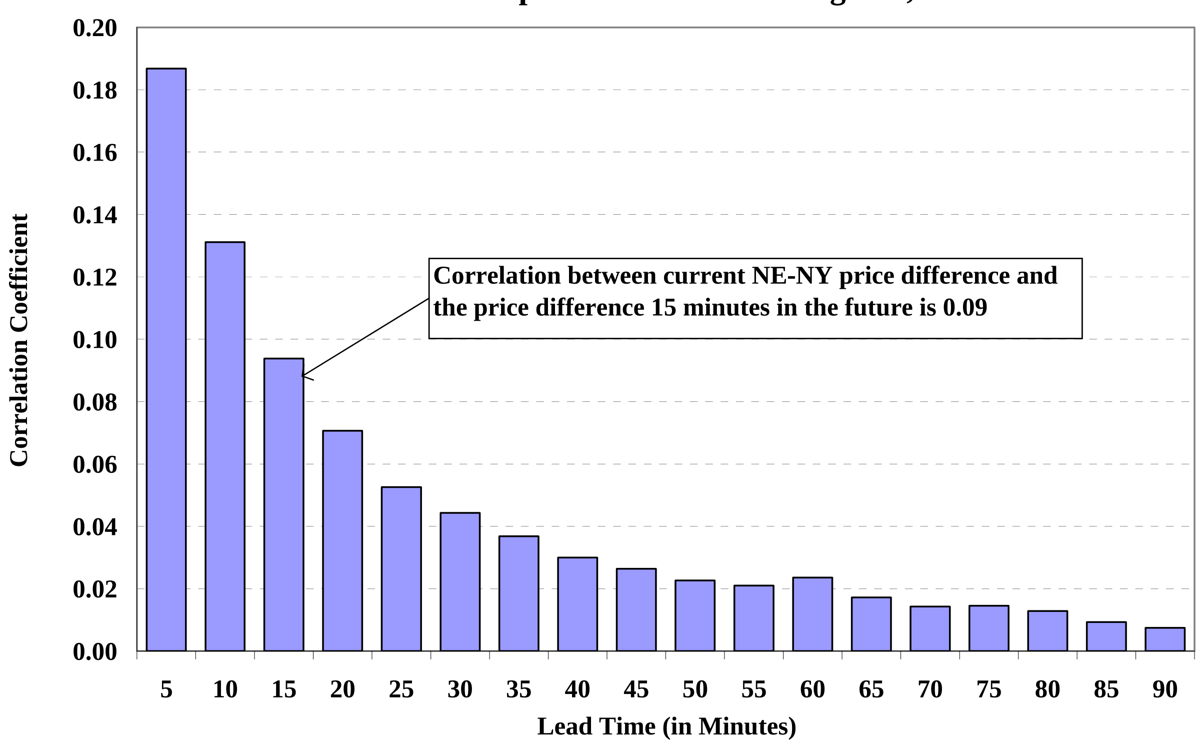
Nevertheless, the difficulty of predicting changes in market conditions in real-time is reflected in the wide dispersion of points most notable on the right side of Figure 8. Forty-three percent of

27 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.

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 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 90 minutes. For example, the correlation of the price difference at the current time and the price difference 15 minutes in the future was 0.09 in 2009.

Figure 9: Correlation Between Price Differences and Lead Time
Interface between Upstate NY and New England, 2009



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 shows that reducing the lead times for scheduling would improve participants' ability to forecast the price differences and determine their schedules. However, it also shows that the correlation remains less than .05 at 30 minutes ahead of real time. This is the shortest scheduling lead time currently used by any RTO. Hence, the likely benefits of reducing scheduling lead-times are small and it is not adequate to address the inter-ISO price convergence issues described above. The next section describes how these issues can be more completely addressed through explicit coordination.

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 or an intrahour scheduling process to achieve full arbitrage.

We employed simulations to estimate the benefits of optimal hourly scheduling of the primary interface between New England and New York from 2006 to 2009. 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. Table 2 summarizes the results of this analysis.

Table 2: Estimated Benefits of Coordinated External Interface Scheduling
Interface Between Upstate NY and New England, 2006 – 2009

	2006	2007	2008	2009
Estimated Production Cost Net Savings (in Millions)	**$17**	**$21**	**$19**	**$10**
Estimated Consumer Net Savings (in Millions):				
New England Customers	**$61**	**$22**	**$25**	**$64**
New York Customers	**$59**	**$177**	**$127**	**$65**
Total for New England and New York Customers	**$120**	**$199**	**$152**	**$129**
During Reserve Shortage Hours	**$16**	**$75**	**$31**	**$13**

The simulations indicate that better coordination would lead to lower average prices and net savings for consumers in both regions. Adjacent regions are brought into better convergence as lower-cost resources in one area displace higher-cost resources in the adjacent area, which lowers the total production costs in the two areas. In each hour, better convergence leads to higher prices in one area and lower prices in the other. However, our simulations indicate that consumers in both areas would benefit from lower average prices because prices generally decrease more in the high-price area than they rise in the low-price area. This result is due to the nonlinear shape of the supply curve in electricity markets, which causes prices to be more responsive to changes in interchange at higher price level than lower prices levels. Estimated consumer net savings averaged $64 million annually for New England customers and $65 million annually for New York customers in 2009. New York's higher share of consumer savings in prior years is primarily attributable to the higher frequency of shortages in New York during those years.

Shortage pricing provisions in both the New York and New England markets have contributed to more efficient pricing when reserve shortages occur. However, shortage pricing increases the importance of coordinating the physical interchange between the ISOs because full utilization of the interfaces can allow the ISOs to avoid unnecessary shortages. Our analysis suggests that ISO coordination of external flows would have reduced consumer costs incurred during reserve shortages by an average of $34 million annually. These benefits tend to be understated because surpluses have prevailed in recent years that have substantially reduced the frequency of shortages. Hence, as capacity margins decrease and the frequency of shortages increase, the total savings for New England customers should increase.

Production cost savings represent the net efficiency benefits of improving coordination between areas. The estimated production cost savings naturally tend to be smaller than estimated consumer net savings. In most cases, a small quantity of lower-cost generators in one area displaces a small quantity of higher-cost generators in the other area, which results in limited production cost savings. Hence, the production cost savings are smaller than the price effects, averaging $17 million annually. ISO New England and the New York ISO have discussed market reforms that would improve the utilization of the New York interface by coordinating the physical interchange, or by allowing participants to submit “spread offers” that would allow transactions to be scheduled intrahour when the spread is larger than the offer. We strongly support these initiatives and believe that they would allow the two markets to capture most of the benefits estimated in this section.

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 either direction between New England and New York, depending on market conditions in each region.

We find that the external transaction scheduling process is functioning properly and that scheduling by market participants tends to improve convergence, but significant opportunities remain to improve scheduled interchange between regions. 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.

ISO New England is working with the New York ISO on two initiatives that are intended to improve the efficiency of scheduling between the two control areas.[28] First, Interregional Transaction Scheduling Coordination will create a system for accepting and clearing intra-hour transactions between control areas based on expected near-term price differences. Second, Market-to-Market Congestion Management Coordination will develop procedures for enabling one ISO to redispatch its internal resources to relieve congestion in the other control area when it is efficient to do so. We strongly recommend that ISO-NE and the New York ISO place a high priority on developing these market enhancements.

28 See Comments of ISO New England Inc.; Docket No. ER08-1281-004; February 2, 2010.

IV. Transmission Congestion and 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.[29]

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 a payment equal 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

29 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.[30] FTRs are auctioned separately for peak and off-peak hours.[31]

In this section, we summarize the congestion costs that have occurred in New England markets and assess two aspects of the performance of the FTR markets. First, we evaluate the net payments to FTR holders. The net payments to FTR holders decreased approximately 75 percent from 2008 to 2009 due to the considerable 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 2009, the congestion revenue collected by the ISO was sufficient to satisfy 97 percent of the obligations to FTR holders (referred to as the “target payment amount”). In 2008, the congestion revenue was sufficient to pay 100 percent of the FTR obligations.

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 2009, FTR prices in the monthly auctions were much more consistent with congestion values in the day-ahead and real-time markets than FTR prices in the annual auction. The annual FTR price significantly over-forecasted congestion-related price differences. The substantial improvement in consistency of FTR prices and congestion values from the annual auction to the monthly auctions is expected because market participants gain additional information about market conditions.

30 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.

31 Peak hours include hours ending 8 to 23, Monday through Friday, not including NERC holidays. Off-peak includes all other hours.

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

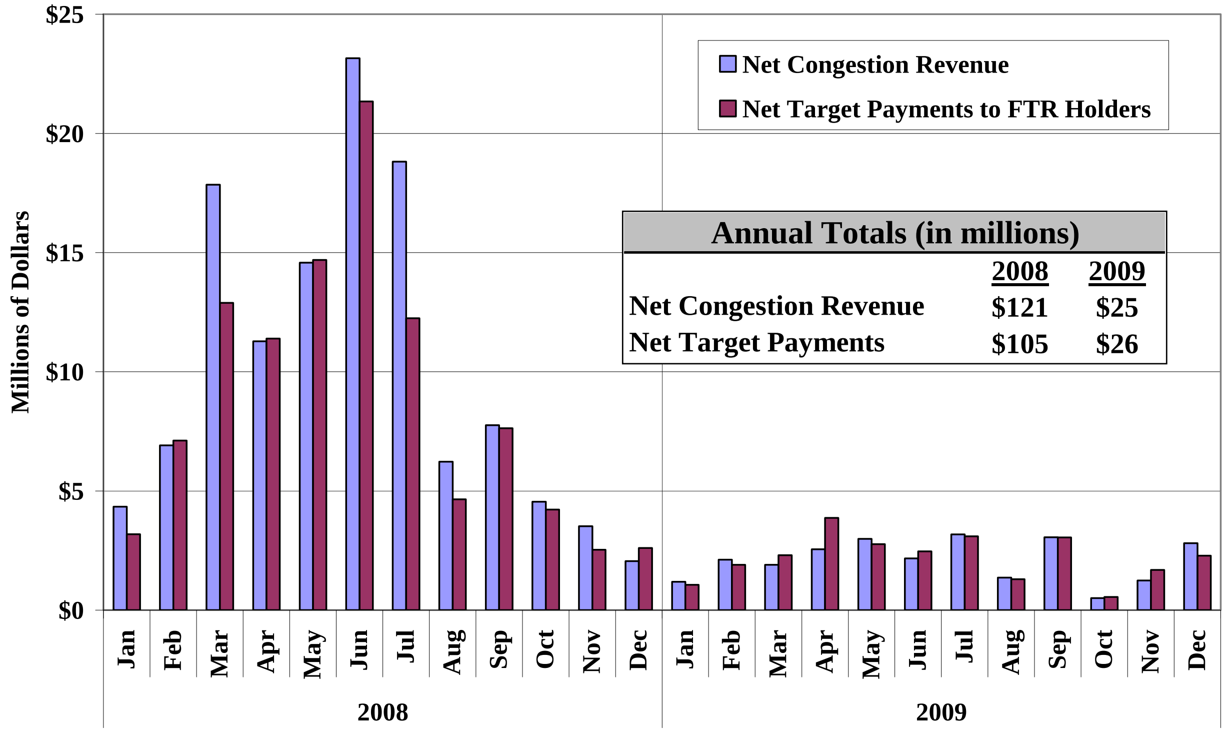
- Day-ahead congestion revenue is equal to the megawatts scheduled to flow across a constrained interface times the shadow price (i.e., the marginal economic value) of the interface.
- Real-time congestion revenue is equal to the change in scheduled flows (relative to the day-ahead market) across a constrained 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.[32]

When the total 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. In months when this occurs, the unpaid FTR 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.

32 For example, suppose 100 MW is scheduled to flow across an interface in the day-ahead market in a given hour, 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 for the hour.

Figure 10 compares the net congestion revenue collected by the ISO with the net target payments to FTR holders in each month of 2008 and 2009. 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 holders and all negative target payments (i.e., payments from FTR holders).

Figure 10: Congestion Revenue and Target Payments to FTR Holders 2008 – 2009



From 2008 to 2009, the net congestion revenue declined from $121 million in 2008 to $25 million in 2009, a 79 percent decrease. Likewise, the net target payments to FTR holders declined substantially from $105 million in 2008 to $26 million in 2009. The sharp reduction in congestion in 2009 was primarily due to:

- Transmission additions in Southwest Connecticut and Lower SEMA, and
- Substantial decreases in fuel prices, which reduce redispatch costs and associated congestion-related price differences.

The patterns of congestion are evaluated in greater detail in subsection B below.

The figure shows that the number of months when net congestion revenues were less than the net target payments to FTR holders increased from just two months in 2008 to five months in 2009. The total net congestion revenues for the 12 months in 2008 were sufficient to fund the net target payments, while net congestion revenues in 2009 were only sufficient to fund 97 percent of the net target payments. As a result, positive target payments to FTR holders were reduced by an average of 3 percent.[33]

The net target payments to FTR holders exceed congestion revenues when the amount of FTRs purchased along a congested transmission corridor is larger than the actual transfer capability in the day-ahead market. For example, assume 1,000 MW of FTRs are sold into a constrained area because that is the normal limit into the area. If the interface is reduced to 700 MW in the day-ahead market due to a transmission outage and the interface is congested, the ISO will only collect 70 percent of the congestion revenue it needs to satisfy the target payments to the holders of the FTRs into the constrained area.

B. Congestion Patterns and FTR Prices

In this section, we evaluate the performance of the FTR markets by comparing 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. When the market is performing well, the FTR prices should converge over time with the actual congestion on the network.

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 2009. The congestion prices shown are calculated for peak hours relative to the New England Hub. Hence, if the congestion price in the

33 In 2009, the ISO collected $25 million in net congestion revenue and $13 million from FTR holders that were obligated to make payments to the ISO. The total collection of $38 million was $1 million less than the $39 million needed to cover positive targeted payments to FTR holders.

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 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. 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.

Figure 11: FTR Auction Prices vs. Day-Ahead and Real-Time Congestion Average Difference from New England Hub in Peak Hours, 2009

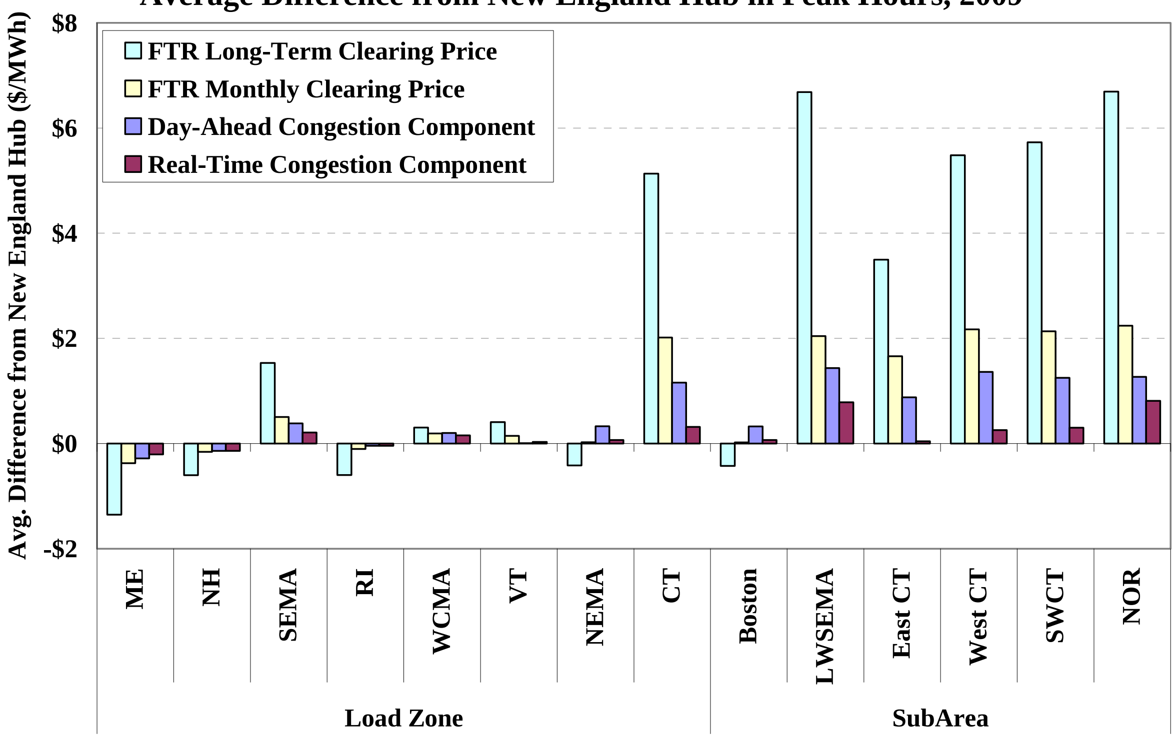


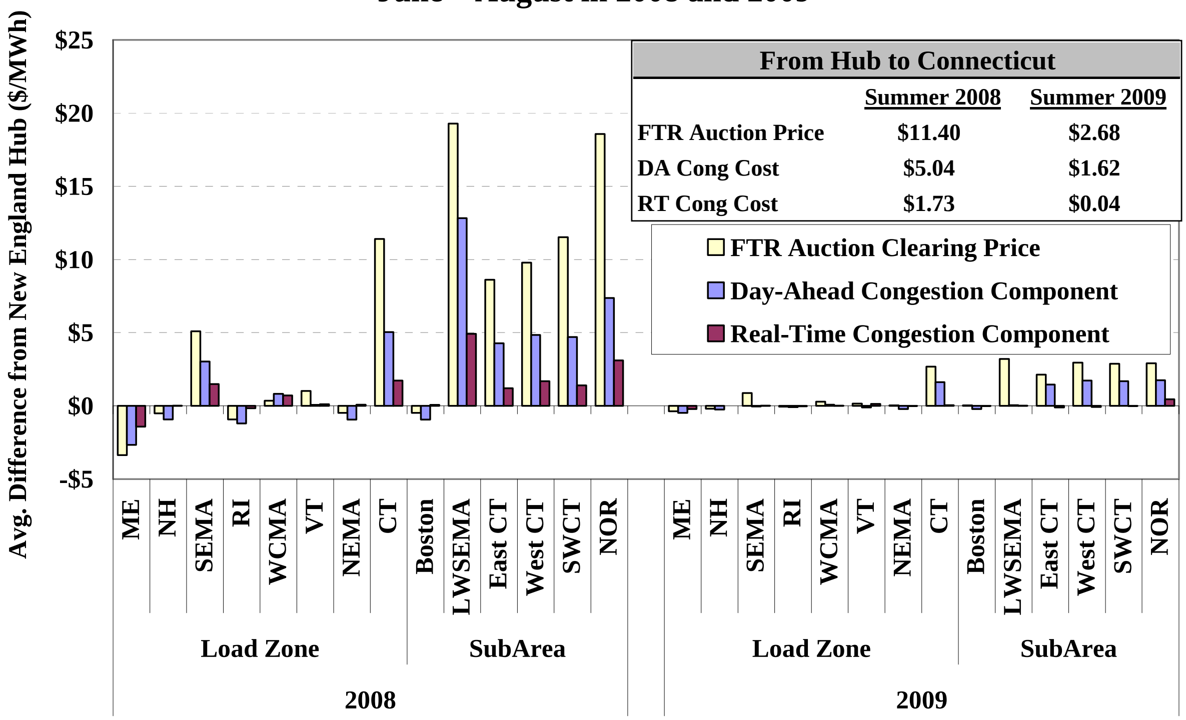
Figure 11 shows that in most areas, monthly FTR prices were relatively consistent with congestion prices in the day-ahead market, while annual FTR prices were less closely correlated with day-ahead congestion prices. For example, the annual FTR prices from the New England Hub to Lower SEMA and the various areas in Connecticut were substantially higher than the day-ahead congestion value. However, the monthly FTR price was only slightly higher than the day-ahead congestion value. 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 monthly auctions.

In fact, the figure shows that annual FTR prices substantially overestimated congestion in most areas. From the Hub to both Lower SEMA and Connecticut, the annual FTR clearing price was more than three times higher than the day-ahead congestion value. These results suggest that market participants underestimated the decrease in congestion associated with the addition of transmission capability in early 2009. As market participants observed less congestion in 2009, they updated their expectations and the monthly FTR prices converged more closely to the day-ahead congestion levels. This is what we would expect in a well-functioning market.

The next analysis presents the same results for the summer months in 2008 and 2009. The analysis focuses on the summer because system peaks generally occur in the summer due to cooling demand. The system peaks are when 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, and real-time congestion for the peak hours during the summer season for the same locations as the previous figure.

In the summers of 2008 and 2009, 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 participants' expectations in the monthly auctions were higher than the 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 more congestion in the real-time market than actually occurred in the summers of 2008 and 2009. The monthly FTR auction prices correctly predicted that there would be very little day-ahead or real-time congestion into Boston in both years.

Figure 12: FTR Auction Prices vs. Day-Ahead and Real-Time Congestion Average Difference from New England Hub in Peak Hours June – August in 2008 and 2009



The over-estimates of congestion in the monthly FTR auctions are likely due in part to the following two factors:

- Load levels in both summers were lower than anticipated, resulting in less congestion in the day-ahead and real-time markets.
- Expectations of congestion were likely influenced by the congestion that occurred the previous summer. Congestion decreased each year from 2007 to 2009, which contributed to the FTRs generally being over-valued in 2008 and 2009.

Given the volatile nature of congestion patterns, we found that FTRs were reasonably valued 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. Additionally, the FTR prices from the annual auction were more consistent with congestion patterns in the previous year than FTR prices from the monthly auction. Therefore, we conclude that the FTR markets performed well in 2009.

V. Reserve and Regulation 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:

- Background on the real-time reserve market;
- Real-time reserve market results;
- 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 to enhance the efficiency of the reserve markets.

A. Background on the 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 1,377 MW in 2009. Based on system conditions, the operator determines how much of the 10-minute reserve requirement to hold as spinning reserves. ISO-NE held an average of 38 percent of the 10-minute reserve requirement in the form of spinning reserves during intervals with binding TMSR constraints in 2009.[34]

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 approximately 2,000 MW in 2009. Since higher quality reserves may always be used to satisfy requirements for

34 The TMSR requirement is binding when a non-zero cost is incurred by the market to satisfy the requirement. This occurred in 1.7 percent of the intervals in 2009.

lower quality products, the entire 30-minute reserve requirement can be satisfied with TMSR or TMNSR.

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, 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. As a result, ISO-NE is able to satisfy the local reserve requirements at a lower cost.

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 is equal to the marginal redispatch cost of 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 52 percent of the Boston requirement during constrained intervals in 2009.

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 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.[35]

35 The RCPF for local 30-minute reserve constraints was changed from $50 per MWh to $250 per MWh, effective January1, 2010.

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 of 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 also reflected in energy clearing prices.[36] 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 $1,000 ($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 has been endorsed by FERC.[37]

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 were $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 modified the RCPF for local reserve zones in January 2010 from $50

---

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

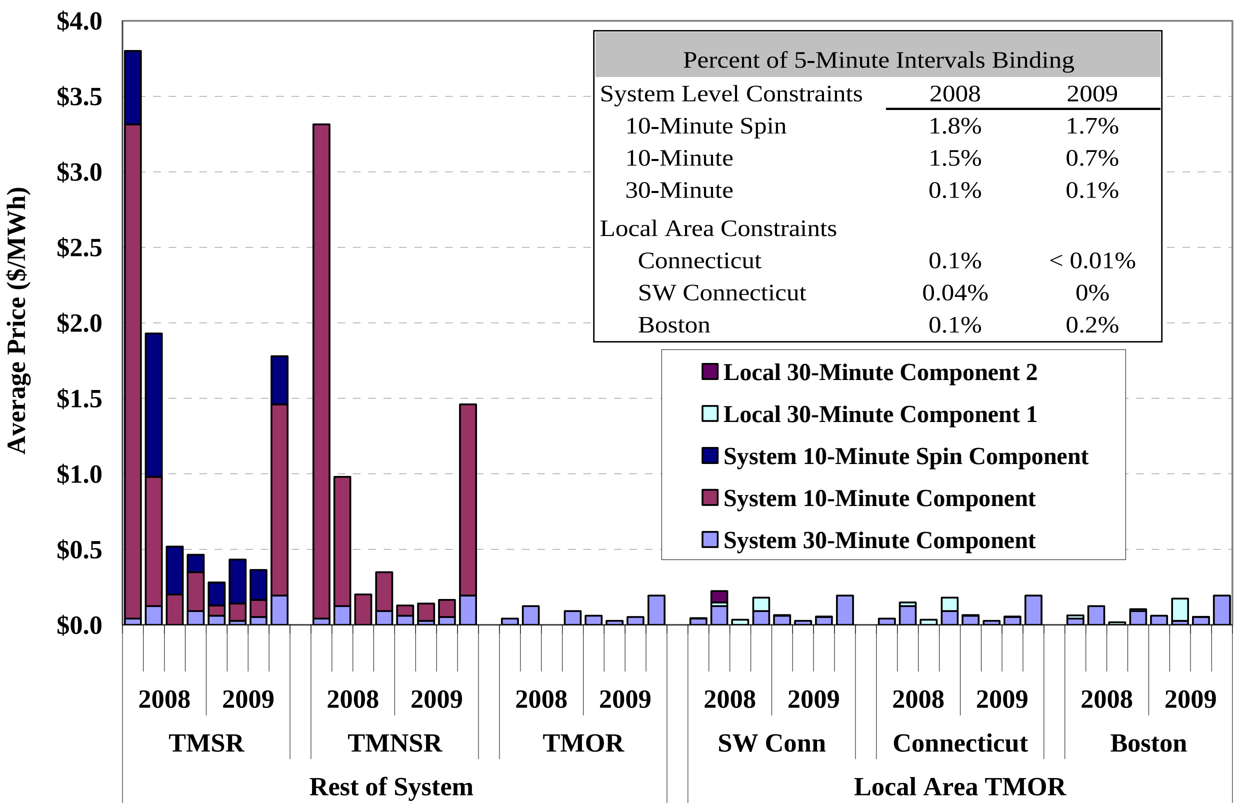
37 *Wholesale Competition in Regions with Organized Electric Markets*, Order No. 719, 73 *Fed. Reg*. 64100 (October 28, 2008), FERC Stats. & Regs. ¶ 31,281 (2008) ("Order No. 719").

per MWh to $250 per MWh, which is sufficiently high to maintain local reserves under such circumstances. The RCPFs are analyzed in greater detail later in Section C.

B. Real-Time Reserve Market Results

Figure 13 summarizes average reserve clearing prices in each quarter of 2008 and 2009. 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: Quarterly Average Reserve Clearing Prices by Product and Location 2008 – 2009



Percent of 5-Minute Intervals Binding		
System Level Constraints	2008	2009
10-Minute Spin	1.8%	1.7%
10-Minute	1.5%	0.7%
30-Minute	0.1%	0.1%
Local Area Constraints		
Connecticut	0.1%	< 0.01%
SW Connecticut	0.04%	0%
Boston	0.1%	0.2%

Outside the local constrained areas, the average TMSR clearing price fell from $3.80 per MWh in the first quarter of 2008 to $0.28 per MWh in the first quarter of 2009 and remained low until the fourth quarter of 2009 when the average price rose to $1.78 per MWh. The TMSR clearing price was highest in the first half of 2008 because binding reserve constraints were more common overnight. Binding reserve constraints occur overnight when less fast-ramping generation is online and available to provide reserves. Pumped storage units may be unavailable when they are in pumping mode. Combined cycle units shutdown overnight when they anticipate that starting up the next morning will be less costly than remaining online overnight. In the first half of 2008, high natural gas prices induced more combined cycle generation to shutdown overnight. The decrease in average TMSR clearing prices in late 2008 and into 2009 was due, in part, to the overall decline in LMPs because the opportunity costs of units providing TMSR are correlated with LMPs.

The TMSR clearing price rose in the last quarter of 2009 because binding reserve constraints became more common around the peak load hours of each day. The increase in TMSR clearing prices was related to the reduction in average surplus capacity, which is the amount of generation online in excess of the energy and reserve needs of the system. The minimum daily surplus capacity fell from an average of 1,300 MW in the first half of 2009 to 750 MW in the second half of 2009.[38] Surplus capacity fell because generator commitments for local reliability needs decreased substantially in 2009, which is discussed in greater detail in Section VII.

In the local areas, TMOR clearing prices were low in both 2008 and 2009. The highest average TMOR clearing price in 2009 was 12 cents per MWh in Boston. This resulted from an average of 4 cents per MWh for the Boston 30-minute reserve component and 8 cents per MWh for the system 30-minute reserve component. In 2009, the TMOR requirements were never binding in Southwest Connecticut and were binding for only three intervals in Connecticut, resulting in the average TMOR clearing prices in both areas equal to the system TMOR clearing prices virtually

38 The minimum daily surplus capacity is the lowest quantity of surplus capacity that was available in any interval on a particular day.

all the time.[39] The low frequency of binding local reserve constraints is primarily due to the transmission upgrades into Boston in the spring of 2007 and multiple upgrades in Connecticut completed between 2007 and 2009.

Average reserve clearing prices were relatively low in 2008 and 2009 because reserve clearing prices were $0 in the vast majority of real-time intervals. This reflects that there is surplus capacity online in most hours that 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 0.7 percent of intervals, implying that the requirement can be met at no cost with surplus capacity in more than 99 percent of intervals. However, when the system-level 10-minute reserve requirement is binding, the clearing price of TMNSR can rise quickly. In 2009, the average TMNSR clearing price was $56 per MWh in intervals when the system-level 10-minute reserve requirement was binding.

C. 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.[40] 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

---

39 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 Boston was 75 cents per MWh. This is composed of 71 cents per MWh for TMSR outside the local areas and the Boston 30-minute reserve component of 4 cents per MWh.

40 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).

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 clearing prices in the process. Hence, it is important to evaluate the RCPF levels periodically to determine whether modifications are warranted. In previous Annual Assessments, 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 increase the local RCPFs.

The ISO conducted a review of the RCPF levels that were used in the local reserve zones. The ISO concluded that the local RCPF of $50 per MWh was not sufficiently high to schedule available resources to satisfy the local reserve requirements under some circumstances, leading the operators to intervene to maintain reserves. Accordingly, ISO-NE increased the local RCPF levels to $250 per MWh in January 1, 2010, which should be high enough to maintain local reserves and set more efficient prices when the system is short of local reserves.

In this section, we evaluate the $50 per MWh RCPF, which was used in 2009, and the new $250 per MWh RCPF to determine how they compare to the costs that the ISO regularly incurs to maintain local reserves. Specifically, the analysis compares the old and new RCPFs to the marginal redispatch costs incurred to maintain local-area reserves in the real-time market in 2008 and 2009.

1. RCPFs and Real-Time Dispatch

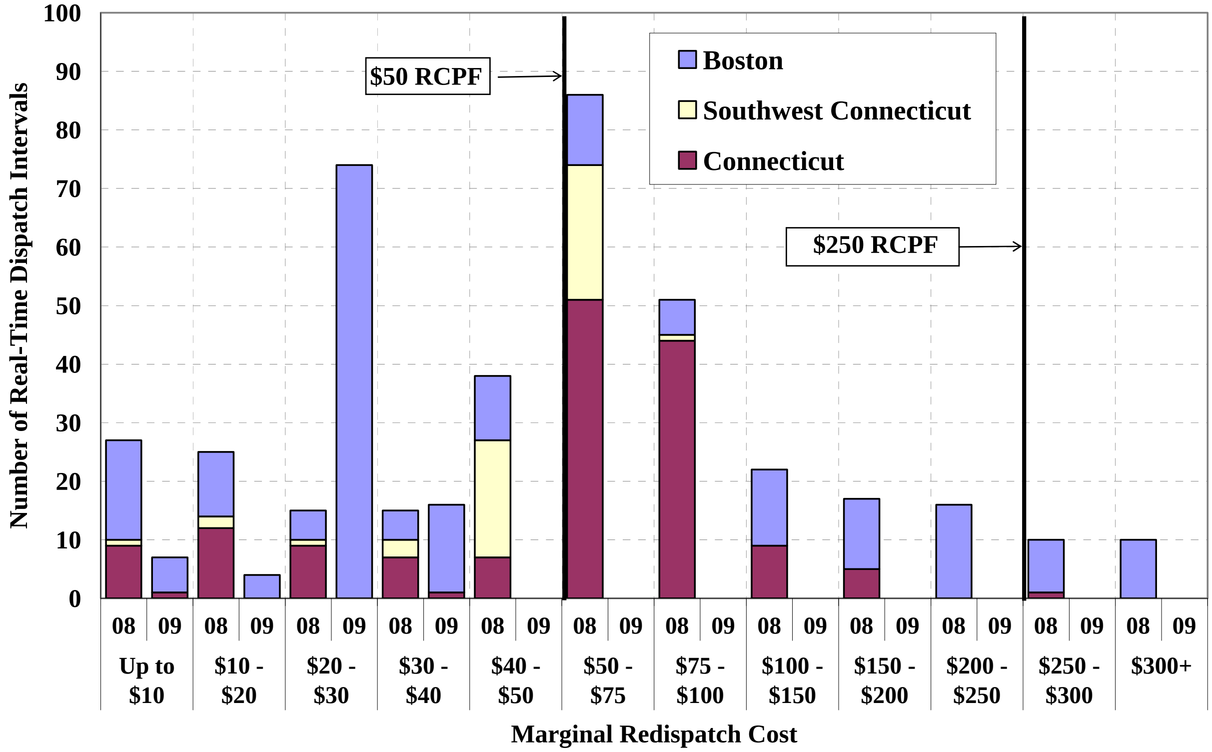
As discussed above, the real-time market may experience a shortage of reserves if the marginal cost of scheduling the available reserves exceeds the RCPF. In such cases, the ISO is required to take additional actions to maintain the required level of reserves if the reserves are available.

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).[41] When possible, the operators use real-time transmission constraints to maintain reserves rather than manual dispatch instructions.

The following analysis compares the $50 per MWh local RCPF and the new local RCPF ($250 per MWh) to the marginal redispatch costs incurred to meet the local-area reserve requirement during the past two years (2008 and 2009). 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 for 2008 and 2009.

41 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).

Figure 14: Marginal Redispatch Costs to Meet Local Reserve Requirements 2008 – 2009



The marginal cost of meeting the local reserve requirements exceeded the $50 per MWh RCPF in more than half of the intervals shown in Figure 14 when redispatch was necessary. In 2008, 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 $50 per MWh RCPF was not sufficiently high to maintain reserves in the local areas under normal operating conditions during 2008, leading the ISO to take additional actions to maintain reserves in a substantial number of intervals. In 2009, the marginal redispatch cost to maintain local reserves never exceeded $50 per MWh, although redispatch was done to maintain local reserves much less frequently in 2009 than in previous years.

The marginal cost of meeting the local reserve requirements rarely exceeded the new RCPF of $250 per MWh. In 2008, 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 new RCPF occurred May 8 and May 27, 2008. In these intervals, the local reserve requirements were satisfied using proxy second contingency constraints, which lead the real-time model to hold reserves on the transmission interfaces into the areas. 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 new RCPF should be sufficient to maintain local reserves when redispatch is necessary. However, the significance of the new RCPF is diminished because local reserve constraints have become very infrequent due to recent transmission upgrades in Boston and Connecticut. If local reserve constraints become more frequent in the future, the use of the new $250 per MWh RCPF will lead to more efficient redispatch and lower energy prices.

D. 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 primarily by modeling local reserve requirements in the real-time market software.

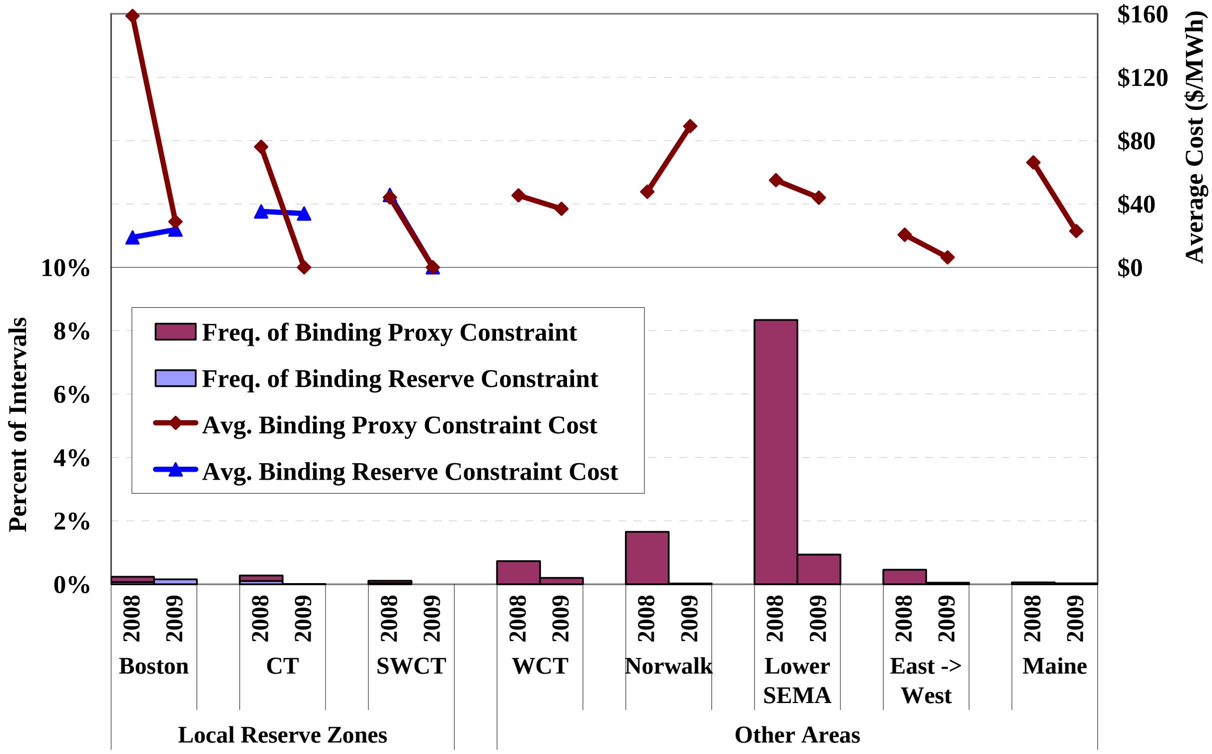
There are other local areas that require the commitment and dispatch of resources to meet local second contingency protection requirements. The ISO's operating procedure for security monitoring discusses the process for maintaining operating reserves in the following eight local areas: the three local reserve zones, as well as West Connecticut, Norwalk-Stamford, Southeast Massachusetts, Western New England, and Maine.[42] In areas requiring local operating reserves

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

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 2008 and 2009. 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.[43]

Figure 15: Summary of Real-Time Market Constraints to Maintain Local Reserves 2008 – 2009



The frequency of intervals when the ISO redispatched the system to meet local reserve requirements fell considerably from 2008 to 2009. In 2008, redispatch to maintain reserves in Norwalk-Stamford, West Connecticut, and Lower SEMA, which were not designated as local reserve zones, was more frequent than maintaining reserves in any of the three designated local

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

reserve zones. In 2009, redispatch to maintain reserves in Lower SEMA was most frequent, although transmission upgrades that were completed in July virtually eliminated the need to redispatch for local reserves.

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. 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, the ISO should 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. The importance of this recommendation is reduced by the fact that redispatch to maintain reserves in the eight local areas has become very infrequent. However, if local reserves needs result in more frequent redispatch in the future, having additional reserve zones defined will lead to more efficient redispatch and lower energy prices.

One could argue that expanding the number of real-time reserve zones would require the ISO to expand the set of local reserve zones that are modeled in the Forward Reserve Market Auction, but this would not be necessary. 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 adequately account for such differences.[44]

---

44 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

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.

E. Locational 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). 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: TMNSR and 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 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

---

clearing prices of TMSR and TMOR at their location. Such settlement rules provide suppliers with incentives to satisfy their Forward Reserve Obligations with higher quality reserve capacity when it is efficient to do so.

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.[45]

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 2009/10 procurement period (October 2009 to May 2010) was held in August 2009. 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.[46]
- For Rest of System (i.e., areas outside Connecticut and Boston), the effective TMOR requirement is 798 MW.[47]
- 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 like Forward Reserve Procurement Periods, adjusted for major changes in the topology of the system or the status of supply resources.

---

45 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.

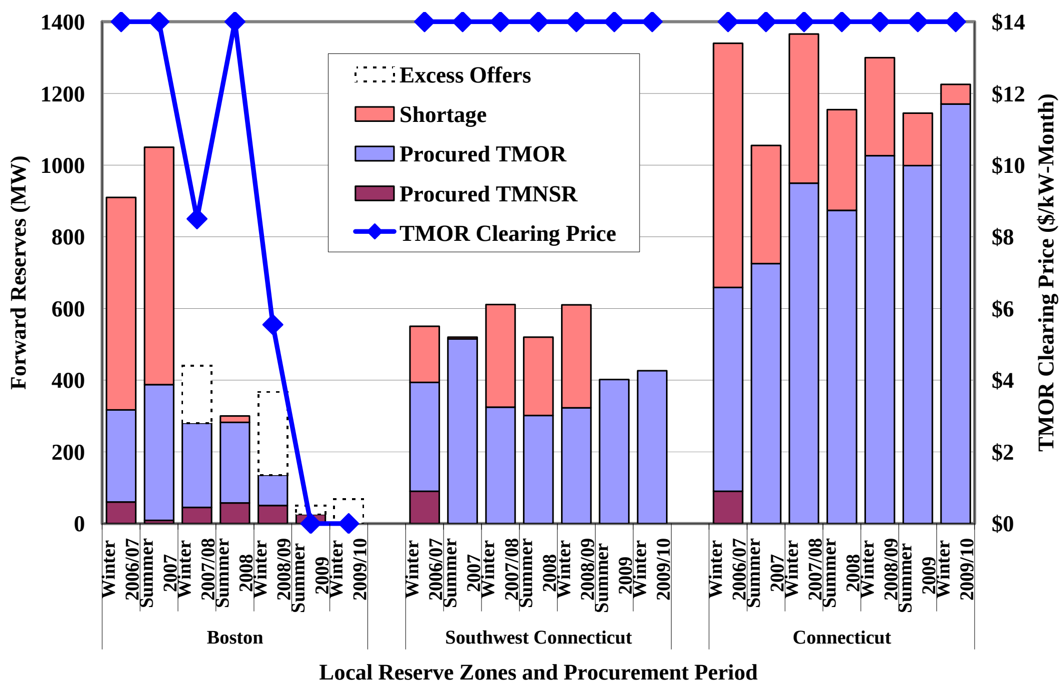
46 Usually, the forecasted largest contingency is the HQ Phase I/II Interconnection and the forecasted second largest contingency is the combination of the Mystic 8 and Mystic 9 generating units.

47 The requirement is 600 MW, although this is multiplied by 1.33 to account for the expected performance of off-line reserve providers.

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.[48]

The following two figures summarize the quantities purchased in the last seven 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 2010



48 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.

In Boston, the local reserve zone requirement was satisfied in four of the seven auctions. In the first two auctions, the local zone requirement for Boston was near 1 GW. However, a substantial amount of transmission capability was added into the Boston area in 2007, leading the ISO to assume between 900 and 1,610 MW of External Reserve Support in the last five auctions. External Reserve Support is the amount of the local reserve zone need that is assumed to be satisfied by the transmission capability into the zone, which reduces the amount that must be satisfied by internal resources. As a result, the amount of local reserves required from internal Boston resources was reduced to less than 300 MW in the last five auctions, including 0 MW (i.e., no need for local resources due to enough External Reserve Support) in the last two 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. In the Summer 2008 auction, the Boston requirement was not met due to a reduction in the quantity of offers from the previous summer procurement period. In the Winter 2008/09 auction, the volume of offers and the External Reserve Support increased, leading the clearing price to fall significantly. In the Summer 2009 and the Winter 2009/10 auctions, the quantity of External Reserve Support was sufficient to meet the local TMOR requirements, leading TMOR and TMNSR to clear at the same prices as in Rest-of-System.

In the seven local reserve procurements for Connecticut shown in Figure 16, the local reserve zone requirement was never satisfied, leading TMOR to clear at the $14 per kW-month price cap in each auction. The shortage quantities have declined over the past seven procurement periods, ranging from as high as 681 MW in the Winter 2006/07 to a low of 55 MW in the Winter 2009/10. The decline in shortage quantities has resulted from sales from new fast-start resources and increased participation by existing fast-start capacity.

The forward reserves procured for Southwest Connecticut are shown both separately and as a subset of the total procurement for Connecticut. Southwest Connecticut was short of forward reserves in the first five procurements shown in Figure 16 but not in the last two. This is due primarily to transmission upgrades into Southwest Connecticut that were brought into service in early 2009. Starting in the Summer 2009 auction, a large portion of the requirement for

Southwest Connecticut was satisfied by External Reserve Support.[49] Although the local requirement was satisfied in the past two auctions, TMOR still cleared at the price cap of $14 per kW-month due to the shortage in Connecticut.

Figure 16 shows that TMNSR has rarely 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-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. For each procurement period, Figure 17 shows the TMOR clearing price, the quantity of TMOR and

49 For Southwest Connecticut, all but 22 MW of the 673 MW requirement in the Summer 2009 auction and all of the 520 MW requirement in the Winter 2009/10 auction were met with External Reserve Support.

TMNSR procured, the shortage quantity if the requirement was not met, and the quantity of excess offers if the requirement was met.

Outside of the local reserve areas, the forward reserve requirements were satisfied in each auction. In all seven 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). Likewise, the Rest of System TMOR price cleared at $0 in the Summer 2009 and the Winter 2009/10 auctions because the requirements were met by procurement for the TMNSR requirements. In the Winter 2006/07 auction, TMNSR and TMOR sold at the same price because the TMNSR requirement was met by the combination of procurement for local areas and the Rest of System TMOR requirement. However, in the subsequent six auctions, TMNSR cleared at a premium over TMOR.

Figure 17: Summary of Forward Reserve Auction for Outside Local Areas Procurement for October 2006 to May 2010

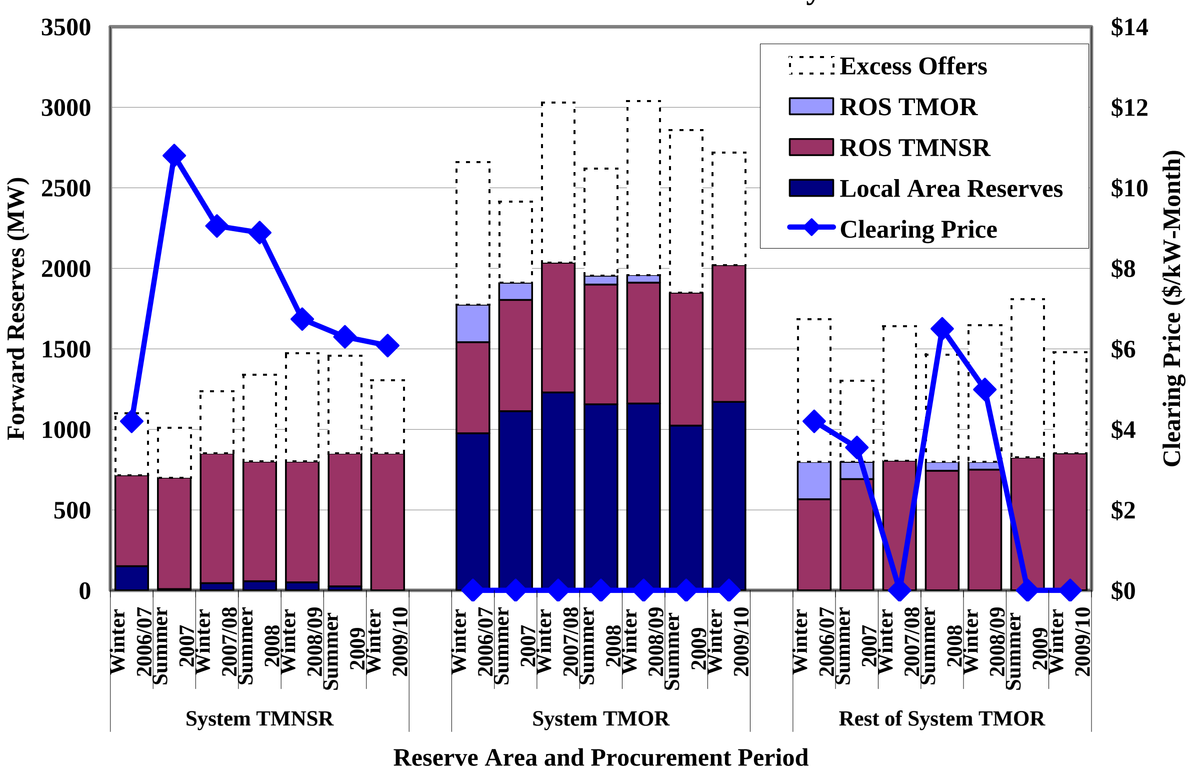


Figure 17 shows that a large share of the TMNSR requirement was procured outside of the local areas. For example, just 25 MW of TMNSR was procured in the local areas in the Summer 2009 auction and none was procured in the Winter 2009/10 auction, even though approximately 275 MW of TMNSR-capable fast-start 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.[50] 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

50 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 $6 per MMbtu, it would result in a Threshold Price of approximately $86 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.

non-affiliated firms was very limited in 2009. Suppliers that do not meet their forward reserve obligations incur a Failure to Reserve Penalty.[51]

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.[52] 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, 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 foregone reserve revenues are the same 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 $100 per MWh and a supplier assigns a generator that has incremental costs of $60 per MWh to provide forward reserves. Because the supplier is required to offer at $100 per MWh, the supplier will not be scheduled to sell energy when the LMP is between $60 per MWh and $100 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 unit that is not a fast-start unit, a supplier may have to commit a

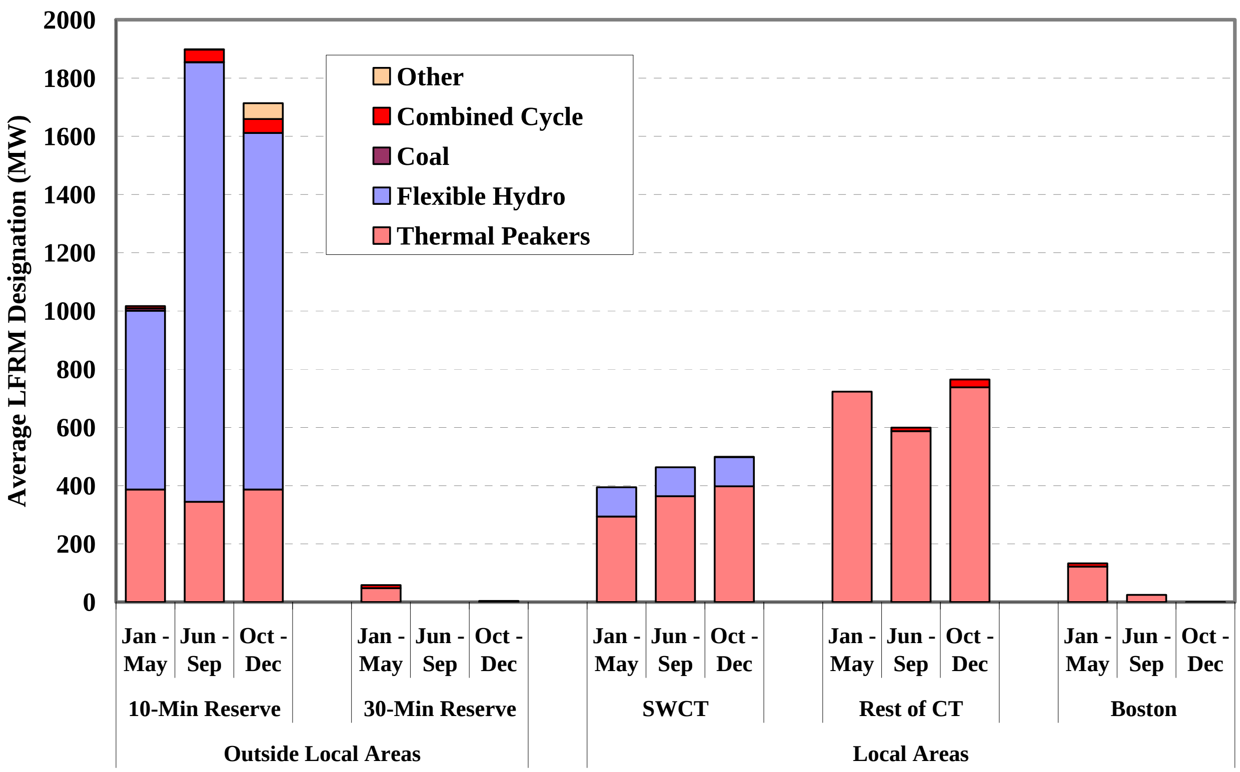
51 The Failure to Reserve penalty is equal to the number of megawatts not reserved times 1.5 times the Forward Reserve Payment Rate, which is the forward reserve clearing price (adjusted for capacity payments) divided by the number of obligation hours in the month.

52 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.

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 2009 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
2009



Approximately 97 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 slower-starting resources.

Combined cycle units were assigned to provide a small portion of the forward reserves in 2009. 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 2009 were ones that are capable of providing offline reserves within 30 minutes.

The average quantity of forward reserve obligations satisfied by coal-fired steam units fell from an average of 19 MW in 2008 to 4 MW in 2009. Coal units have two characteristics that can make them relatively efficient providers of forward reserves under certain market conditions. First, 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. However, some suppliers may not be comfortable offering this range from their coal-fired resources. Second, it is frequently economic to commit coal-fired units so suppliers do not face significant costs from committing them uneconomically. However, this factor is mitigated when natural gas prices fall to relatively low levels, as they did in 2009.

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 slower-starting resources do not sell forward reserves because the expected costs of providing forward reserves exceed the price cap. However, slower-starting 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 units that do not have 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.

Some had expected 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.

F. 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 30 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 five years: from 143 MW in 2005 to 134 MW in 2006, 129 MW in 2007, 121 MW in 2008, and 89 MW in 2009.

In this report, we evaluate two aspects of the market for regulation. Section 1 reviews the overall expenses from procuring regulation. Section 2 explains how regulation providers are selected and examines the pattern of supply offers from regulation providers. The end of this section summarizes our conclusions and recommendations related to the regulation market.

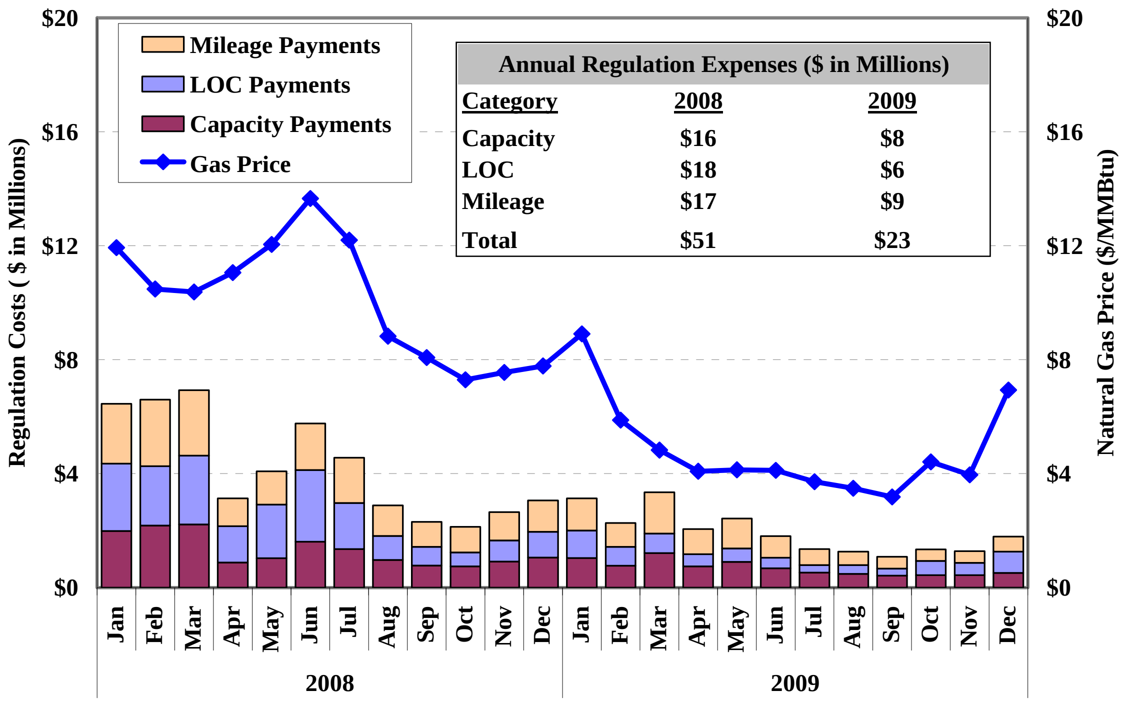
1. Regulation Market Expenses

Resources providing regulation service receive the following payments:[53]

- *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 2008 and 2009. The figure also shows the monthly average natural gas prices.

Figure 19: Regulation Market Expenses
2008 – 2009



53 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."

This figure shows that each category of expenses accounts for approximately one-third of total regulation expenses. Total regulation expenses declined 55 percent from 2008 to 2009. This decrease is primarily due to the comparable decrease in natural gas prices from 2008 to 2009. Accordingly, the figure shows that variations in monthly regulation market expenses were correlated with changes in the monthly average natural gas price.

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 more frequently during periods of low gas prices. This increases the availability of low-priced regulation offers and leads to lower regulation expenses. Third, lower fuel prices normally reduce the opportunity costs for units to provide regulation service, which is consistent with the $12 million decrease in regulation opportunity cost expenses from 2008 to 2009.

Changes in natural gas prices and commitment patterns led to changes in offer patterns that explain much of the fluctuations in regulation market expenses in 2008 and 2009. Offer patterns are examined in more detail in the following section.

2. 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 Mileage Payment* – This is equal to the estimated capacity payment.
- *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.
- *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.

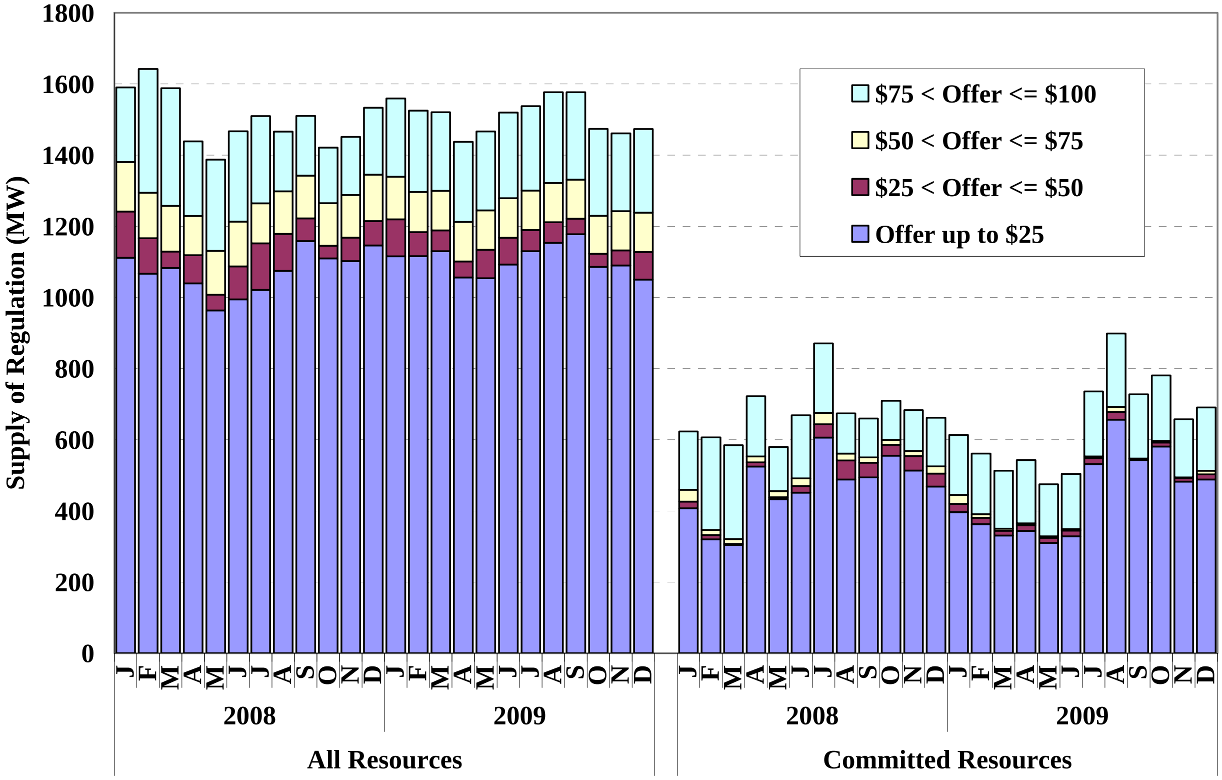
The ranking process iterates until the set of resources selected to provide regulation does not change for two consecutive iterations.[54]

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 2008 and 2009 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 the regulation offer prices and quantities over the past two years were relatively consistent throughout the period. The quantities of total regulation offers varied typically between 1,400 MW and 1,600 MW in each month, averaging approximately 1,500 MW in both 2008 and 2009. The portion of regulation offers in each price range was also consistent over the past two years. On average, almost three quarters of of the regulation offers were below $25 per MWh, 5 percent were between $25 and $50 per MWh, 7 percent were between $50 and $75 per MWh, and 15 percent were more than $75 per MWh.

54 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
2008 – 2009



The right panel shows the changes in offer quantities and prices that more directly determine market outcomes since only offers from committed resources can be selected. On average, approximately 45 percent 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 increased substantially in the last six months of 2009 for at least two reasons. First, very low natural gas prices led to more commitment of combined cycle units, which provide most of the regulation capability in New England. Second, the substantial reduction in supplemental commitment of slow-ramping steam units for reliability also led to more commitment of combined cycle units.

During 2008 and 2009, 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, supply is sometimes 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.

3. Future Potential Changes

The ISO launched a pilot program in November 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.[55] 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 14 percent of the average hourly regulating requirement in 2009. 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.

Finally, although the current market performed relatively well in 2009, the ISO may wish to consider co-optimizing the regulation market with the energy and operating reserve markets. Given the complex interaction of the regulation market with the energy market, particularly with

55 See "Appendix J – Alternative Technologies Regulation Pilot Program" of "Market Rule 1" for detail.

respect to commitment decisions made in the day-ahead market, co-optimizing these markets would improve the scheduling, commitment and pricing in these markets.

G. 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 RTO to determine the amount of reserves that are held on resources inside a local area versus the amount of reserves that are imported to the area in the real-time market, reducing the overall cost of meeting the local reserve requirements. During reserve shortages, the real-time market sets efficient LMPs and reserve clearing prices according to the RCPFs.

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 97 percent of the resources assigned to satisfy forward reserve obligations in 2009 were fast-start resources capable of providing offline reserves.

Overall, the regulation market performed competitively in both 2008 and 2009. On average, approximately 650 MW of available supply competes to provide less than 100 MW of regulation service. The significant excess supply generally limits competitive concerns in the regulation market.

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

- Although local reserves were rarely constrained in 2009, the cost of maintaining operating reserves in local areas frequently exceeded the $50 per MWh local RCPF in previous years. Accordingly, the RCPF was increased to $250 per MWh in January 2010 to reduce the cost of real-time dispatch and to more accurately reflect the cost of maintaining local reserves in real-time prices.
- 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.

- ✓ Although local requirements in these areas were binding relatively infrequently in 2009, efficient modeling will become more beneficial if they bind more frequently in the future.
- ✓ Therefore, the ISO should consider creating additional local reserve zones in areas that are currently managed using proxy second contingency transmission limits if these constraints bind more frequently.

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.
  - ✓ However, our analysis of operating reserves, congestion, and forward capacity all indicate a substantial surplus in these areas.
- 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, which is not necessary 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.
- The ISO may wish to consider the long-term viability of the forward reserve market because:
  - ✓ It has not achieved one of its primary objectives (to lower NCPC by purchasing forward reserves from high-cost units frequently committed for reliability);

- ✓ It has produced price signals that are not consistent with the prevailing surpluses in the local areas (although this will be resolved if the external reserve support for the local areas continues to rise to reflect the new transmission investment); and
- ✓ It may be largely redundant with the locational requirement in FCM.

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

- Although the current market performed relatively well in 2009, the ISO may wish to consider co-optimizing the regulation market with the energy and operating reserve markets.
  - ✓ Given the complex interaction of the regulation market with the energy market, particularly with respect to commitment decisions made in the day-ahead market, co-optimizing these markets would improve the scheduling, commitment and pricing in these markets.

VI. 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 2009. 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.

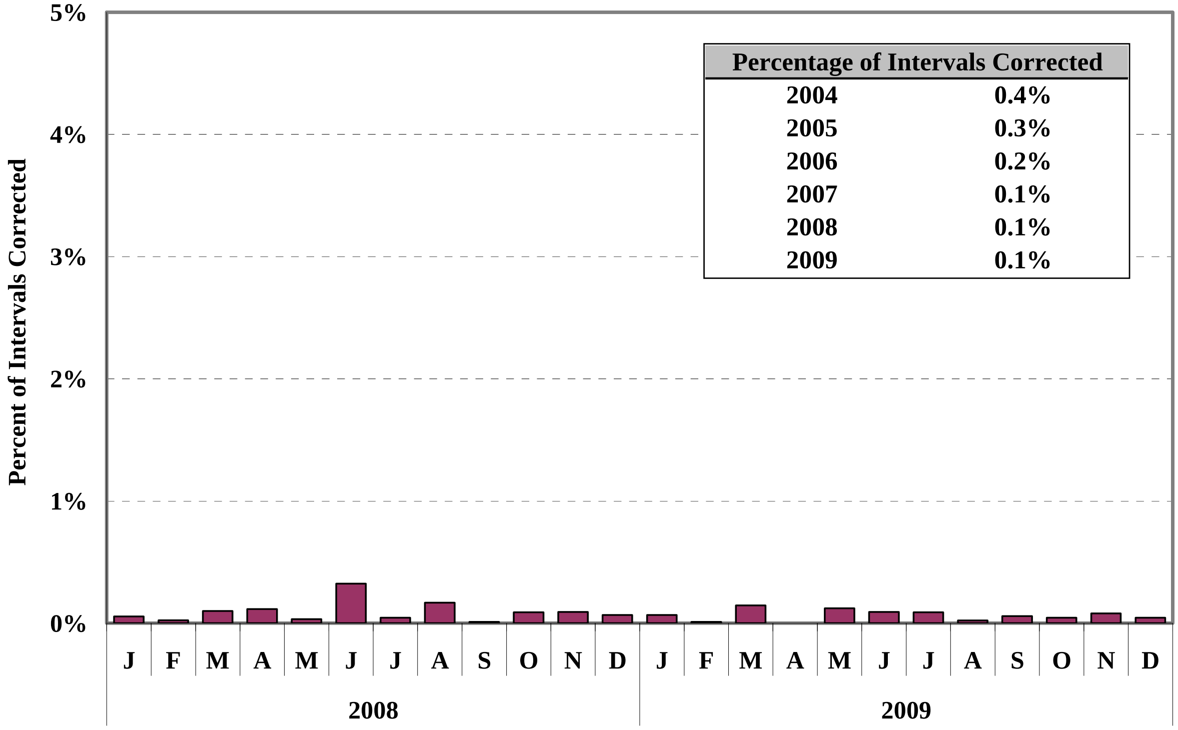
It is particularly important for the market to set efficient real-time prices and dispatch signals during shortages of operating reserves. ISO New England uses RCPFs to set real-time clearing prices during operating reserves shortages. This is discussed in greater detail in Section V of this report, which evaluates the reserve markets.

A. Real-Time Price Corrections

This subsection evaluates the rate of real-time price corrections during 2009. 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, 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 2008 and 2009. The inset table summarizes the annual rate of price corrections in the past six years.

Figure 21: Rate of Real-Time Price Corrections 2008 – 2009



Percentage of Intervals Corrected	
2004	0.4%
2005	0.3%
2006	0.2%
2007	0.1%
2008	0.1%
2009	0.1%

The figure shows that real-time price corrections were infrequent in both 2008 and 2009. 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 steadily over the past six years. It is notable that approximately 63 percent of the intervals that experienced price corrections in 2009 were due to issues with the real-time software's Dead Bus Logic, which affects the LMPs at very few pricing nodes.[56] 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 six 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 notification. This enables them to provide valuable offline reserves. 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 challenges 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 2009.

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

The ISO's real-time dispatch software, called Unit Dispatch System ("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

---

56 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 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.

instructions to each generator for the next dispatch interval based on a short-term forecast of load and other operating conditions.[57] Most commitment decisions are made in the day-ahead timeframe prior to the operation of UDS. UDS' primary function 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.[58] It is more efficient to allow UDS to start fast-start generators than to rely exclusively on operators to manually commit such units.[59]

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.[60] 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.

---

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

58 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.

59 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.

60 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).

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-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 (i.e., 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 even 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 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 generator 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 eligible to set the LMP since it is at its minimum output level, 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 total offer and the real-time market revenue, resulting in additional uplift charges to the market.

2. Evaluation of Fast-Start Deployments by UDS in 2009

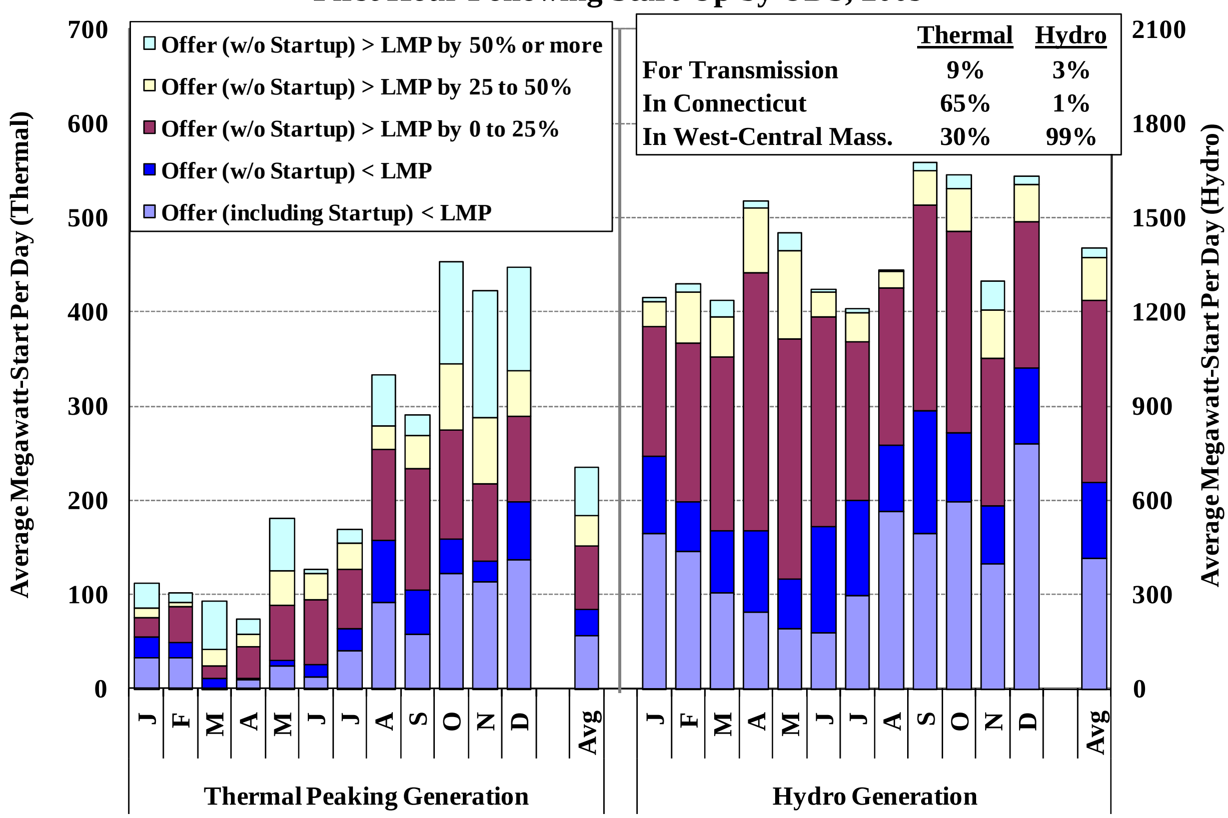
The following two analyses assess the efficiency of real-time pricing during periods when fast-start units were deployed in merit order in 2009. The first figure shows the quantities of fast-start capacity deployed economically by UDS on average each day. It also examines the extent to which the real-time LMP revenues that such units receive are consistent with their total offers. The second figure evaluates how real-time prices would be affected if the average total offers were fully reflected in real-time LMPs.

Figure 22 summarizes the consistency of the average total offer[61] of fast-start generators that were deployed economically by UDS with the average real-time LMP over the initial

61 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.

commitment period, which is 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 (including Startup) < 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 four categories that exclude the start-up component of the offer. These categories are shown according to the size of the difference between the average total offer and the average real-time LMP. This comparison is shown separately for hydro and thermal peaking generators in each month of 2009. The table in the figure shows the shares of peaking capacity that were started in Connecticut and in West-Central Massachusetts. The table also shows the share of peaking capacity that was started in import-constrained areas.[62]

Figure 22: Comparison of Real-Time LMPs to Offers of Fast-Start Generators First Hour Following Start-Up by UDS, 2009



62 Capacity is included in the "For Transmission" category if the congestion component of the LMP at the peaking unit's node is greater than the congestion component at New England Hub by $1 per MWh or more.

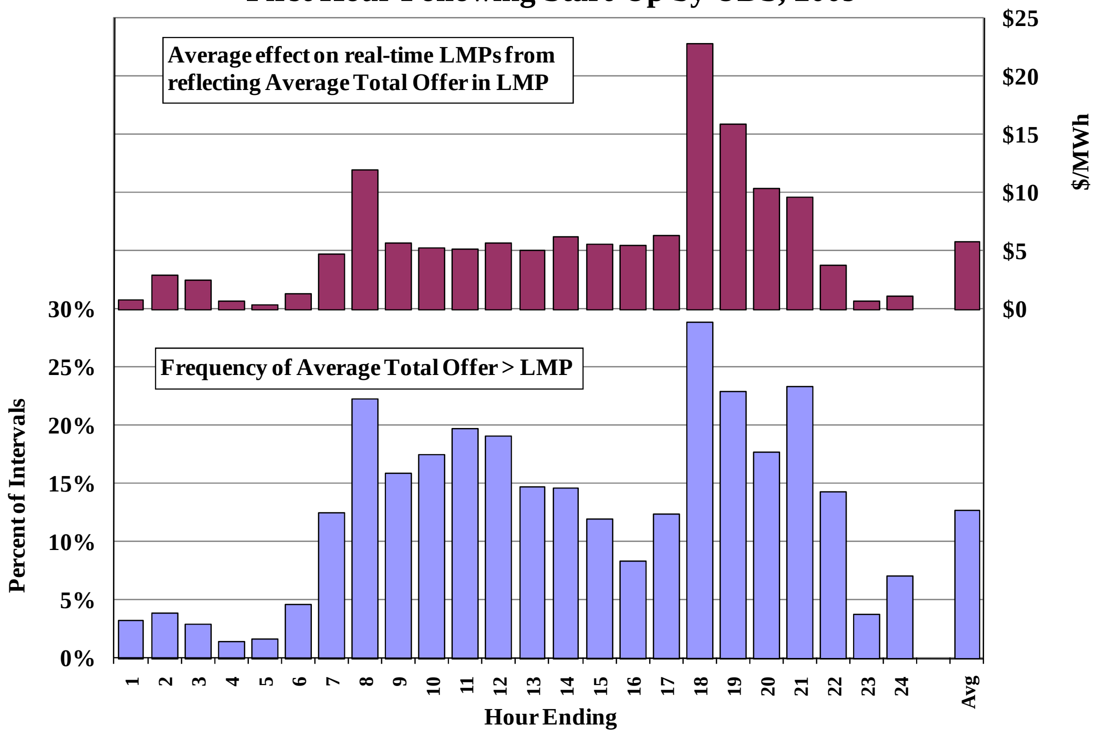
The figure shows that flexible hydro generation accounted for the majority of fast-start generation that was started by UDS in 2009. This was primarily associated with 1,600 MW of pumped storage capacity located in West-Central Massachusetts. The share of starts where the average total offer was higher than the real-time LMP was 70 percent in 2009. This ratio is lower for thermal peaking resources as discussed below, which is likely due to the fact that the thermal peaking generators are generally less flexible than hydro generators.

Figure 22 shows that roughly 235 MW of thermal peaking generation were started each day, 76 percent of which exhibited a total offer (including start-up costs) greater than the average real-time LMP. Even when start-up costs are excluded, 64 percent of the thermal peaking generation started exhibited offers that exceeded the average LMP over the minimum run time. Thermal peaking generators are deployed in a relatively limited set of hours. Nonetheless, they are frequently the marginal source of supply to the system in the hours that they run, making it particularly important to reflect the full cost of their deployment in real-time LMPs.

Figure 22 shows that the average amount of thermal peaking capacity that was started by UDS increased considerably in the last five months of 2009, from 123 MW per day in January to July to 389 MW per day in August to December. The increase in commitment of thermal peaking units occurred as the surplus capacity online in real-time decreased in the second half of 2009. As detailed in Section VII of the report, surplus online capacity fell in 2009 because the amount of capacity committed after the day-ahead market (i.e., supplemental commitment) to satisfy local reliability requirements fell sharply following several major transmission upgrades in Southeast Massachusetts and Connecticut. If the amount of surplus capacity remains low in the future, fast-start units will continue to play an increased role in satisfying real-time demands.

The previous figure shows that the full costs of the thermal peaking units are frequently higher than real-time LMPs. This indicates that fast-start units that are committed in economic merit order usually rely on NCPC payments to recoup their full offer costs. More importantly, it indicates that real-time prices do not accurately reflect the marginal cost of serving real-time demand, which affects the economic signals provided by the day-ahead and forward markets in New England. The following analysis examines how real-time energy prices would be affected if the average total offers of such units were reflected in real-time LMPs.[63] The analysis summarizes the portion of the fast-start units' costs that were not fully reflected in real-time LMPs in 2009. The lower portion of Figure 23 shows how frequently thermal and hydro fast-start units were started economically by UDS and their average total offers were greater than the LMP during the minimum run time in 2009.[64] The figure excludes fast-start units that were started in import-constrained areas, since the LMP of the fast-start unit during such events would be representative of only a limited area of New England.[65] The upper portion of the figure shows the difference between the average total offer and the real-time LMP from such periods averaged over the year by time of day.

Figure 23: Difference Between Real-Time LMPs and Offers of Fast-Start Generators First Hour Following Start-Up by UDS, 2009



63 If a gas turbine from the earlier example was started with a total offer of $115/MWh when the LMP was $75/MWh, this analysis would assume the unit would increase the LMP by $40 per MWh. Other lower-cost gas turbine started in the same hour would not affect prices because they are inframarginal.

64 If multiple fast-start units are started at one time, the analysis uses the one with the largest difference between the average total offer and the real-time LMP, which is usually the highest-cost unit.

65 The area is treated as import-constrained if the congestion component of the LMP at the fast-start unit's node is greater than the congestion component at New England Hub by $1 per MWh or more.

Figure 23 shows that fast-start units were deployed economically by UDS when their average total offer was greater than the real-time LMP in a substantial portion of hours. Such hours were most frequent from hours-ending 7 to 22, particularly around the morning peak (hours-ending 8 to 12) when load picks up rapidly and the evening peak (hours-ending 18 to 21). Ramp demands are highest on the system during these periods so fast-start generation is sometimes needed to meet these demands.

Overall, Figure 23 shows that fast-start units were started economically by UDS when their average total offer exceeded the real-time LMP over the minimum run time in 13 percent of all hours, and as high as 29 percent in hour 18. It also shows that if the average total offers were fully reflected in the energy price in these hours, the average real-time LMP would increase approximately $6 per MWh. The effect would be largest in hour-ending 18 when the average LMP would rise by $23 per MWh. These values are likely overstated, however, because they ignore the likely market responses to the higher real-time prices:

- Incentives to purchase more in the day-ahead market would increase, which would increase the amount of lower-cost generation committed in the day-ahead market.
- Net imports would increase from neighboring control areas, particularly New York.

Hence, the actual effect on real-time LMPs from more efficient pricing during fast-start deployments would be smaller than the effects reported in Figure 23, which does not consider the market response to more efficient real-time prices. Importantly, these responses would substantially improve efficiency because higher-cost peaking generation would be displaced by lower-cost intermediate generation and lower-cost imports. In addition, the changes in the market's economic signals that would result by causing peaking resources to set prices more reliably when they are marginal would improve economic efficiency by facilitating more efficient contracting and investment over 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.[66] 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 constraint is "relaxed", which is discussed later in this section.

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

---

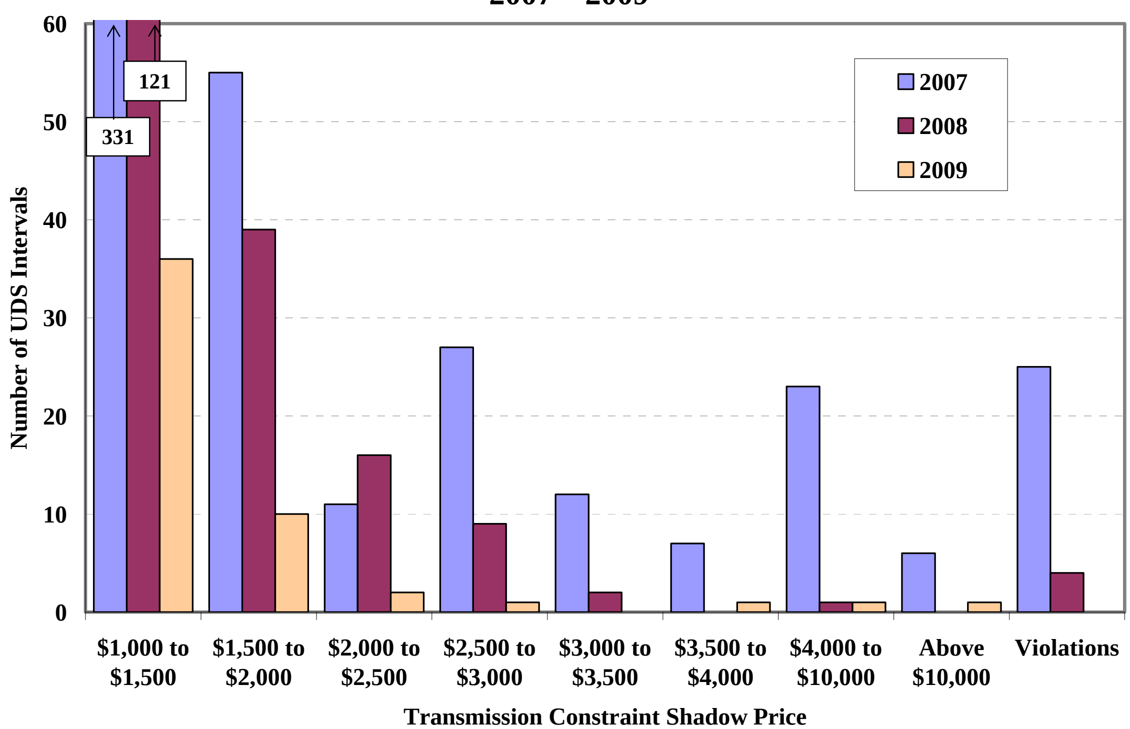
66 UDS will not use redispatch options that exceed the level of the Transmission Constraint Penalty Factor, although the penalty factor is ordinarily set to a very high level. Also, UDS will not redispatch resources that have a sensitivity factor with a magnitude of less than 2 percent relative to the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

- 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 may necessitate immediate curtailment of firm load to maintain reliability.[67] 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 24 illustrates the significance of periods when transmission congestion was particularly acute from 2007 to 2009.

Figure 24: Frequency of UDS Intervals with High Redispatch Costs
2007 – 2009



67 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.

The figure shows that the frequency and the cost of acute congestion have declined substantially since 2007. 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 and 52 such UDS intervals in 2008 and 2009, respectively, down approximately 61 percent and 90 percent from 2007.

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. Instead of using the relaxation algorithm to set LMPs, LMPs could be based on the shadow price of the violated constraint, which would equal the constraint penalty factor. This would require setting the transmission constraint penalty factors to lower levels that reflect the reasonable 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.

Using lower transmission constraint penalty factors would also be beneficial in some cases when extremely costly redispatch is not warranted for reliability. For example, Figure 24 shows several intervals when the transmission limit was resolved with a shadow price exceeding $10,000 per MWh, resulting in LMPs that exceeded $5,000 per MWh at several nodes. 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. For context, we would note that the New York ISO uses $4,000 per MWh for an analogous purpose, the Midwest ISO generally uses $2,000 to $3,000, and the Southwest Power Pool uses $2,000.

In a previous report, we evaluated how lower transmission constraint penalty factors would have affected flows in the intervals in 2007 when the shadow price exceeded $10,000 per MWh. We found that if a much lower transmission constraint penalty factor had been used, it would have had a relatively small impact on the flows over the constrained facility.[68] For example, in the most acute interval, lowering the transmission constraint penalty factor to:

- $6,000 per MWh would have led to a 1 percent violation of the limit,
- $3,000 per MWh would have led to a 6 percent violation of the limit, and
- $1,000 per MWh would have led to a 9 percent violation of the limit.

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 modest violations that would not likely 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 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.[69] This approach would also safeguard the market from the current algorithm which can cause violated constraints to be substantially under-priced. Hence, we believe the ISO should evaluate potential enhancements to the current

68 See, *e.g*., “2008 Assessment,” pages 74-76.

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

operating procedures that would establish reasonable Transmission Constraint Penalty Factors and allow them to set LMPs when a constraint is in violation. Given the extremely low levels of congestion recently in New England and the fact that no constraints were violated in 2009, the benefits from such a change will be limited in the near term. For that reason, we have not made this a recommendation in this report.

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 three Forward Capacity Auctions was composed of demand response capability rather 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:[70]

- 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.

70 The current Demand Response Programs are set to expire on May 31, 2012. The ISO is working with stakeholders to develop new demand response programs that will allow resources to be paid for being price-responsive.

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

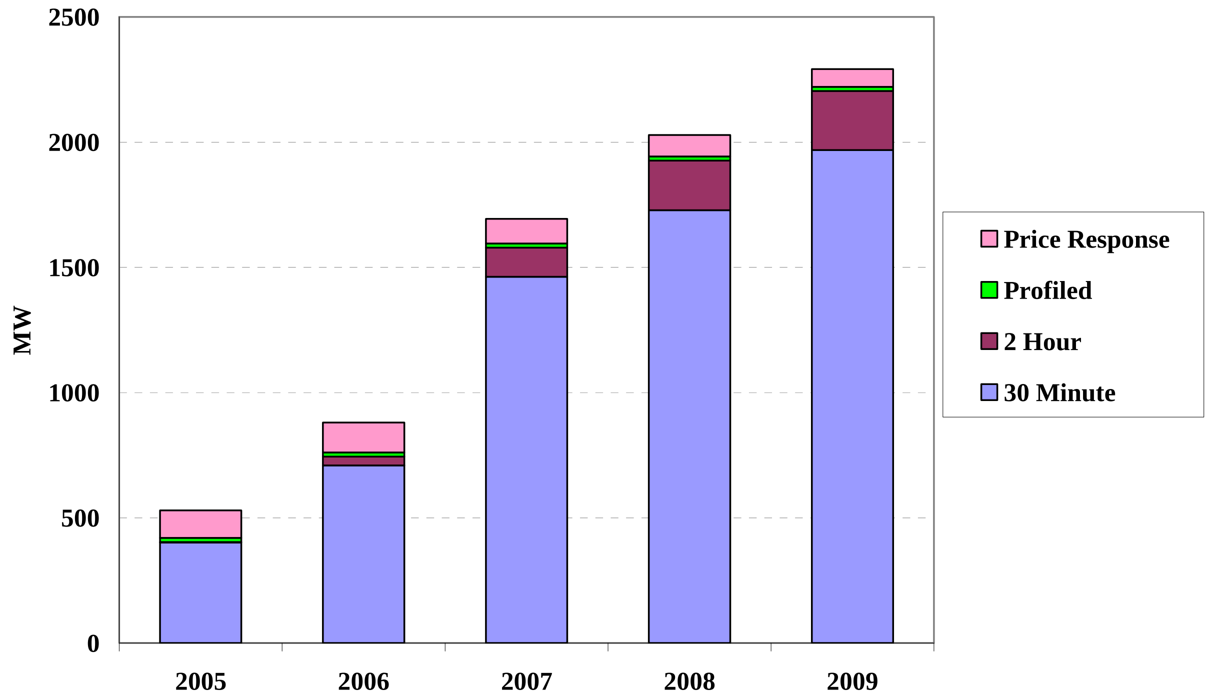
Demand response participation has surged in New England in recent years. Figure 25 shows the quantity of resources enrolled in each of the four demand response programs from 2005 to 2009.[72]

The quantity of enrolled resources increased substantially from 530 MW in 2005 to 2,292 MW in 2009. Most of demand response resources (86 percent) were enrolled in the Real-Time 30-Minute Demand Response Program by the end of 2009, 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 five years, increasing from 2 MW in 2005 to nearly 240 MW in 2009. The enrollment in the Real-Time Price Response Program decreased over the period, from 110 MW in 2005 to 71 MW in 2009.

71 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.

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

Figure 25: Real-Time Demand Response Program Enrollments 2005 – 2009



Initially, the programs described above were set to expire on June 1, 2010, at the start of the first Capacity Commitment Period under the FCM. However, new demand response programs have not been fully designed and implemented, so the expiration date for the current programs was extended to June 1, 2012. 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.[73] More than 2,700 MW of demand resources sold in the first three Forward Capacity Auctions ("FCAs"). Hence, FCM is serving to encourage the rapid development of demand response resources in New England. These FCM results are discussed in detail in Section IX.

73 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.

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.[74,75] These inflexibilities can raise efficiency issues regarding real-time pricing.

In particular, the activation of demand response can depress real-time prices substantially below the marginal cost of the foregone consumption by the demand response resources. In 2008 and 2009, 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 development of new demand response programs.

Resources were activated relatively frequently under the Real-Time Price Response Program in 2009. 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 2009, the average real-time clearing price was substantially lower than the average cost of activating these resources. Of the 354 hours when these resources were activated, the clearing price at the New England Hub was less than $100 per MWh in 91 percent of the hours and less than $70 per MWh in 71 percent of the hours.

74 “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.

75 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.

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.

The current real-time emergency demand response programs will expire on June 1, 2012, so it will be important to address these real-time pricing issues in the development of new demand response programs. Because non-dispatchable 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, it will be important to have efficient mechanisms for setting real-time clearing prices when demand response is activated.

3. Conclusions and Recommendations

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.

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.

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.[76]

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[77] 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

76 This assumes the generators are offered at marginal cost.

77 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.

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 (0.3 percent) 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.

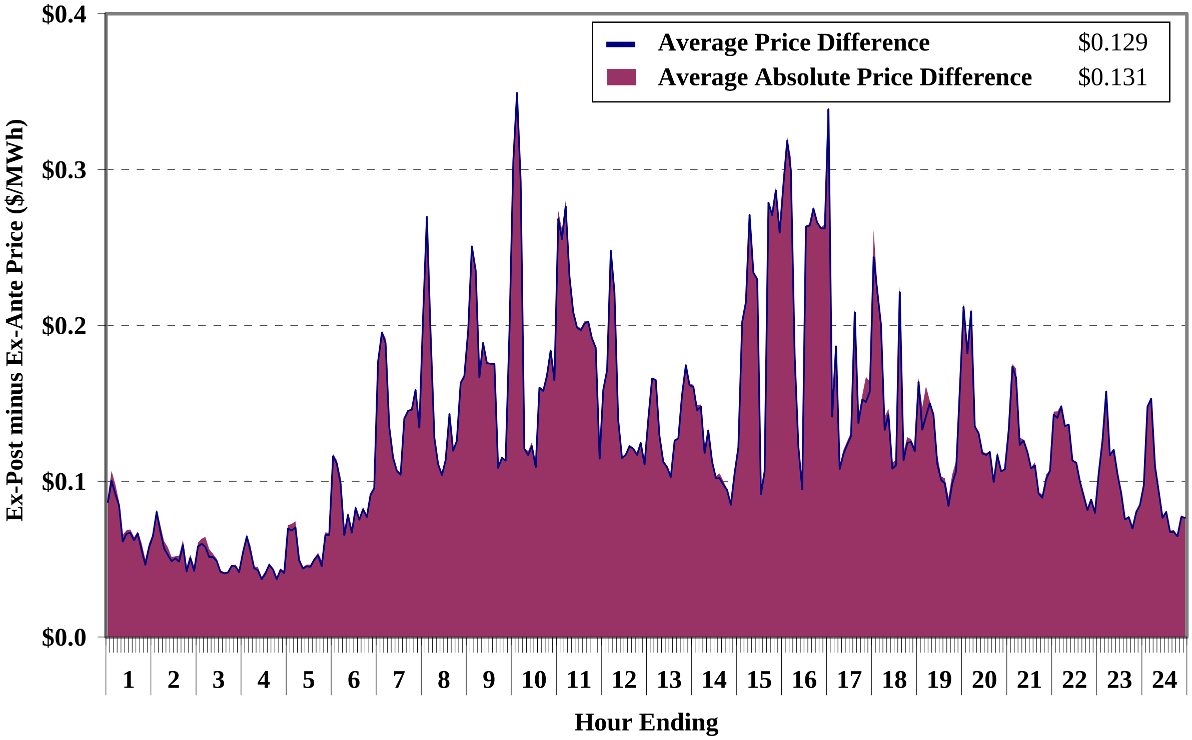
1. Persistent Differences Between Ex Ante and Ex Post Prices

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 2009 at a location close to the New England Hub.[78] 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 of day.

The average differences between the ex post and ex ante prices were relatively small in 2009. 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 $0.13 per MWh higher than ex ante prices at this location in 2009, down from $0.32 per MWh in 2008.

78 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.

Figure 26: Average Difference Between Five-Minute Ex Post and Ex Ante Prices 2009



The persistent bias shown in Figure 26 results from a combination 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, which are discussed in Part 3 of this subsection, 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. Due to the specific implementation in New England, this theoretical concern is rarely manifested.

3. Ex Post Pricing 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[79] in the ex post pricing.

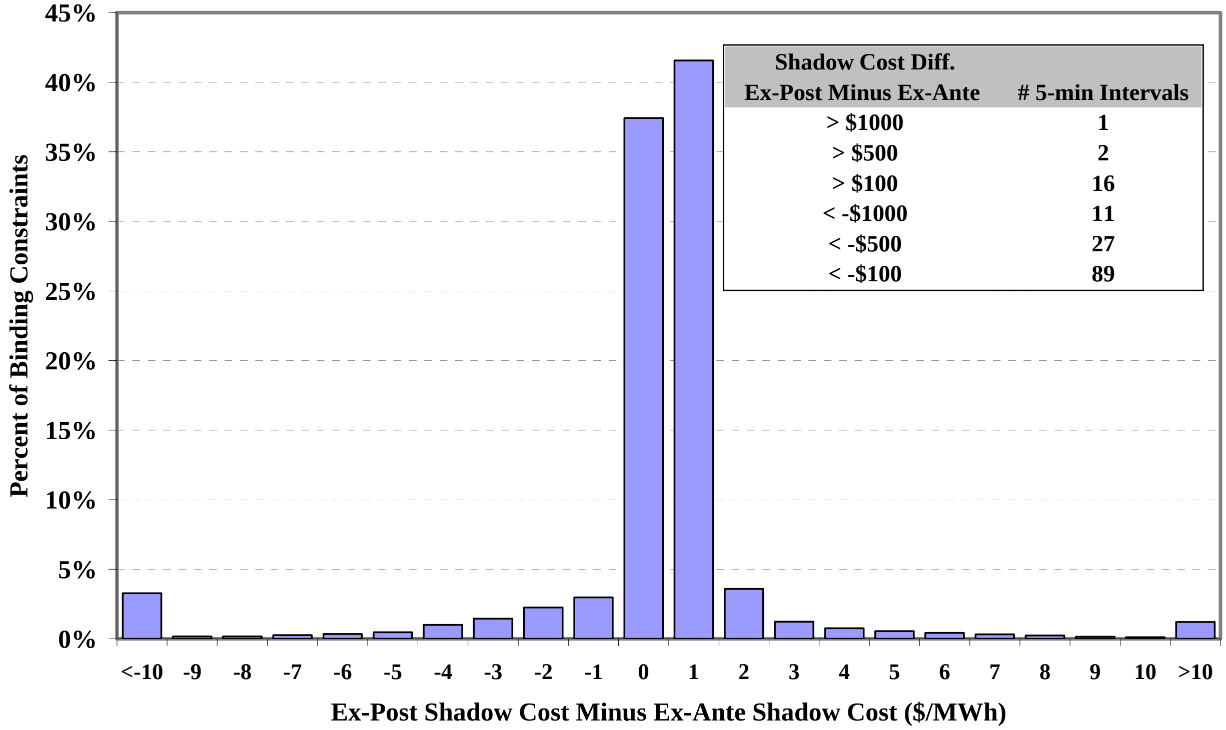
For example, suppose a combustion turbine with an incremental offer of $150 per MWh and an amortized start-up and no-load cost of $100 per MWh is started in order to resolve a load pocket constraint. Suppose that there is also a $50 per MWh unit in the load pocket that is dispatched at its maximum level. 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 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 2009. 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.

---

79 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.

Figure 27: Difference in Constraint Shadow Costs Between Ex Post and Ex Ante All Binding Constraints, 2009



Shadow Cost Diff. Ex-Post Minus Ex-Ante	# 5-min Intervals
> $1000	1
> $500	2
> $100	16
< -$1000	11
< -$500	27
< -$100	89

The average difference was not significant in 2009. Nearly 96 percent of all differences were within $10 per MWh and almost 90 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 19 intervals during which ex post shadow prices were at least $100 per MWh higher than ex ante prices, and 127 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 ex post pricing:

- 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 2009. Our evaluation leads to the following conclusions and recommendations:

- Price corrections were very infrequent in 2009, which reduces uncertainty for market participants in the New England wholesale market. Further, a large share of the price corrections that did occur affected a very small number of pricing nodes.
- Fast-start generators are routinely deployed economically, but the resulting costs are often not fully reflected in real-time prices. This leads to inefficiently low real-time 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 congestion management process does not always result in efficient prices when transmission constraints are in violation and sometimes causes the ISO to incur 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.
- Demand response programs help 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 allow the costs of non-dispatchable demand response resources to be reflected in clearing prices when there is a capacity deficiency or when a deficiency is avoided by the activation of the 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.

VII. System Operations

To maintain the reliability of the system, sufficient resources must 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 largely satisfy these requirements in a manner that is economically 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 and Out-of-Merit Generation – Reliability commitments make additional resources available to operate in real time, and they increase the amount of generation that runs out-of-merit in real time;
- Surplus Generation – The amount of capacity from online or fast-start resources in excess of the system's energy and operating reserve requirements; and

- Uplift Expenses – This examines the financial charges that result from out-of-market commitment and reliability agreements.

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).

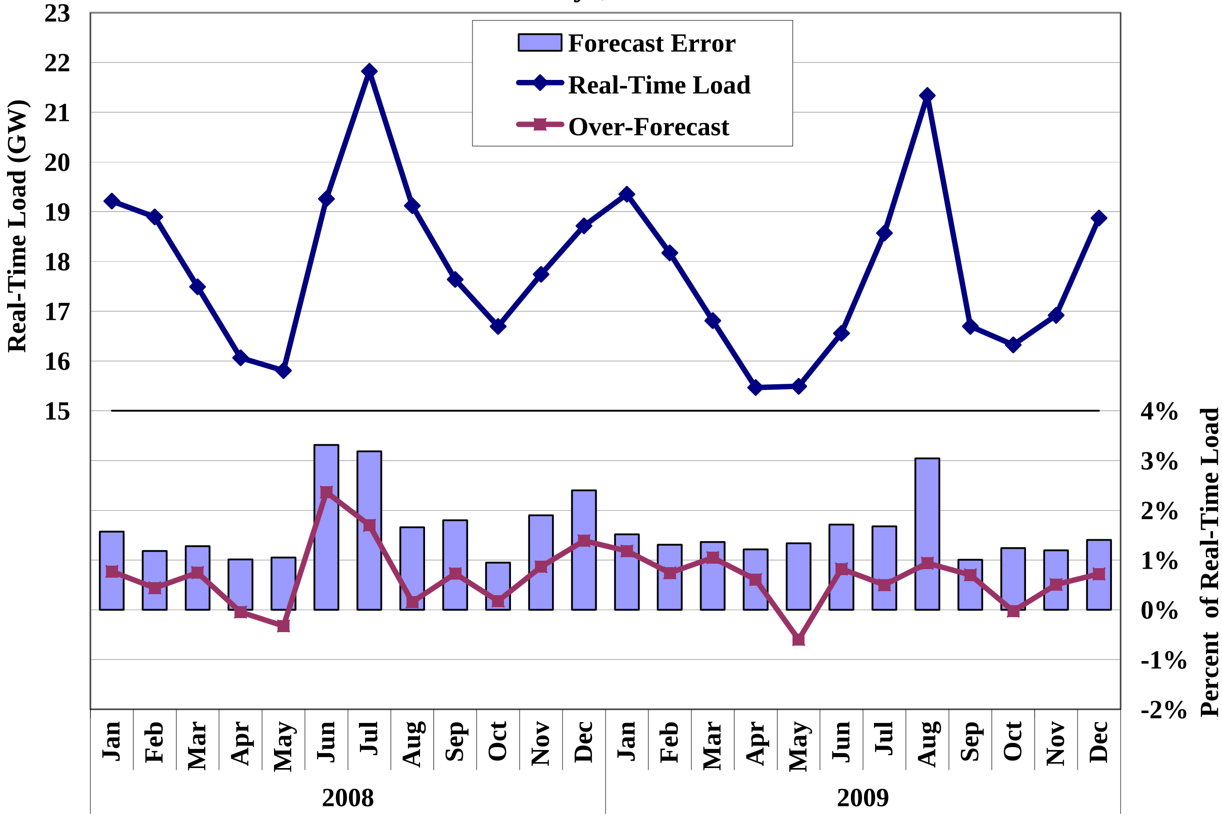
Accurate load forecasts promote efficient scheduling and unit commitment. Inaccurate load forecasts can cause the day-ahead market and/or the ISO to commit too much or too little capacity, which can affect prices and uplift. Therefore, it is desirable for the day-ahead forecast to accurately predict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

Figure 28 summarizes daily peak loads and two measures of forecast error on a monthly basis during 2008 and 2009. The *Over-Forecast* is the percentage by which the average day-ahead forecasted daily peak load exceeded the average real-time daily peak load[80] 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

80 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.

day-ahead forecasted daily peak load and the actual daily peak load, expressed as a percentage of the average actual daily peak load.

Figure 28: Average Daily Peak Forecast Load and Actual Load Weekdays, 2008 – 2009



The figure shows a characteristic pattern of high loads during the winter and summer and mild loads during the spring and fall. Overall, loads were significantly lower in 2009 than in prior years. The annual peak load of 25.1 GW occurred on August 18, 2009, down 4 percent from 26.1 GW in 2008. The average load also declined 4 percent, from 15.2 GW in 2008 to 14.6 GW in 2009.

The ISO's day-ahead load forecasts are highly consistent with actual load. However, the ISO had a tendency to over-forecast load. In 2009, only May and October showed that load was under-forecasted on average. The average over-forecast was comparable in the two years -- 0.7

percent in 2008 and 0.6 percent in 2009. The ISO regularly evaluates the performance of its load forecasting models to ensure there are no factors that bias the forecast unjustifiably.[81]

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 2008 to 2009, the average forecast error improved slightly from 1.8 percent to 1.5 percent. The improvement was more evident during the summer months when the forecast error declined from 2.7 percent in 2008 to 2.1 percent in 2009. However, these levels of forecast error are relatively small. Hence, we find that the load forecasting performance of the ISO remains good overall.

B. Commitment for Local Congestion and Reliability

In New England, several load pockets have historically imported 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")).

81 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.

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.

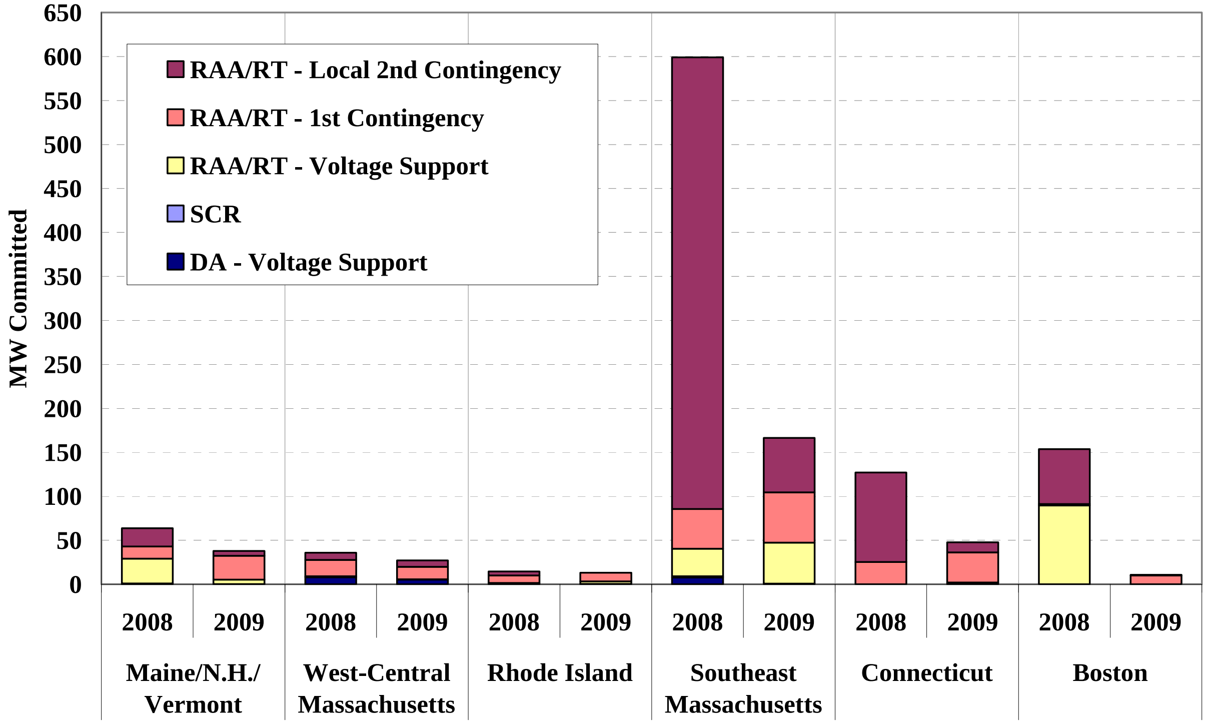
After the day-ahead market, the ISO may need to commit generators with high commitment costs to meet local reliability requirements. Once the commitment costs have been incurred, these generators may be inexpensive providers of energy and reserves in the local area. Because these commitment costs are 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 needed to meet local requirements, so it is important to continue evaluating potential market improvements. This section summarizes the pattern of supplemental commitment for local reliability in the past two years and discusses several initiatives by the ISO to reduce the frequency and effects of supplemental commitment.

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 2008 and 2009.[82] The figure shows the entire capacity 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, 2008 – 2009



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 2009 compared to the prior year. The amount of supplemental commitment decreased 70 percent overall, from an average of 1,000 MW in 2008 to only 300 MW in 2009. Nearly all of the decline occurred in

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

Southeast Massachusetts, Boston, and Connecticut. In Southeast Massachusetts, the average amount of supplemental commitment declined from 600 MW in 2008 to less than 170 MW in 2009, accounting for 62 percent of total decrease. In Boston, the average amount of supplemental commitment fell from 150 MW in 2008 to only 10 MW in 2009. In Connecticut, the average amount of supplemental commitment fell from 130 MW in 2008 to 50 MW in 2009.

In Southeast Massachusetts, the substantial reduction in supplemental commitment was primarily due to the transmission upgrades in Lower Southeast Massachusetts that were brought into service in early July 2009. Historically, supplemental commitment was frequent there because the units needed to ensure local reliability were usually not economic at day-ahead price levels. In order to maintain sufficient reserves, the ISO usually required to commit at least one large unit. As a result, the unit(s) were frequently committed for local reliability and substantial NCPC payments. With the recent transmission upgrades, the ISO no longer needs to commit additional generation for local reliability in this area after the day-ahead market, which substantially reduced the amount of supplemental commitment and associated uplift costs.

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.[83] 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.[84] 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, which was completed and fully placed in service in early 2009. These upgrades significantly increased the

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

84 The behavior change is discussed later in detail in Section VII.B.4. of the "2008 Assessment".

transfer capability into and within Southwest Connecticut, reducing the need to commit local capacity for reliability.

2. 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 70 percent from 2008 to 2009 due to substantially reduced commitment in Lower Southeast Massachusetts and Connecticut. These reductions were primarily due to transmission upgrades in these areas that were completed in 2009, which have reduced the need to commit generation for reliability in these areas.

These transmission upgrades have substantially affected operations in several other ways as well that are discussed in this section. Subsection C illustrates that the quantity of out-of-merit dispatch (i.e., capacity producing output at a cost greater than the LMP) has decreased significantly. Subsection D shows that the amount of surplus online capacity has decreased, which has affected real-time prices. Subsection E reports that the total amount of uplift charges resulting from reliability-committed units and reliability agreements has declined.

C. Out-of-Merit Generation

Out-of-merit generation occurs in real time when energy is produced from an output range on a unit whose 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-run 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.

Out-of-merit generation occurs 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 when they are not economic in the day-ahead market, so their energy offer tends to be higher than the LMP. Third, a unit may be dispatched out-of-merit in real time to satisfy reliability requirements, although this accounts for a very small share of the total out-of-merit generation.[85]

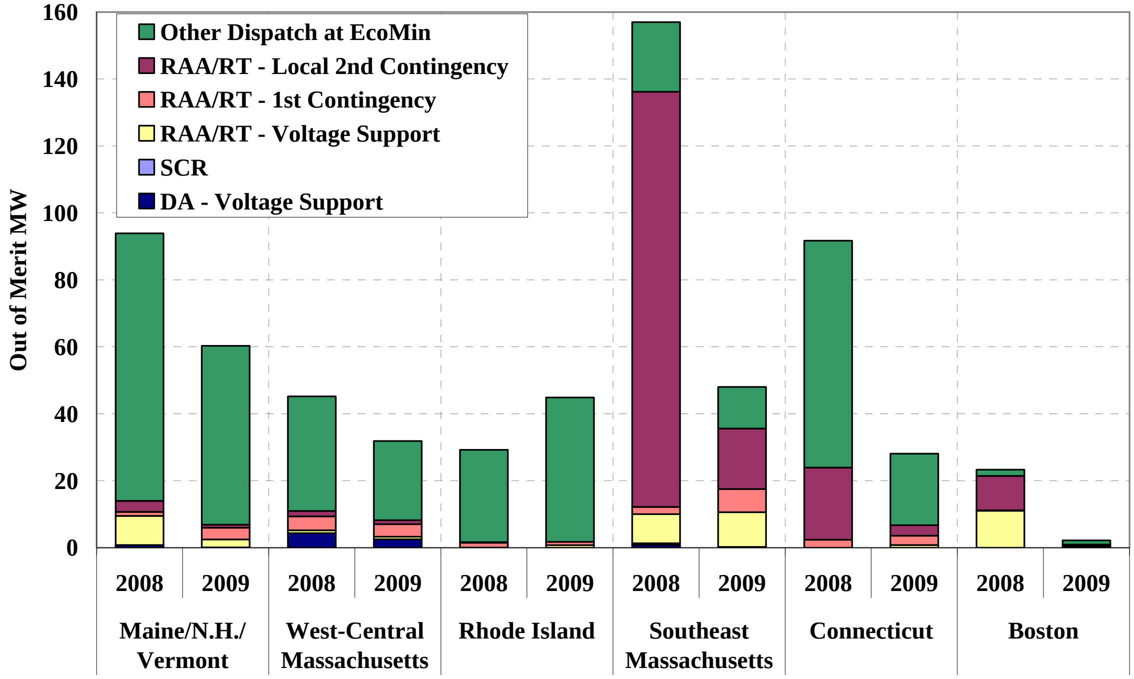
Figure 30 summarizes the average out-of-merit generation by location during peak hours (weekdays 6 AM to 10 PM) in 2008 and 2009. The figure shows five categories of out-of-merit generation on units that are committed (and occasionally dispatched) for reliability reasons.[86] The figure also shows an "other dispatch" category that includes generation from units that were economically committed but are running at their EcoMin.

In most regions, Figure 30 shows that most of the out-of-merit generation 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.

85 Similar to the supplemental commitments, operators may request certain units to run at higher levels than would result from their energy offers. 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.

86 Day-ahead commitments that are flagged for Local Second Contingency are excluded from this category if they occur as the result of market-based scheduling activity. Likewise, day-ahead commitments that are flagged for Voltage Support are excluded from this category if they would have been economically committed.

Figure 30: Average Hourly Out-of-Merit Generation
Weekdays 6 AM to 10 PM, 2008 – 2009



The average quantity of out-of-merit generation from units committed for local reliability (including first contingency, second contingency, voltage support, and SCR) declined 71 percent, from an average of 208 MW in 2008 to 60 MW in 2009. The amount of out-of-merit energy from non-local reliability units (i.e., Other Dispatch at EcoMin) declined 33 percent, from an average of 232 MW in 2008 to 155 MW in 2009.

The changes in out-of-market dispatch that occurred in 2009 tracked the changes in supplemental commitments and were caused by the same underlying factors. The reduced commitment for reliability in Southeast Massachusetts, Connecticut, and Boston led to proportionate reductions in out-of-merit energy in those zones.

The reduced commitment for reliability has also helped reduce the quantity of out-of-merit generation from economically committed units. This is because the reduced commitment for reliability has led to a decline in the quantity of surplus online capacity (i.e., capacity that is not required to satisfy energy and ancillary services requirements), which has reduced the frequency of units operating at their EcoMin levels during peak hours.

Hence, the reduction in supplemental commitment for reliability in 2009 has been beneficial because it has:

- Reduced the quantity of out-of-merit generation from units committed for reliability; and
- Reduced the quantity of out-of-merit generation from economically committed units.

These results suggest that generating units are being dispatched more consistently with real-time LMPs. This is beneficial because it improves incentives for efficient dispatch from the lowest-cost resources.

D. Surplus Capacity and Real-Time Prices

Under normal operating conditions, the available online and fast-start capacity is more than sufficient to satisfy load and reserve requirements, which suggests that some surplus capacity will exist in almost every hour. This is a normal outcome in a properly functioning market. Surplus capacity does not raise concerns unless inflated by inefficient commitments by the ISO or market participants.

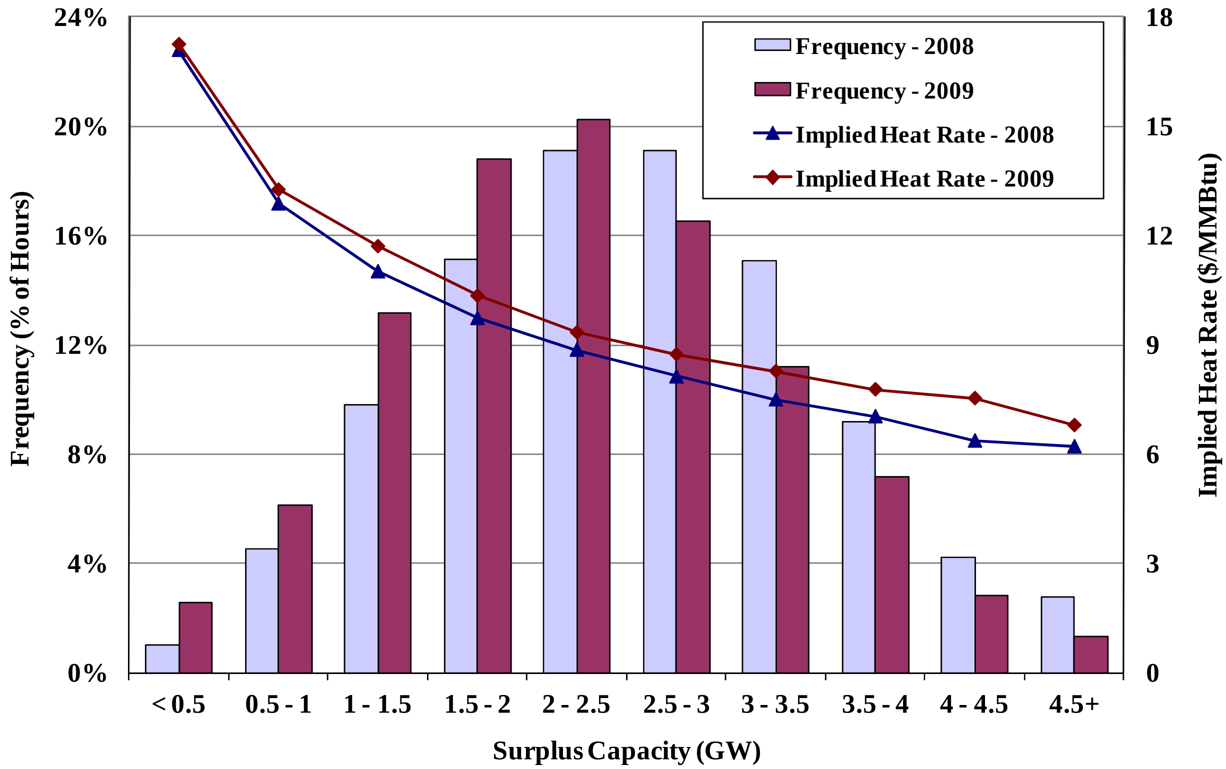
Surplus capacity is also important because it determines what resources are available to respond to changes in real-time operating conditions, which can change unexpectedly. Accordingly, the quantity of surplus capacity exhibits a strong negative correlation with real-time energy prices. This section evaluates the pattern of surplus capacity and real-time energy prices. In this report, we define "Surplus Capacity" as the amount of capacity that is online or capable of starting within 30 minutes in excess the amount required to meet load and reserve requirements. Hence, surplus capacity is equal to:

Online Reserves + Offline Reserves Deployable in 30 min – TMOR Requirement

Figure 31 summarizes the relationship of Surplus Capacity to real-time energy prices at New England Hub in each hour of 2008 and 2009. Each bar shows the frequency of hours when Surplus Capacity was in the range of values shown on the horizontal axis. For example, there was 1.0 to 1.5 GW of Surplus Capacity in approximately 10 percent of the hours in 2008 and 13 percent of the hours in 2009. The lines show the average real-time implied marginal heat rate at

New England Hub in the hours that correspond to each range of Surplus Capacity. For example, in hours when there was 1.0 GW to 1.5 GW of Surplus Capacity, the average real-time implied marginal heat rate was 11.0 MMbtus per MWh in 2008 and 11.7 MMbtus per MWh in 2009. The implied marginal heat rate is shown in order to normalize real-time energy prices for changes in natural gas prices during 2008 and 2009.[87]

Figure 31: Surplus Capacity and Implied Marginal Heat Rates Based on Real-Time LMPs at the Hub in Peak Hours, 2008 – 2009



The figure shows a strong correlation between the quantity of surplus capacity and the implied marginal heat rate in real time. In 2009, the average implied marginal heat rate was highest (17.3 MMbtus per MWh) in hours with less than 0.5 GW of surplus capacity and lowest (6.8 MMbtus per MWh) in hours with more than 4.5 GW of surplus capacity.

87 In this section, the implied marginal heat rate in a particular hour is equal to the real-time LMP divided by the natural gas index price.

After normalizing for variations in natural gas prices and the level of surplus capacity, the figure shows that average real-time prices were relatively consistent from 2008 to 2009. However, the figure shows a modest (0.2 to 1.2 MMbtus per MWh) premium on implied marginal heat rates in 2009 relative to 2008. The increase from 2008 to 2009 is due to several factors that are discussed in Section II.

The figure shows significant reductions in the amount of surplus capacity from 2008 to 2009, which have contributed to higher real-time LMPs. The percentage of hours when there was less than 0.5 GW of surplus capacity increased from 1.0 percent in 2008 to 2.6 percent in 2009. This is a significant development, since these hours exhibited substantially higher than average price levels with average implied marginal heat rates above 17 MMbtu per MWh. Likewise, the percentage of hours when there was less than 2 GW of surplus capacity increased from 30 percent in 2008 to 41 percent in 2009. These reductions are largely due to the sharp reduction in supplemental commitments for local reliability needs that are discussed earlier in this section. Although these commitments were necessary in the past, they tended to increase system-wide surplus capacity levels.

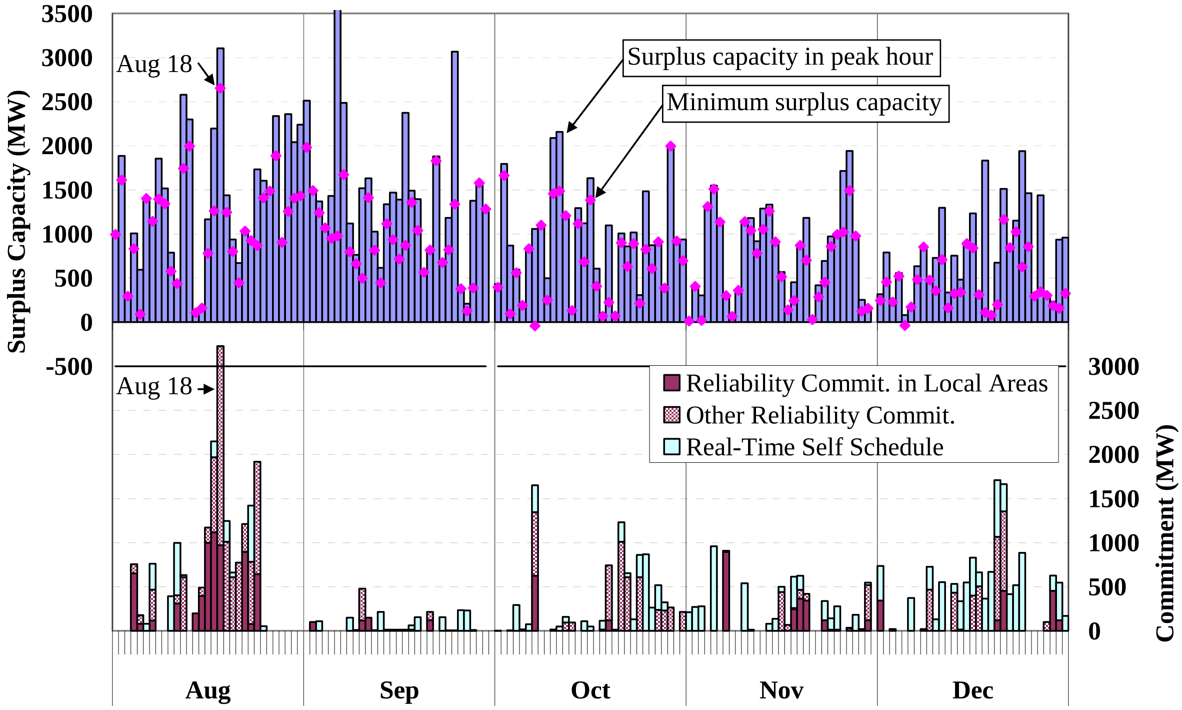
This evaluation shows how real-time LMPs are affected by the amount of surplus capacity, which decreased substantially in 2009. To the extent that real-time LMPs are higher as a result of more efficient commitment and dispatch, it is an indication of more efficient market operations and real-time pricing. Nevertheless, we (i.e., the External Market Monitoring Unit) and the Internal Market Monitoring Unit review market outcomes to identify potential exercises of market power, and address them appropriately to ensure that prices and other market outcomes are competitive and efficient.

E. Supplemental Commitments and Surplus Capacity

Given the effect of surplus capacity on prices, it is important to evaluate the supplemental commitments made by the ISO and self-commitments made by market participants after the day-ahead market. Inefficient commitments in either area will tend to distort real-time prices, while inefficient supplemental commitments by the ISO will also lead to increased uplift costs.

As detailed above, transmission upgrades have substantially reduced the need for the ISO to commit generation to satisfy local reliability requirements. By July 2009, the ISO's need to commit supplemental resources for local reliability had largely been eliminated. However, the ISO must still periodically make commitments to satisfy New England's system-wide reliability requirements. To evaluate the effectiveness of this process, Figure 32 shows the supplemental commitments and self-scheduled commitments by day in the bottom panel, and the surplus capacity in the peak load hour and the minimum surplus capacity in any hour of each day.

Figure 32: Daily Supplemental Commitments and Surplus Capacity August to December, 2009



In general, large quantities of commitments for system-wide reliability needs that lead to large surplus capacity levels would raise a concern because such commitments generally raise costs to New England's customers and distort real-time prices. Figure 32 shows that the ISO did not frequently make large quantities of supplemental commitments after the local reliability issues in Southeast Massachusetts were addressed in July 2009. When the ISO has made supplemental commitments for system-wide capacity needs, the minimum surplus capacity levels have been low on a majority of these days.

However, there have been a number of days when large quantities of supplemental commitments have resulted in large quantities of surplus capacity. The most notable day was August 18, 2009 when supplemental commitments totaling roughly 3200 MW resulted in surplus capacity in real time of more than 3000 MW and $1.5 million of NCPC for the day. We worked with ISO-NE staff to understand the factors that led to the high levels of supplemental commitments on that day relative to the minimum amounts that were ultimately necessary to satisfy ISO New England's reliability requirements. We found that there were a number of factors on that day that contributed to the unusually large capacity surplus.

It is important to recognize that uncertainties tend to have a bigger effect in New England due to the limited quantity of fast-start generating resources. This causes the ISO in some cases to have to rely on slower-starting units that must be notified well in advance of the operating hour when uncertainty regarding load, imports, and generator availability is high. For example, most of the commitments on August 18 were slow-starting units that were made overnight, more than 12 hours before the forecasted peak.

The two assumptions in the reliability commitment process that made the largest contributions to the over-commitment on August 18 were:

- The "desired capacity surplus" that operators have the discretion to determine to account for concerns regarding generator availability, load forecast errors, or other factors;[88] and
- The assumed imports and exports.

In general, the desired capacity surplus should be minimized since the operating reserve requirements are set at levels that should ensure reliability. Adding a non-zero desired capacity surplus introduces an inconsistency between the market requirements and the operating requirements. However, we recognize that conditions can sometimes arise that would justify an increase in the desired capacity surplus.

88 The operators have the discretion to commit surplus generation when they believe it is necessary to deal with uncertainty as stated in the System Operating Procedure, Perform Reserve Adequacy Assessment, Section 5.3.2.3, "The Forecaster may commit additional Generators as needed for reliability (anticipated storms, hurricanes or other conditions that affect Bulk Power System reliability)."

With regard to the import and export assumptions, we believe that improvements are possible based on our review. In general, these assumptions are based on the day-ahead scheduled transactions. By committing generation to support day-ahead transactions, they are treated as firm and we understand from the ISO that the operators generally do not curtail day-ahead exports. This treatment of the day-ahead exports in the capacity evaluation process raises potential efficiency concerns because:

- The participants are not obligated to schedule the exports in real time, which could render the units committed to support them unnecessary (which happened on Aug. 18).

- The value of the day-ahead exports may not justify the commitment costs of the supplemental commitments made to support them.

Hence, the ISO should consider whether its assumptions regarding imports and exports in its capacity evaluation process could be improved. Additionally, if the ISO New England moves ahead with the New York ISO in implementing intra-hour scheduling, it should consider alternative assumptions for its capacity evaluation. These scheduling improvements should rationalize the physical flow between the two markets. This should, in turn, allow the ISO to rely more heavily on the markets to cause power to flow into New England when and if shortages occur. Therefore, committing generation to support day-ahead imports and exports would no longer be necessary.

F. 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 2008 and 2009. 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.[89] In 2009, 31 percent of the capacity in Connecticut was covered under reliability agreements.
- Local Second Contingency Protection Resources – In 2009, 95 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.[90] 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 2009, 88 percent of this uplift was allocated to Real-Time Deviations throughout New England.[91] 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.

---

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

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

91 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, uninstructed generation deviations from day-ahead schedules, and virtual supply schedules.

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

Table 3: Allocation of Uplift for Out-of-Market Energy and Reserves Costs 2008 – 2009

	Millions of Dollars	
Category of Uplift	**2008**	**2009**
Reliability Agreement		
Connecticut	$110	$61
Other Areas	$19	$24
Local Second Contingencies		
Connecticut	$24	$2
Boston	$11	$0.3
Southeast Massachusetts	$143	$14
Other Areas	$4	$1
Special Case Resources	$2	$1
Voltage Support	$29	$5
Other (Mostly Local First Contingencies)	$44	$31
Total	$387	$139

Overall, uplift charges fell significantly from $387 million in 2008 to $139 million 2009, a decrease of 64 percent. The largest decline occurred in uplift charges for local second contingency protection, which fell 91 percent from $182 million in 2008 to only $17 million in 2009. These sharp reductions in uplift charges were primarily due to three factors:

- Supplemental commitment for reliability fell substantially from 2008 to 2009, reducing the amount of capacity receiving NCPC payments.
- Fuel prices fell significantly from 2008 to 2009, reducing the portion of NCPC payments that are related to the fuel costs of individual generators.
- Reliability agreement costs fell 34 percent from 2008 to 2009, primarily due to the expiration of reliability agreements for several units during this period.

In 2009, 61 percent of the uplift charges were associated with reliability agreements. Hence, uplift charges are expected to fall further in 2010 when the remaining reliability agreements are scheduled to expire.

G. Conclusions and Recommendations

We conclude that the ISO's operations to maintain adequate reserve levels in 2009 were reasonably accurate and consistent with the ISO's procedures. The amount of capacity committed for reliability decreased significantly in Lower Southeast Massachusetts, Connecticut and Boston. This was primarily due to transmission upgrades that substantially increased the import capability into these areas and greatly reduced the need to commit additional resources after the day-ahead market to satisfy local reliability requirements.

We regularly review patterns of supplemental commitment and the resulting out-of-merit generation because they generally lead to the following four market issues:

- 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.
- Large and volatile uplift charges that can be 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.

To ensure that these issues are minimized, we recommend that the ISO review its assumptions and processes for determining that additional commitments are necessary to satisfy its reliability requirements. In particular, the ISO should consider the assumptions it currently makes regarding the imports and exports that will be scheduled in real time. 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.

We also recommend several changes in Sections V and VI that would help the real-time prices of energy and reserves better reflect the costs of maintaining reliability during tight operating conditions. 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 system reliability criteria.

VIII. Competitive Assessment

This section evaluates the competitive performance of the New England wholesale markets in 2009. This type of assessment is particularly important for LMP markets. While LMP markets increase overall system efficiency, they can 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.[92] In this section we address four main areas:

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

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).

92 See, *e.g.*, Section IX of the "2008 Assessment".

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 price in the spot market.[93] 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.

93 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 price. This subsection 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 (*e.g.*, electricity from a coal-fired plant is substitutable 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 and slower-starting 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 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 as energy produced in real time for our analysis.

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, which is in Southwest Connecticut;
- Boston; and
- Lower SEMA.

This subsection analyzes the seven geographic areas listed above using the following structural market power indicators:

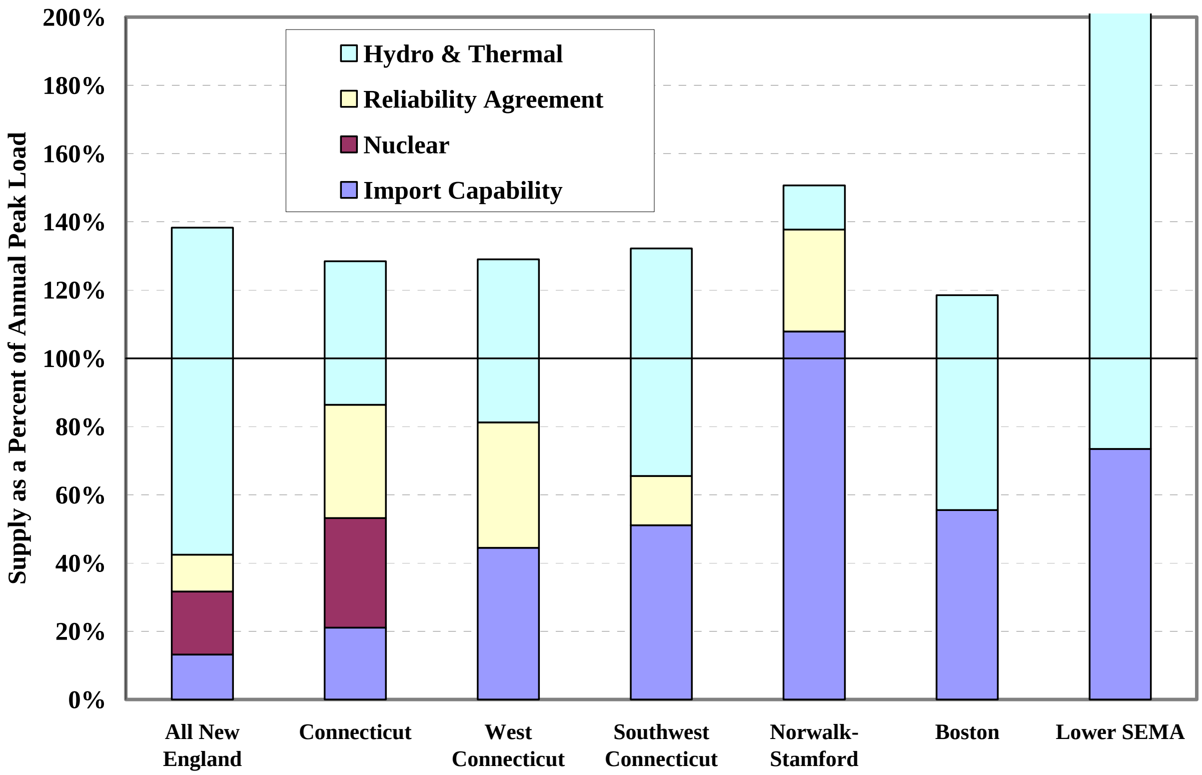
- Supplier market shares;
- Herfindahl-Hirschman indices; and
- Pivotal supplier indices.

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 33 shows several categories of supply relative to the load in each of the seven regions of interest.

Figure 33: Supply Resources versus Summer Peak Load in Each Region 2009



For each region, Figure 33 shows import capability[94] 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 2009 peak load, although a substantial quantity of additional capacity (typically around 2,000 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.

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

In each region shown in Figure 33, the relative shares for categories of internal supply increased slightly from 2008 to 2009. This is because the summer peak load levels fell slightly from 2008 to 2009 and because there were very few changes to the supply of internal resources in each region.

However, the amount of import capability into several regions changed considerably from 2008 to 2009. The import capability into Southwest Connecticut increased from 40 percent of peak load in 2008 to 51 percent in 2009. The import capability into Norwalk-Stamford increased from 73 percent of peak load in 2008 to 108 percent in 2009, effectively eliminating it as an area of significant concern for market power. This was primarily due to the completion of transmission upgrades under Phase II of the Southwest Connecticut Reliability Project (Middletown to Norwalk Project) in 2009, which significantly improved the transmission system infrastructure and increased the transfer capability into the Norwalk-Stamford and Southwest Connecticut area. Although the import capability into Lower SEMA also increased substantially in 2009 due to transmission upgrades, the figure shows the import capability to Lower SEMA prior to completion of the project.[95] For the other regions shown in Figure 33, changes in import capability from 2008 to 2009 were primarily attributable to the differences in network topology, generation patterns, and load patterns during the peak load hours in the two years.

Figure 33 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 15 percent, ranging from 19 percent in Boston to more than 100 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 because the lost revenue on withheld

95 The figure shows the import capability into the Lower SEMA area in early July of 2009 right before the short-term transmission upgrades in Lower SEMA were completed. This upgrade effectively eliminated constraints into Lower SEMA.

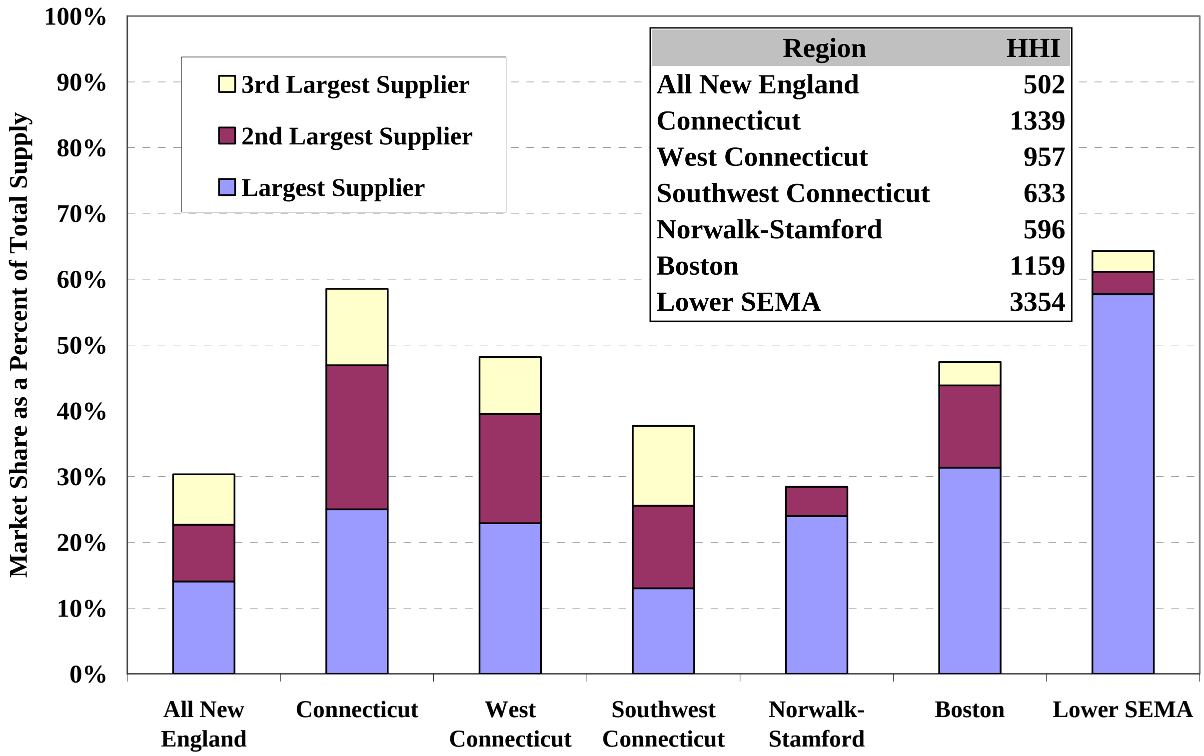
units can be very costly. Nuclear generators typically cannot be dispatched up and down in a manner that would allow the owner of the unit to profitably withhold. Thus, the owner of nuclear generation would have to also own significant amounts of non-nuclear capacity that could be withheld from the market. 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 Unit 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 2009. 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 2009, further 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 34 shows the market shares of the largest three suppliers coinciding with the annual peak load hour on August 18, 2009. 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 34 indicates a substantial variation in market structure across New England. In all New England, the largest supplier had a 14 percent market share in 2009. In the load pockets, the largest suppliers had market shares ranging from 13 percent in Southwest Connecticut to 58 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 2009, while Southwest Connecticut had three native suppliers with comparable market shares.

Figure 34: Installed Capacity Market Shares for Three Largest Suppliers August 18, 2009



The HHI figures suggest that only Lower SEMA is highly concentrated.[96] The HHI for Norwalk-Stamford is 596, which is relatively low for most product markets. This is counter-intuitive since there are only two suppliers in the area. However, because its load can be entirely

96 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.

served by imports, the need for local suppliers is very limited. Of the remaining areas, Connecticut and Boston have the highest HHI statistics with 1339 and 1159, respectively.

While HHI statistics can be instructive in generally indicating the concentration of the market, they alone 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, they do not consider demand conditions, load obligations, or the heterogeneous effects of generation on transmission constraints based on their location. In the next subsection, we 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.[97, 98] 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

97 The DOJ and FTC evaluate the *change* in HHI as part of their 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.

98 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.

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 power in granting market-based rates.[99] 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.

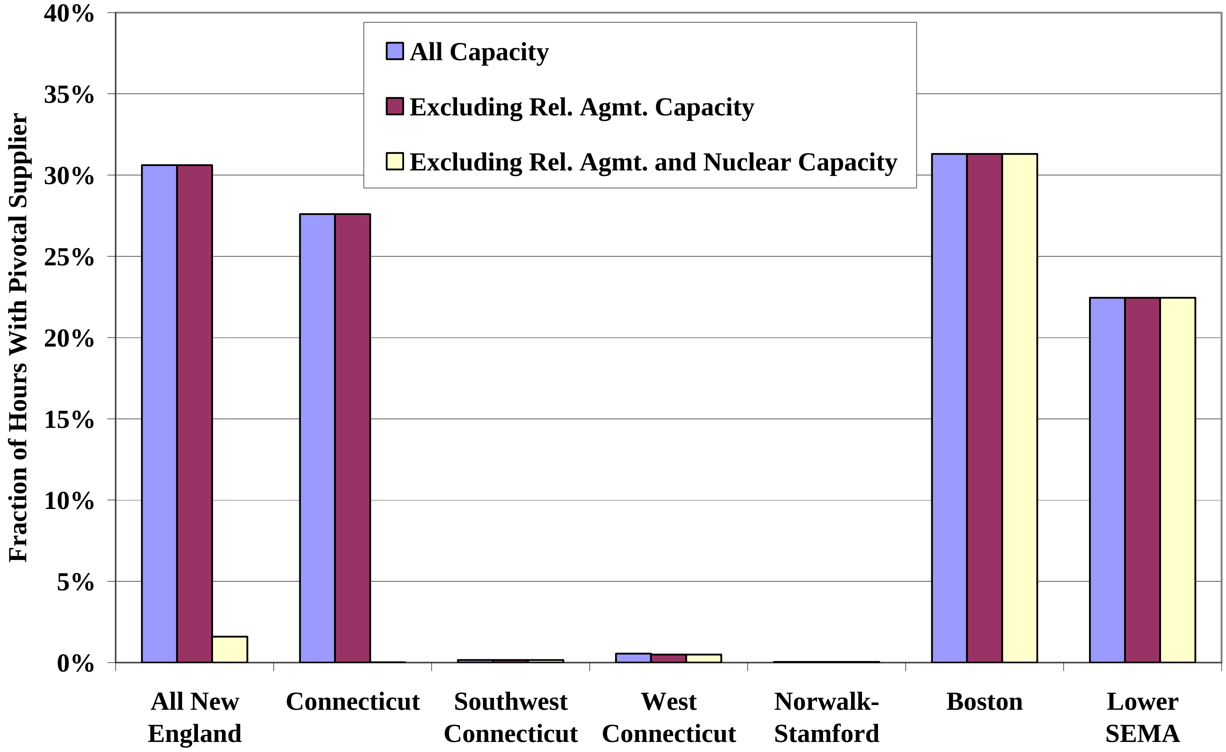
Even small suppliers can be pivotal for brief periods. For example, all suppliers are pivotal during periods of shortage. This does not mean that all suppliers should be deemed to have market power. As described above, suppliers must have both the *ability* and *incentive* to raise prices to have market 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 35 shows the portion of hours where at least one supplier was pivotal in each region during 2009. 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 four of the seven areas shown in Figure 35. The areas that do not raise potential concerns are Norwalk-Stamford, 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.

99 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 35: Frequency of One or More Pivotal Suppliers by Type of Withheld Capacity 2009



The figure shows that at least one supplier was pivotal in Lower SEMA for approximately 22 percent of the hours in 2009. However, all of these hours occurred during the first six months of 2009. After the transmission upgrades were completed in early July, no supplier was ever pivotal in Lower SEMA because the import capability is now sufficient to satisfy local needs. Likewise, in Norwalk-Stamford, the pivotal frequency declined from 12 percent in 2008 to zero in 2009, primarily due to the transmission upgrades made under Phase II of the Southwest Connecticut Reliability Project.

Potential local market power concerns are most acute in Boston, where one supplier 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 2009.

Although Connecticut had a pivotal supplier in 28 percent of the hours in 2009, 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 in order to raise the clearing prices paid for its nuclear production.[100] 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 2009, the market power conclusions depend primarily on how nuclear capacity affects the incentives of large suppliers. Excluding nuclear capacity from the pivotal supplier analysis for all of New England would substantially reduce the pivotal frequency (from 31 percent to 2 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.

In Connecticut and all of New England, the pivotal frequency rose notably from the prior year, from 17 percent to 28 percent in Connecticut, and from 13 percent to 31 percent in all of New England.[101] These increases were partly driven by the substantial declines in reliability commitment, which greatly reduced the supply margin in these areas compared to 2008.[102]

The pivotal supplier summary in Figure 35 indicates the greatest potential for market power in 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.

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

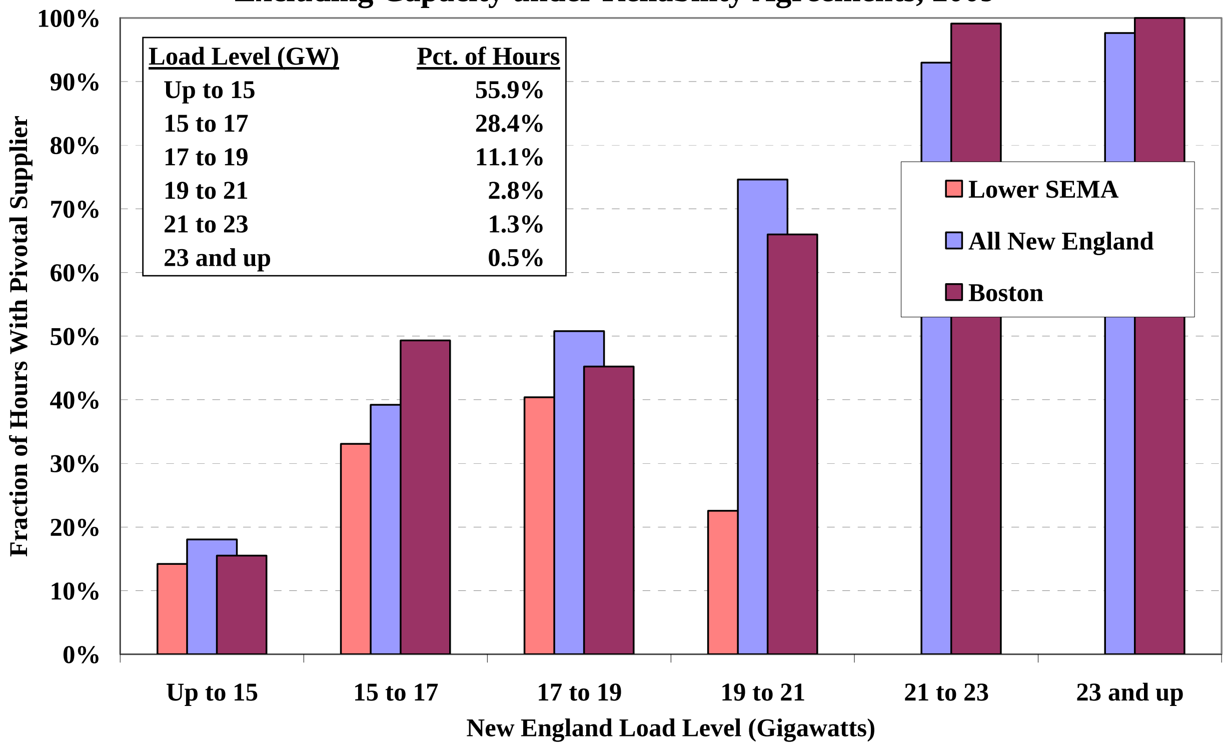
101 We reported a pivotal frequency of 7 percent in 2008 for all of New England because the analysis assumed lower than actual reserve requirements. The analysis in this report uses the actual reserve requirement.

102 The changes in reliability commitment pattern are discussed in detail in Section VIII.

As described above, 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 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 36 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, Boston on the right 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 there were very few instances of a supplier being pivotal during 2009. 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 36: Frequency of One or More Pivotal Suppliers by Load Level Excluding Capacity under Reliability Agreements, 2009



A supplier in Boston was pivotal in at least 45 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 17 GW. The pivotal frequency fell below 20 percent in Boston and all of New England during hours when load was below 15 GW in New England. The pivotal frequency in Lower SEMA fell to zero when load exceeded 21 GW in all of New England because these conditions all occurred after the transmission upgrades were placed in service in July 2009. These upgrades increased the import capability sufficiently to cause the largest supplier to not be pivotal.

Based on the pivotal supplier analysis in this subsection, market power is no longer a concern in Lower SEMA due to transmission upgrades, but is most likely to be a concern in Boston and all of New England when load exceeds 17 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 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 operating and maintenance costs). 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 Monitor 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 Monitor 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 real-time market outcomes for fast-start units and day-ahead market outcomes for slower-starting 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, it may lead the unit to 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 subsection, 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 37 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 subsection, the largest supplier can expect that its capacity will be pivotal in most hours when load exceeds 15 GW.

Figure 37: Average Output Gap by Load Level and Type of Supplier Boston, 2009

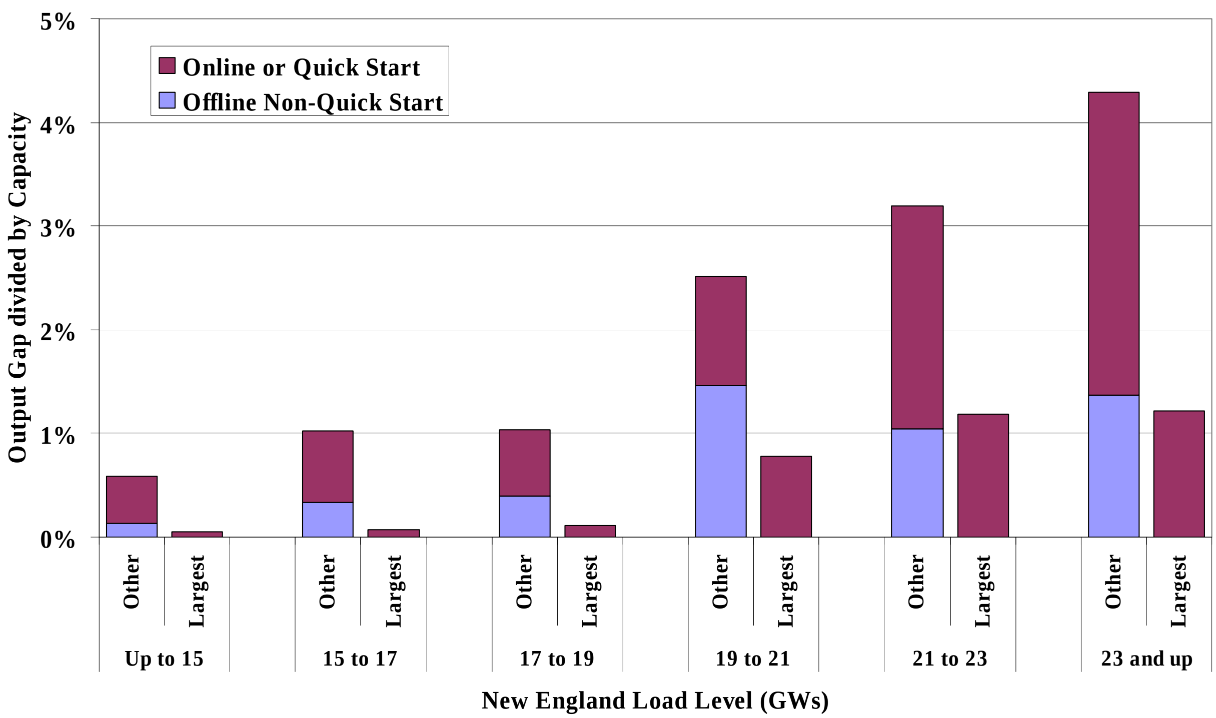


Figure 37 shows that the overall amount of output gap for the largest supplier in Boston was small as a share of total demand in 2009, although the amount of output gap increased at higher load levels.

The level of output gap for the largest supplier was down considerably from 2008. In prior reports, we identified a pattern of output gap that was likely the result of offers above marginal cost that had the effect of increasing the NCPC payments received by several generators. In April 2008, revisions to the reliability requirements for Boston area voltage led to sharp reductions in the frequency of supplemental commitment for reliability. After the reduction in commitment for reliability, the conduct of the largest supplier changed in a manner that significantly reduced the output gap in 2009.[103]

The output gap for the other suppliers was only about one percent when the New England load was below 19 GW, but it rose as load increased, ranging from approximately 2 to 4.5 percent. The increased level of output gap for the other suppliers during high load periods was related to several small units that had issues with their primary fuel supply. This caused them to have to switch to the more expensive alternative fuel in order to reach full output. The output gap associated with these units was not concentrated in periods of congestion and likely did not have a significant price impact. Therefore, these results do not raise significant competitive concerns.

3. Output Gap in Lower SEMA

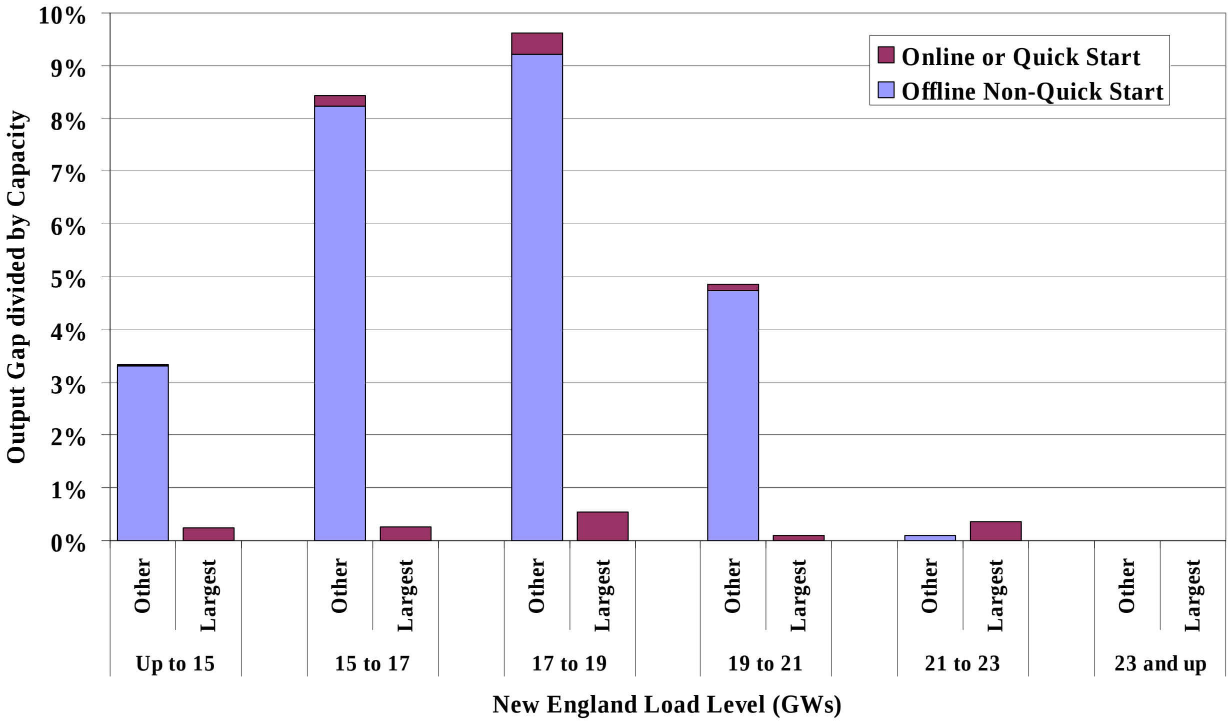
Lower SEMA was the most import-constrained load pocket in New England before the transmission upgrades were completed in July 2009. The largest supplier (who owns approximately 90 percent of the capacity) in Lower SEMA was pivotal in the majority of hours during the first half of 2009. We focus particular attention on 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 38 shows the output gap identified in Lower

103 See, *e.g.*, Section IX.E of the "2008 Assessment".

SEMA in 2009 by load level. The output gap is shown separately for the largest supplier and for other suppliers.

Figure 38: Average Output Gap by Load Level and Type of Supplier Lower SEMA, 2009

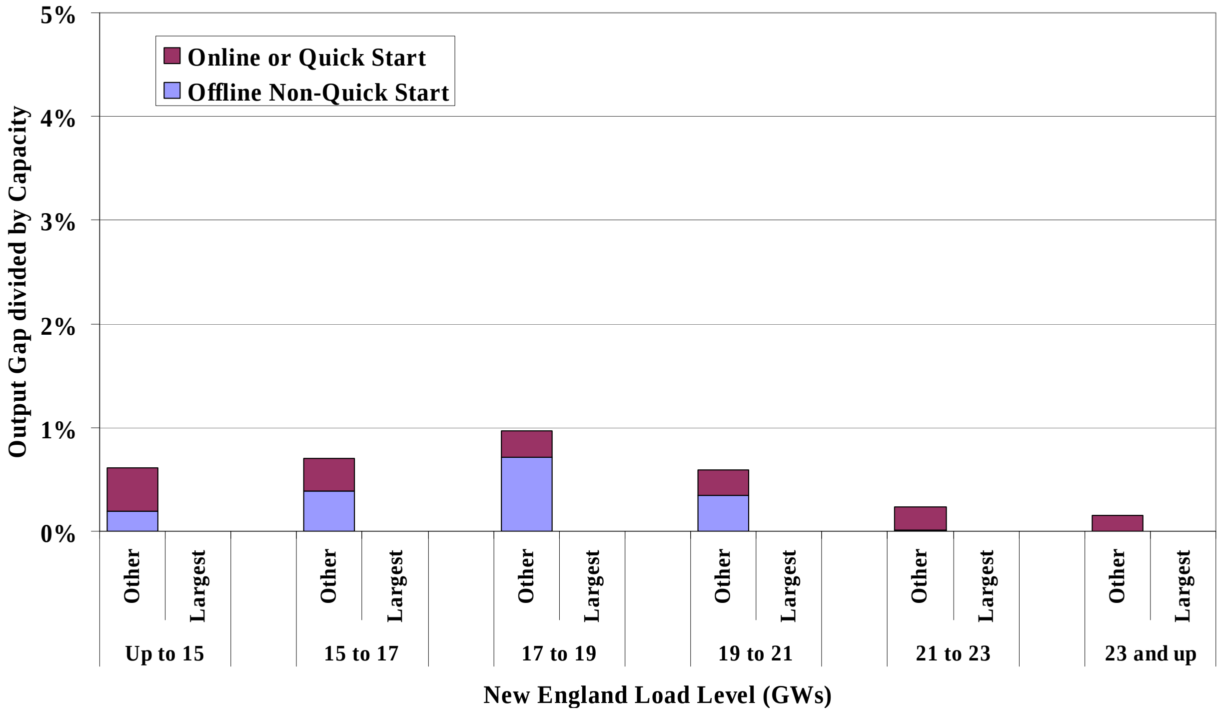


Although the pivotal supplier analysis identified Lower SEMA as a region with the potential for market power, the figure shows that the output gap of the largest supplier's online and fast-start units was very low in 2009. These results clearly indicate that the largest supplier's offers have consistently been priced below the Conduct Threshold for Economic Withholding, which is identified in Appendix A of Market Rule 1. The output gap for the other suppliers in Lower SEMA reached nearly 10 percent of their capacity during moderate load conditions. This output gap was related to a small dual-fueled unit with a reference level based on natural gas that frequently uses oil.

4. Output Gap in Connecticut

In this subsection, we examine potential economic withholding in Connecticut. Historically, Connecticut has been import-constrained, although the pivotal supplier analysis does not raise significant concerns about the potential exercise of market power in 2009 in Connecticut. Figure 39 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 39: Average Output Gap by Load Level and Type of Supplier Connecticut, 2009

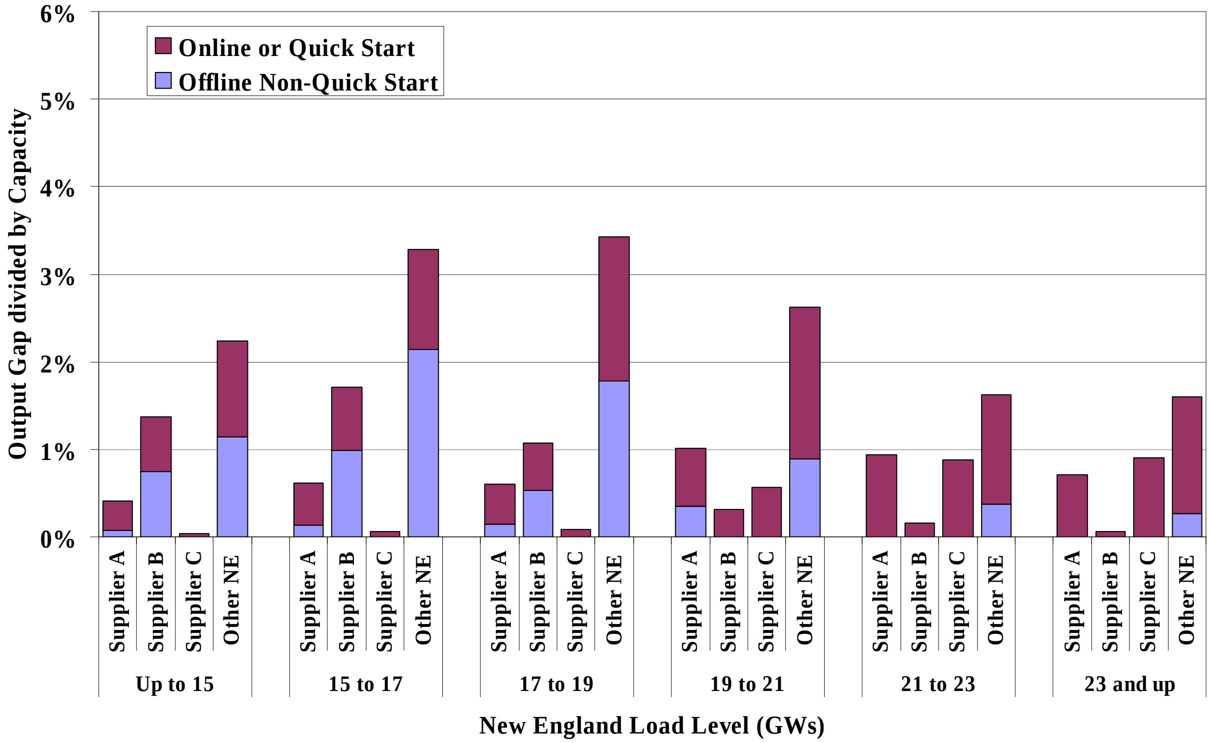


The pivotal supplier analysis indicated that the largest supplier in Connecticut was pivotal in less than 1 percent of all hours. Figure 39 also shows that the total output gap of all suppliers was very low relative to the total capacity in Connecticut. The largest supplier owns exclusively nuclear capacity and had no output gap in 2009. The other suppliers also produced very little output gap. Given these amounts, these results do not raise concerns regarding economic withholding in Connecticut.

5. Output Gap in All New England

Figure 40 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 31 percent of the hours during 2009 (excluding capacity under reliability agreements). Suppliers B and C are the second and third largest suppliers in New England and were each pivotal during less than 1 percent of the hours. All other suppliers are shown as a group for reference.

Figure 40: Average Output Gap by Load Level and Type of Supplier
All New England, 2009



The figure shows that the region-wide output gap was generally low for each of the four categories of supply. Suppliers A and C exhibited small output gap levels 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 17 GW. It is notable that the output gap levels for the three largest suppliers were lower than the output gap levels of all other suppliers.

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 2009.

D. Physical Withholding

This section of the report examines declarations of forced outages and other non-planned deratings to determine if there is any evidence that the suppliers are exercising market power. In this analysis, we evaluate the four geographic markets examined in the output gap analysis above: Boston, Lower SEMA, Connecticut, and all of New England.

In each market, we examine forced outages and other deratings by load level. The "Other Derate" category includes any reduction in the hourly capability of a unit from its maximum seasonal capability that is not logged as a forced outage or a planned outage. 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 41 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 19 GW. We compare these statistics for the largest supplier to all other suppliers in the area.

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 2009, 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. Under low load conditions, this operating practice does not raise competitive concerns and is consistent with competitive conduct.

Figure 41: Forced Outages and Deratings by Load Level and Supplier
Boston, 2009

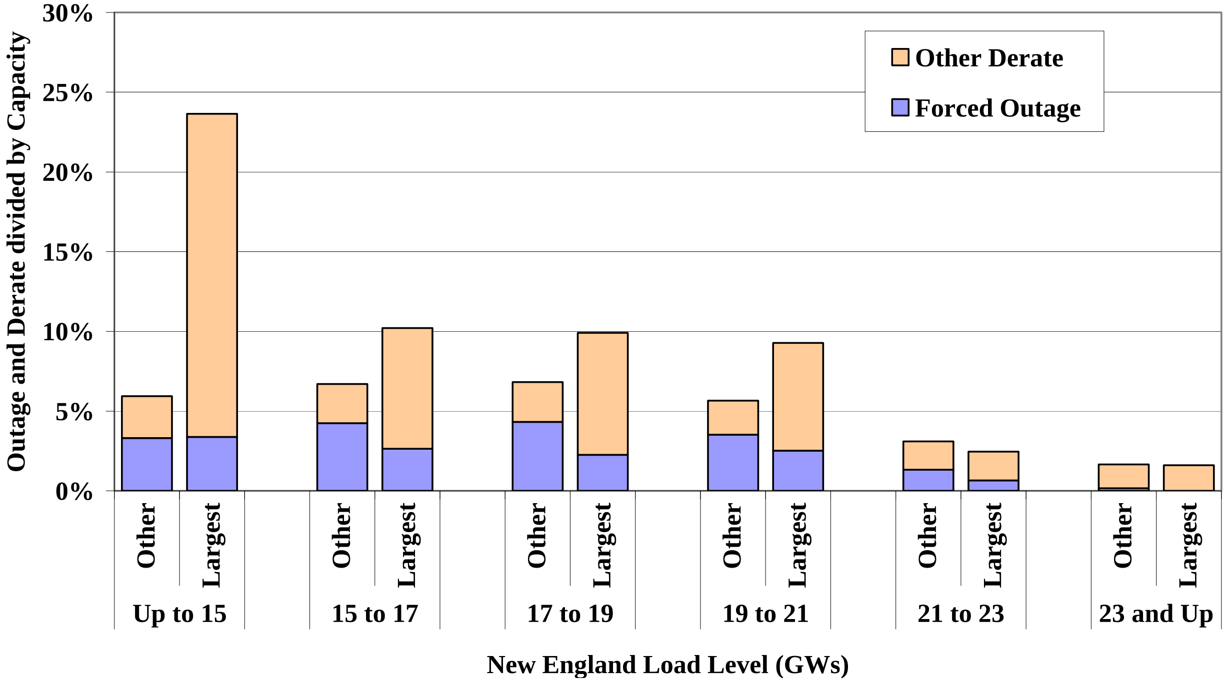
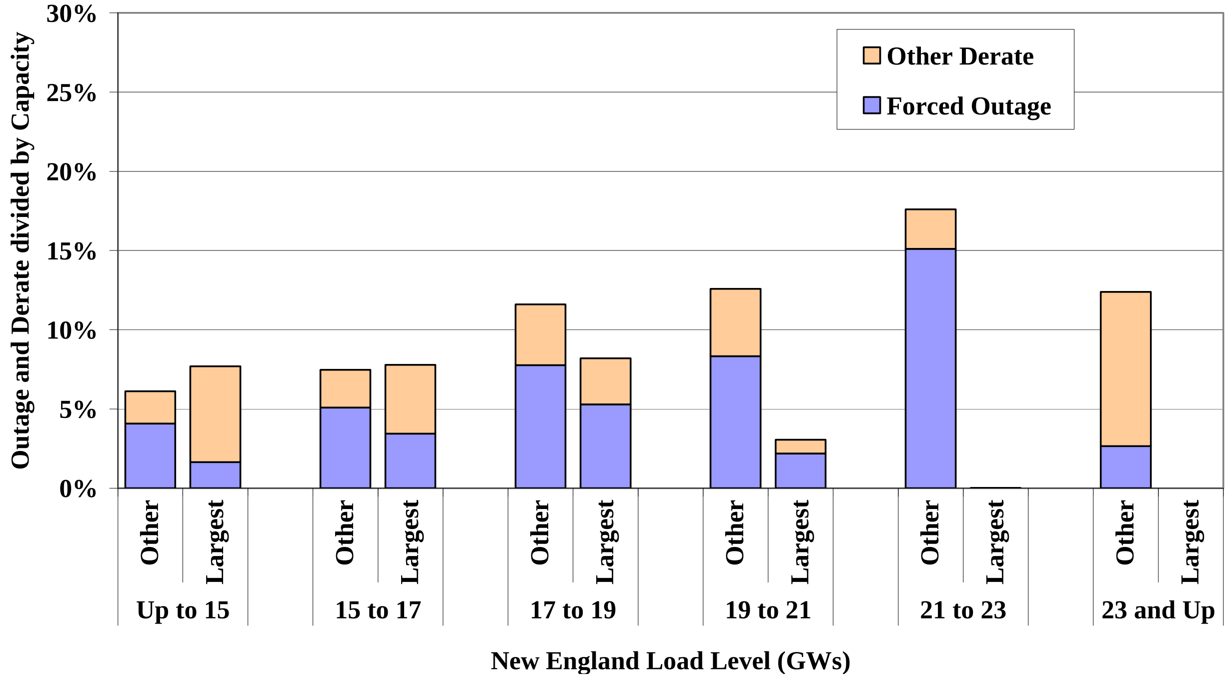


Figure 41 shows a pattern of deratings and outages consistent with expectations in a competitive market. Although levels of outages and deratings for the largest supplier were high at low load levels, they were comparable to other suppliers when load exceeded 21 GW (when withholding is most likely to 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 42 summarizes declarations of forced outages and other deratings in the Lower SEMA area by load level in 2009. These statistics are shown for the largest supplier compared with the other suppliers in the area.

Figure 42: Forced Outages and Deratings by Load Level and Supplier
Lower SEMA, 2009



Based on the pivotal supplier analysis for Lower SEMA, the largest supplier was pivotal in 22 percent of the hours in 2009. The pivotal hours, however, all occurred in the first half of 2009. No supplier has been pivotal in the Lower SEMA area since July 2009 due to transmission upgrades. The rate of forced outages and other deratings for the largest supplier was modest during low load periods and declined considerably when load increased, falling well below the level of other suppliers. Overall, the outage and deratings results for Lower SEMA do not raise competitive concerns.

3. Potential Physical Withholding in Connecticut

Figure 43 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 43: Forced Outages and Deratings by Load Level and Supplier
Connecticut, 2009**

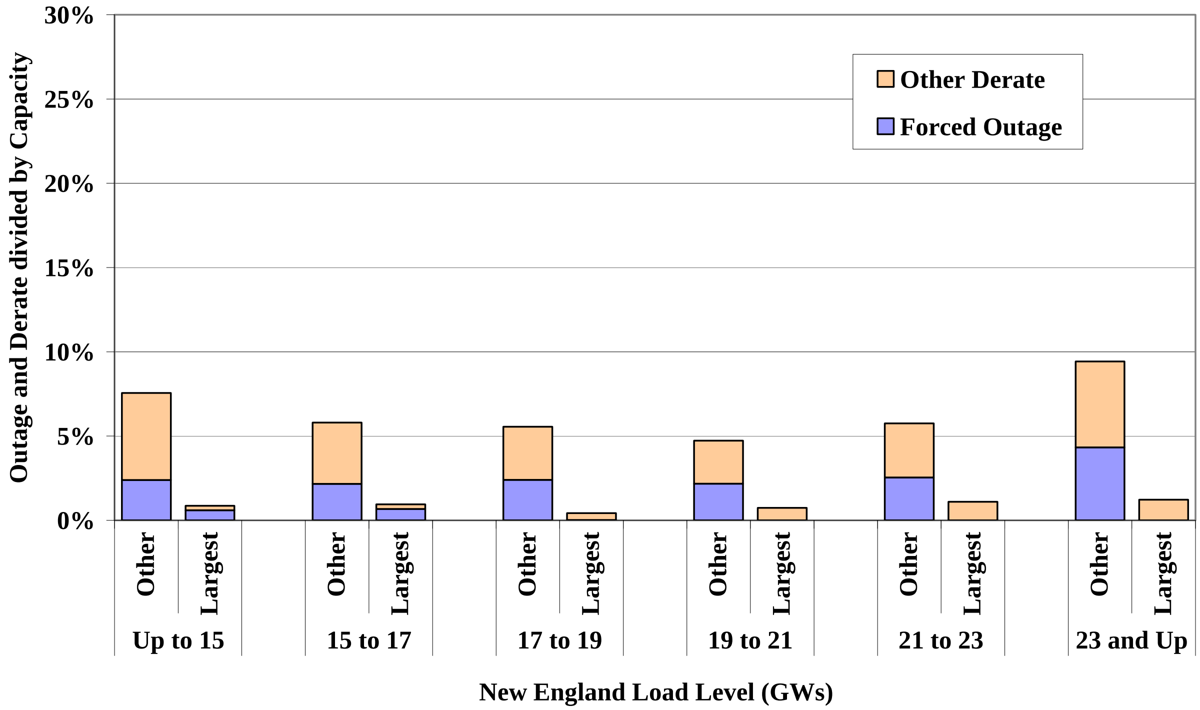
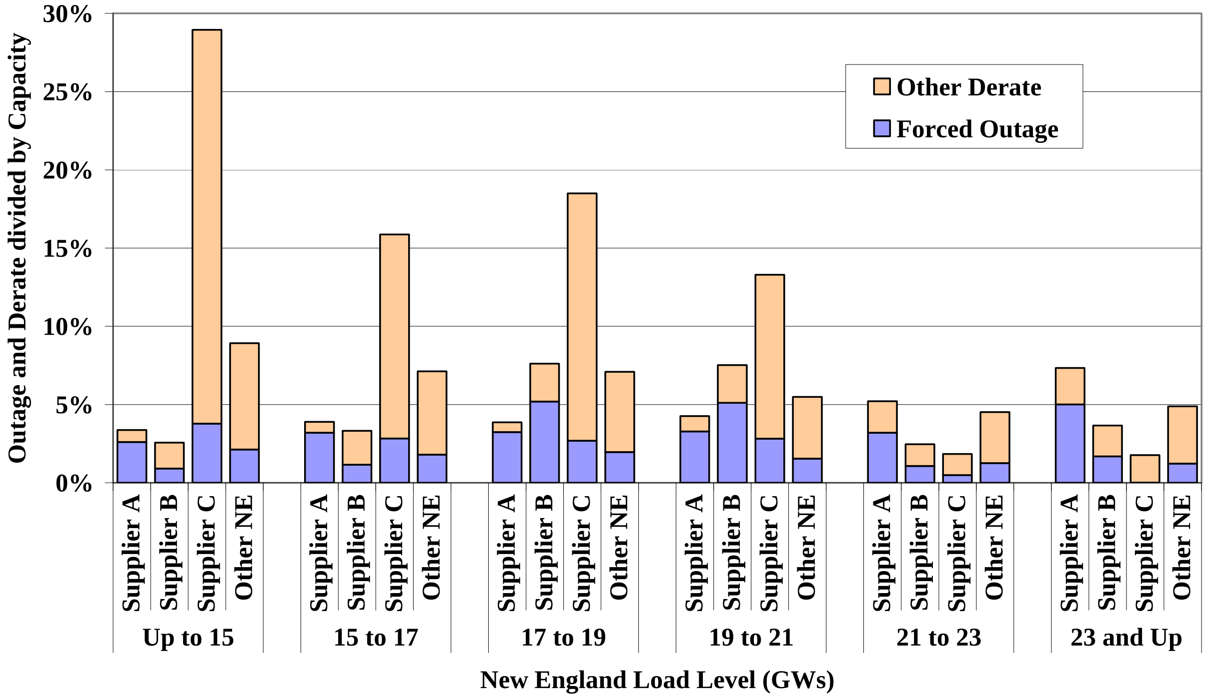


Figure 43 shows that the physical derating and forced outage quantities for the largest supplier and all other suppliers in Connecticut were moderate under all load conditions in 2009 and especially low during high load conditions. Hence, these deratings and outages do not raise concerns about physical withholding in Connecticut.

4. Potential Physical Withholding in All New England

Having analyzed each of the major constrained areas in New England, Figure 44 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 31 percent of the hours during 2009 (excluding capacity under reliability agreements). Suppliers B and C are the second and third largest suppliers in New England and were each pivotal during about 1 percent of the hours. All other suppliers are shown as a group for comparison purposes.

Figure 44: Forced Outages and Deratings by Load Level and Supplier
All New England, 2009



Supplier A and Supplier B exhibited rates of forced outages and other non-planned deratings that were moderate under all 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 21 GW, but were substantially higher at lower load levels, especially when load was less than 15 GW. Supplier C is also the largest supplier in Boston. The pattern for Supplier C was explained earlier 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. 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 2009. 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. Hence, ISO-NE should continue to monitor closely for potential economic and physical withholding in 2010, particularly in constrained areas as the reliability agreements expire.

In previous reports, we identified at least one supplier engaged in conduct to inflate its NCPC payments when it was needed for local reliability. Due to the decline in commitment for local reliability, such conduct was not a significant concern in 2009. However, to more effectively address this type of conduct in the future, the ISO filed modifications to the mitigation measures that were accepted by FERC in October 2009.

IX. Forward Capacity Market

ISO New England has had an installed capacity market since it began operations in 1998, but the original 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 with FERC and approved in 2006, established a new market mechanism to attract and maintain sufficient resources to satisfy New England's long-term resource planning requirements efficiently.

As discussed in more detail below, ISO-NE proposed an initial set of FCM reforms at the end of 2008 ("FCM Phase II") [104] and in June 2009, the Internal Market Monitoring Unit released a report addressing specific FCM design issues.[105] In response to issues raised in both the FCM Phase II filing and in the INTMMU report, ISO-NE filed revisions to the ISO-NE Tariff on February 22, 2010 to address the FCM design.[106] The Commission has conditionally approved these reforms, but established a paper hearing process to address a number of outstanding issues.

This section of the report provides an overview of the FCM design and evaluates the outcomes of the first three auctions. This section also discusses FCM reforms recently proposed by ISO-NE and conditionally accepted by FERC.

A. Background on the Forward Capacity Market

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

104 *ISO New England, Inc. and New England Power Pool*, Various Revisions to FCM Rules Related to Bilateral Contracts and Reconfiguration Auctions, Docket No. ER09-356-000 (December 1, 2008).

105 "Internal Market Monitoring Unit Review of the Forward Capacity Market Auction Results and Design Elements," ISO New England, Inc. Market Monitoring Unit (June 5, 2009).

106 *ISO New England, Inc. and New England Power Pool*, Various Revisions to FCM Rules Related to FCM Redesign, Docket No. ER10-787- 000 (February 22, 2010).

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 Cost of New Entry ("CONE"), is the estimated capacity revenues that would be necessary for the investment to be profitable.[107]

In an efficient market, the investments with the lowest Net CONE will be the first to occur. The capacity price should clear at a level that is higher than the Net CONE of the 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 to meet system planning requirements. The resulting clearing price provides a signal to the market of the value of capacity.

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

- *Installed Capacity Requirement* – The FCM procures the Net Installed Capacity Requirement ("NICR")[108] of the New England Control Area and the capacity judged necessary to achieve regional reliability standards in the Capacity Commitment Period, which begins three years after the auction.
- *Local Sourcing Requirement* – Before each auction, the existing[109] installed capacity in each zone, less retirement and export bids, are compared to the zone's Local Sourcing Requirement ("LSR").[110] 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

107 Cost of New Entry has a specific meaning in the context of FCM, which is defined in Market Rule 1, Section 13.2.4.

108 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.

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

110 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.

zones are always modeled in the auction. When the zonal requirements are modeled, the FCM produces locational prices that reflect the value of capacity in each zone.

- *Alternative Price Rule* – This provision is designed to set the clearing price at a more efficient level when Out-Of-Merit capacity sales (i.e., new capacity entry from resources selling below their costs) distort the outcome of the auction.
- *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.

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 de-list bids (the price below which a supplier will not sell its capacity) and export bids are subject to review by the market monitor prior to the FCA in order to address potential withholding by suppliers. 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

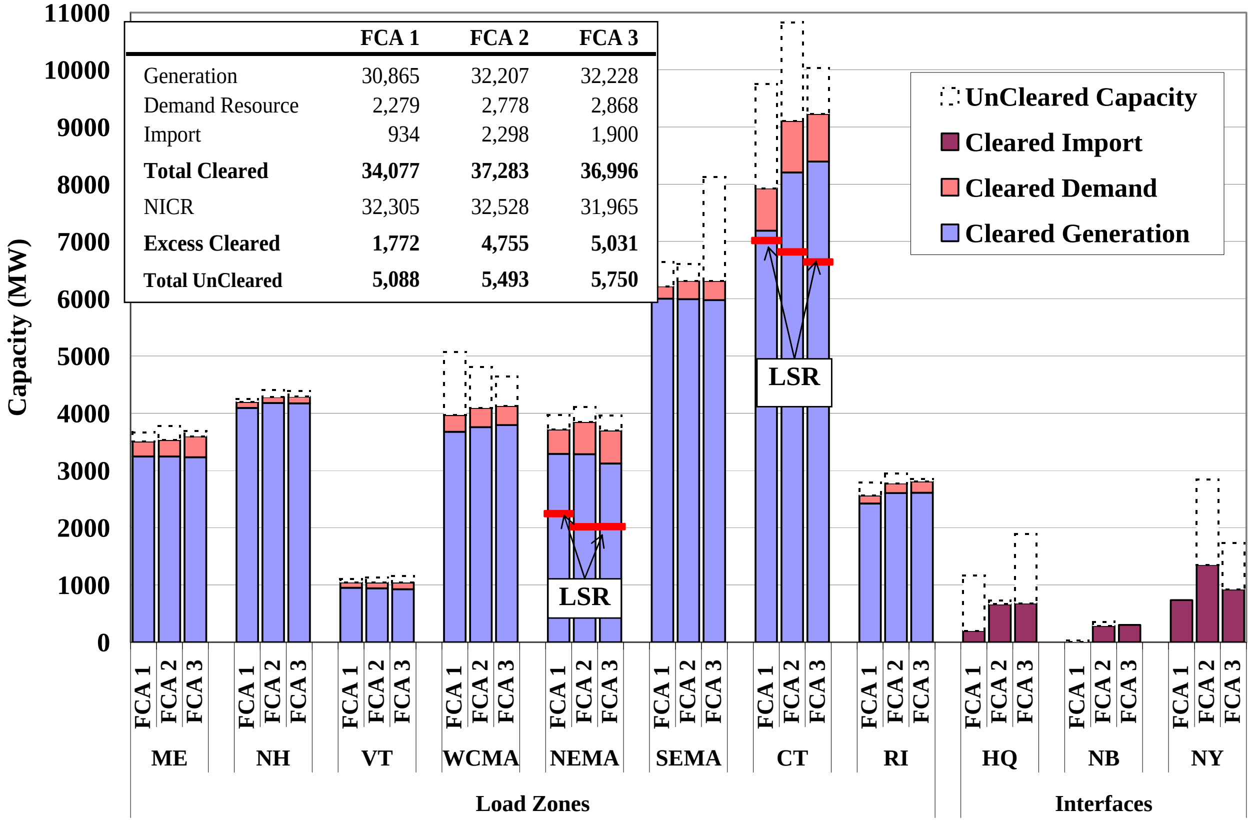
Three FCAs have been held to date: the first in February 2008 for the commitment period of 2010/2011 (“FCA1”), the second in December 2008 for the commitment period of 2011/2012 (“FCA2”), and the third in October 2009 for the commitment period of 2012/2013 (“FCA3”). In each auction, there was a substantial surplus of capacity over the NICR. Accordingly, each auction cleared at the floor price: $4.50 per kW-month in FCA1, $3.60 per kW-month in FCA2, and $2.95 per kW-month in FCA3.

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 all three auctions, but there was no price separation between Maine and the rest of New England. This section summarizes and evaluates the overall results of the first three FCAs, the de-list bids of existing suppliers, and the procurement of new capacity.

1. Summary of Capacity Procurement

Figure 45 summarizes the procurements in the first three 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.[111] The amounts of cleared resources are shown relative to the LSRs for Connecticut and NEMA and relative to the NICR for all of New England.

Figure 45: FCM Auction Clearing Summary by Location
FCAs 1, 2, & 3



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 1.8 GW. In FCA2, over 37 GW of resources were procured, exceeding the NICR by

111 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.

approximately 4.8 GW. In FCA3, nearly 37 GW of resources were procured, exceeding the NICR by more than 5.0 GW.

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, 2,300 MW in FCA2, and 2,600 MW in FCA3 for Connecticut; and
- 1,500 MW in FCA1, 1,800 MW in FCA2, and 1,700 MW in FCA3 for NEMA.

In each auction, a substantial amount of qualified resources did not clear. New proposed resources accounted for more than 80 percent of the uncleared capacity. 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, 99 percent in FCA2, and 101 percent in FCA3 (i.e., cleared generating resources alone exceeded NICR in FCA3) . 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, 9 percent of the NICR in FCA2, and 9 percent of the NICR in FCA3. 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, which 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 and 1,900 MW in FCA3.[112]

Substantial excess capacity cleared in the first three auctions as a result of the price floor that was originally stipulated in the Settlement Agreement. The price floor was originally supposed

112 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.

to be eliminated after FCA3, although this was extended until after FCA4 by the ISO's recent filing.

2. De-list 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 de-list bids to indicate they intend to de-list (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 de-list bid. The ISO reviews de-list bids and may reject them for reliability needs or in accordance with the mitigation rules.

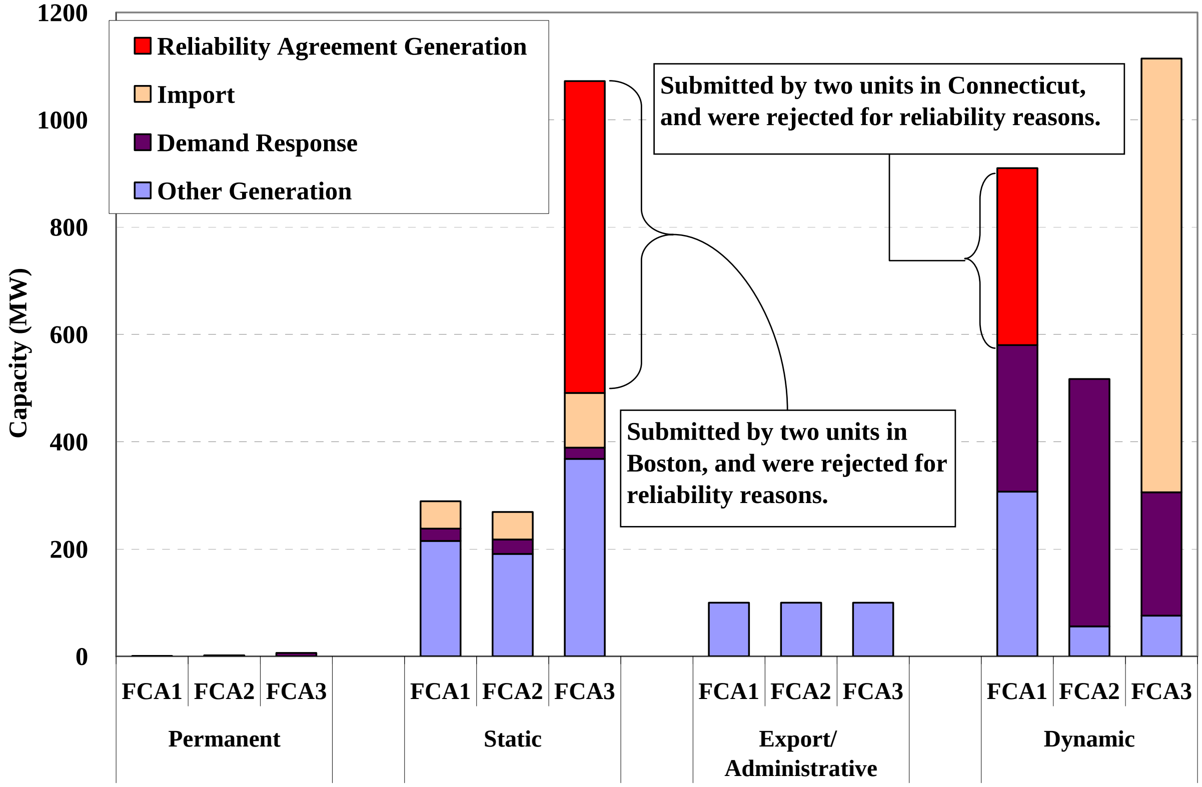
Figure 46 evaluates several categories of de-list bids in the first three FCAs. The figure shows four categories of de-list bids: permanent, static, export or administrative, and dynamic.[113] De-list bids are also separated according to the type of resource: generation not currently under a reliability agreement, generation currently under a reliability agreement, demand response resources, and imports.

Figure 46 shows that accepted de-list bids increased from an average of 930 MW in FCA1 and FCA2 to 1710 MW in FCA3. Dynamic de-list bids accounted for the majority of accepted de-list bids in all three auctions. Dynamic de-list bids are the only type of de-list bid that can be submitted during an auction. Other types of de-list bids must be submitted prior to the auction. De-list bids that are less than 80 percent of CONE are not subject to mitigation, while bids above

113 Each category of de-listde-list bid is defined in *Market Rule 1, Section 13.2.5.2*.

80 percent must be approved by the market monitor as consistent with the net going forward costs of the resource.[114]

**Figure 46: Summary of De-list Bidding by Type
FCAs 1, 2, & 3**



Internal generation accounted for approximately 64 percent of the de-listed capacity in FCA1, 39 percent in FCA2, and 32 percent in FCA3. In each auction, approximately 100 MW of capacity de-listed in order to support capacity exports to New York over the Cross-Sound Cable. In FCA3, two units in Boston and one unit in SEMA de-listed their entire capacity. Otherwise, most of the de-listed generation capacity was associated with small output ranges on individual units.

114 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*.

Four de-list bids were rejected in the first three auctions. Two bids associated with 330 MW of generation in Connecticut were rejected in FCA1, while two bids associated with 581 MW of generation in Boston were rejected in FCA3. All four de-list bids were rejected when the ISO determined in its Transmission Security Analysis that the units were needed for reliability in Connecticut or Boston. Since the rejected de-list bids were substantially smaller than the excess cleared capacity for all of New England in both cases, the auctions would have cleared at the price floor with or without the rejected bids and the decisions to reject did not affect the auction clearing prices in FCA1 or FCA3. However, the rejection of the de-list bids in Connecticut and Boston has highlighted an issue with the method of determining the LSRs.

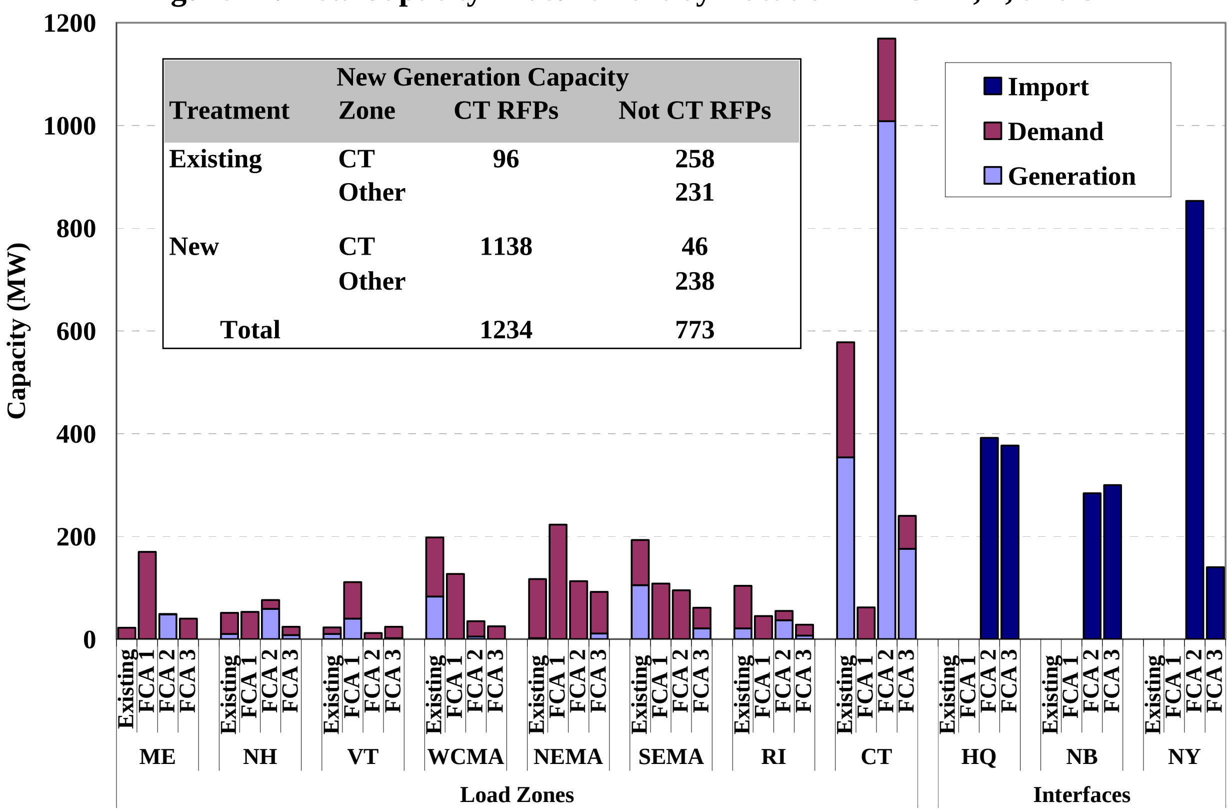
The Connecticut and Boston LSRs were much lower than the capacity requirements that were implied by the Transmission Security Analysis, which is the basis for rejecting de-list bids. As a result, de-list bids were rejected to protect Connecticut and Boston area reliability even though the Connecticut LSR was satisfied by nearly 1,200 MW in FCA1 and the Boston LSR was satisfied by over 1,600 MW in FCA3. In principle, markets should be always designed to satisfy the reliability needs of the system, which allows market prices to accurately reflect these needs. Accordingly, the ISO filed to modify the LSR criteria to be consistent with Transmission Security Analysis used to determine whether a de-list bid should be rejected for zone-level reliability. FERC approved the change for use beginning with FCA4.

Other than the two generators in Connecticut whose de-list bids were rejected, none of the generation that is currently under reliability agreements attempted to de-list in the first three auctions. During 2009, seventeen units (2,700 MW in total), were under reliability agreements. Although the majority of this capacity was retained in the first three capacity commitment periods without the use of out-of-market payments, one of the two units in Connecticut whose de-list bids were rejected in FCA1 will still receive out-of-market payments in order to cover the difference between the auction clearing price and their de-list bid price. It is positive that the use of out-of-market payments by the ISO to retain capacity will be greatly reduced beginning with the first capacity commitment period.

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 that 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. Figure 47 shows the amounts of new capacity that were procured in the first three FCAs by load zone or external interface. Capacity is divided by resource type: generation, demand response, and import capacity. We also distinguish the capacity based on whether it received existing treatment in FCA1 or it cleared in FCA1, FCA2, or FCA3.[115]

Figure 47: New Capacity Procurement by Location in FCA 1, 2, and 3



	New Generation Capacity		
Treatment	Zone	CT RFPs	Not CT RFPs
Existing	CT	96	258
	Other		231
New	CT	1138	46
	Other		238
Total		1234	773

---

115 Resources 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 de-listde-list bids rather than supply offers.

To determine whether new capacity entered due to the FCM revenue, the table in the figure identifies the quantity of capacity contracted under The Connecticut DPUC Request for Proposals (RFP) that may receive additional capacity payments beyond those from the FCM.[116]

In the first three FCAs, nearly 5,400 MW of new capacity was procured from generation, demand response resources, and imports.[117] The following discussion reviews and evaluates the procurements of new capacity by resource type that are shown in Figure 47.

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 the future if the floor price is no longer used. Similarly, the amount of capacity that de-lists in order to export may increase in the future if 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

116 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.

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

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.[118] The fact that other types of new resources must bear different obligations can inefficiently bias the 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 (2,007 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 the first three FCAs. 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 more than 1,200 MW of the new investment in generation received additional payments under the RFPs of the Connecticut DPUC and almost 500 MW are resources that received existing treatment, which indicates that their entry decisions were not contingent on the outcome of the 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.

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 284 MW of new generating resources

118 Peak Energy Rents are the revenues a generator earns in the real-time energy market during shortage events.

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 three 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 third Capacity Commitment Period exceeds the New England capacity requirement by nearly 16 percent. FCM has provided strong incentives for the sale of new capacity by demand response resources and importers. However, if the price floor is no longer used after FCA4, the amount of excess capacity purchased in the auction will be eliminated.

It is too early to determine whether the FCM will efficiently facilitate investment in new generation when it is needed. The prevailing surplus has caused the auction to clear at the floor price, which is well below most estimates of the Net CONE for new generation. Therefore, the market has not needed to facilitate investment in new generation resources.

4. Forward Capacity Auction Results – Conclusions

The Forward Capacity Market was successfully introduced by ISO New England in 2008 with no significant operational issues or procedural problems. The qualification processes and the auctions have occurred on schedule. 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 2013. 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 supplemental payments under the RFPs of the Connecticut DPUC. 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. These issues are discussed in greater depth in following subsection of this report.

Large quantities of demand response resources have entered at prices well below the Net CONE of new generation, which is a notable outcome of the first three 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) distorts investment incentives in favor of these resources. The ISO is planning to facilitate stakeholder discussions to design demand response programs to replace the current programs which are scheduled to expire in 2012. Hence, we recommend that the redesign should also address the inconsistency of the obligations of generation and demand response capacity resources.

C. Proposed Forward Capacity Market Reforms

1. Introduction and Summary

The Forward Capacity Market was proposed by ISO-NE and subsequently approved by the Commission in 2006. It established a market mechanism to attract and maintain resources to satisfy New England’s long-term resource planning requirements in an efficient manner. There has been active discussion and debate about reforming the FCM during 2009. At the end of 2008, ISO-NE proposed an initial set of FCM reforms (“FCM Phase II”).[119] In June 2009, the Internal Market Monitoring Unit released a report (“the INTMMU Report”) addressing specific FCM design issues.[120] In response to issues raised in both the FCM Phase II filing and in the

119 *ISO New England, Inc. and New England Power Pool*, (December 2008), Op. Cit..

120 ISO New England, Inc. Market Monitoring Unit (June 2009), Op. Cit..

INTMMU report, ISO-NE filed revisions to the ISO-NE Tariff on February 22, 2010 in order to amend the FCM design.[121]

Improving FCM price signals and the economic efficiency of the FCM has been the focus of the recent debate and FERC filing. In this regard, proposed changes to the FCM design have focused on two aspects:

- The treatment of "out-of-market" or "OOM" capacity, which is capacity that is offered into the market below the full cost of making such capacity available (this is usually the result of the supplier receiving supplemental revenue not generally available to others in the market); and
- The modeling of capacity zones, which allows the FCM to reflect the cost of reliability in transmission-constrained locations. Currently, these zones are formed only in certain circumstances.

The ISO-NE proposals made in February 2010 were directed at both of these market design issues. In the first instance, the ISO proposed to mitigate the price effects of OOM capacity by removing (at least partially) the OOM supply when setting prices in the Forward Capacity Auction ("FCA"). This involved changes to the Alternative Price Rule ("APR"). In the second instance, the ISO proposed to trigger the formation of capacity zones in a wider range of circumstances in order to improve the correspondence between FCA outcomes and reliability requirements.

We intervened in the ISO-NE proceeding at the Commission and provided comments,[122] which generally concluded that the ISO's proposed changes would increase the efficiency of the FCM.

---

121 *ISO New England, Inc. and New England Power Pool,* (February 2010), Op. Cit..

122 Our involvement in the proceeding was prompted, in part, by requests by the New England Power Generators Association ("NEPGA") and by a smaller group of New England generators (calling themselves the "Indicated Generators"), some of whom also belong to NEPGA. These groups requested that Potomac Economics, as the External Market Monitoring Unit (EMMU), conduct an independent analysis of the proposed amendments to the FCM rules in accordance with section 10.4.2 of the New England Power Pool ("NEPOOL") Participants Agreement. Under section 10.4.2, NEPOOL participants can request the EMMU to exercise its discretion to analyze and comment on important market-related issues. The NEPGA letter requested that we evaluate the proper design of the APR. The letter by the Indicated Generators asked that we evaluate the proposed changes to the modeling of zonal capacity and the associated pricing rules for such zones.

However, we explained that certain elements of the FCM design interfered with a fuller realization of its efficiency objectives. These related to both the APR and to modeling of capacity zones.

We generally found that the proposed revisions contained in the ISO's amendments would not fully satisfy this objective and we made specific recommendations regarding certain aspects of the APR changes.[123] Therefore, we supported a stakeholder process to develop long-term solutions in these areas, including market monitoring and mitigation provisions.

In April 2010, the Commission issued an order in the docket.[124] The Commission concluded that the rule changes have not been shown to be just and reasonable. However, because of the upcoming FCA in August 2010, the Commission accepted the proposed revisions and established a "paper hearing" to further evaluate them.

a. Summary of Recommendations

In the remainder of this section, we provide our analysis of the FCM design and our recommendations for improvement. We focus particular attention to those areas that are the subject the paper hearing ordered by the Commission. Hence, we address the Alternative Price Rule (Part 2), the modeling of capacity zones (Part 3), and the need for improved market power mitigation measures that should be considered in conjunction with other FCM reforms (Part 4).

Based on our analysis, we recommend the following:

- Modify demand response resources' obligations to be comparable to obligations of generation resources and imports.
- Revise the APR or any replacement provisions such that they more fully mitigate the price effects of OOM entry and do not treat rejected de-list bids as OOM.
- Permanently model the capacity zones in order to allow capacity prices to reflect local capacity requirements.

---

123 In particular, we explained our objection to certain elements of APR-1 and APR-3 and we recommended that the Commission not approve these two changes.

124 Order on Forward Capacity Market Revisions and Related Complaints, Docket ER10-787-000, *et al.* (Issued April 23, 2010).

- Modify market power mitigation measures to be effective given the changes in the market design.

b. Forward Capacity Market Design Principles

In evaluating the FCM design, it is important to establish an objective that is based on sound economic principles and to judge the market design in accordance with this objective. The objective of the FCM is to establish prices that will facilitate efficient decisions to invest in, retire, and maintain capacity in New England, as well as efficient decisions to import or export capacity to or from New England. Such prices should be set by the marginal cost of supplying capacity by the marginal supplier, whether the marginal supply is a new resource or an existing resource.

In our evaluation of the two main design issues identified above (OOM capacity and zonal modeling), we assess whether the existing market design meets this efficiency objective and, if not, what revisions may serve to advance the objective. As discussed in more detail below, the current market design (which includes the rule changes accepted by the Commission in April), does not fully satisfy this economic efficiency objective.

c. The Forward Capacity Auction

The FCA is a uniform-price, descending-clock auction. A participant with existing capacity may bid to "de-list" by indicating the minimum price they would accept to remain under the FCM obligations. New capacity makes offers into the FCM by indicating the price it requires to enter the market.[125] These FCM supply bids and offers are stacked in ascending order and the auction price is reduced until the total quantity of supply willing to sell at that price equals the forward capacity requirement (or until the auction price reaches the floor price). This price is paid to all remaining bids and offers. The auction results in an efficient allocation when all new capacity offers and all de-list bids are "competitive." The term "competitive" in this context means that offers for new capacity reflect the payment necessary to earn just enough in the auction

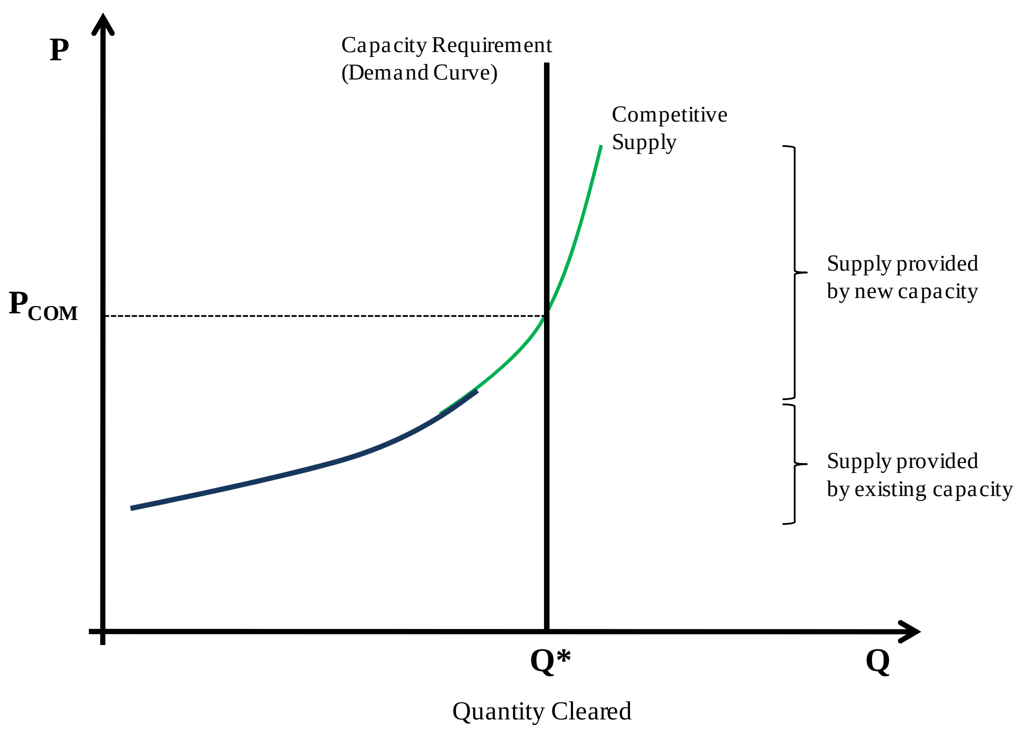
125 Both the de-list bids and new capacity offers essentially are offers to supply, but we refer to them as bids and offers in order to retain the language used to describe the FCM in the ISO-NE Tariff.

(recognizing expected energy and ancillary services revenues) to cover annualized entry costs. For existing units, a competitive de-list bid is one that will reflect the benefits and the cost avoided by de-listing the unit. Depending on the unit, this may be:

- The going-forward cost that can be avoided by retiring the unit minus the expected net revenues from the energy and ancillary services markets;
- The going-forward cost that can be avoided by temporarily shutting down the unit ("mothballing") minus the expected net revenues from the energy and ancillary services markets;
- The expected revenues of selling the de-listed capacity in the neighboring market; and
- The value of the Peak Energy Rent adjustment that is avoided by de-listing.

These competitive bids and offers result in an efficient outcome because they accurately reflect the marginal cost of satisfying New England's reliability requirements. Figure 48 shows a graphical depiction of the FCA.

Figure 48: Clearing the Forward Capacity Auction (Competitive Case)



In this figure, P_{COM} is the competitive price and Q^* is the cleared capacity. As the figure shows, the auction mechanism can be illustrated in the standard supply-demand framework. Demand is depicted as a vertical curve due to the fact that it essentially remains unchanged regardless of the price.[126] The supply curve is relatively flat for the initial quantities of capacity because many resources essentially participate in the market regardless of the clearing price, indicating that they will supply even at a relatively-low clearing price. The higher-priced green segment at the top of the supply curve represents the supply offers from new resources. The implicit (and reasonable) assumption is that new capacity is more expensive to supply than existing capacity. The intersection of the supply curve and the vertical demand curve establishes the clearing price.

2. Out-of-Market Capacity

a. Alternative Price Rule

The Alternative Price Rule is intended to mitigate the distorting effects of out-of-market ("OOM") capacity on the FCA prices. OOM capacity generally refers to the capacity offered in the FCM that earns OOM revenues through mechanisms that are not generally available to comparable units in New England, revenues from such sources as long run capacity contracts or special tax reductions. When a capacity resource enjoys access to these other revenue sources, the amount it will accept to enter the FCM will be lower than that required by a comparable entrant that has no access to such OOM revenues.

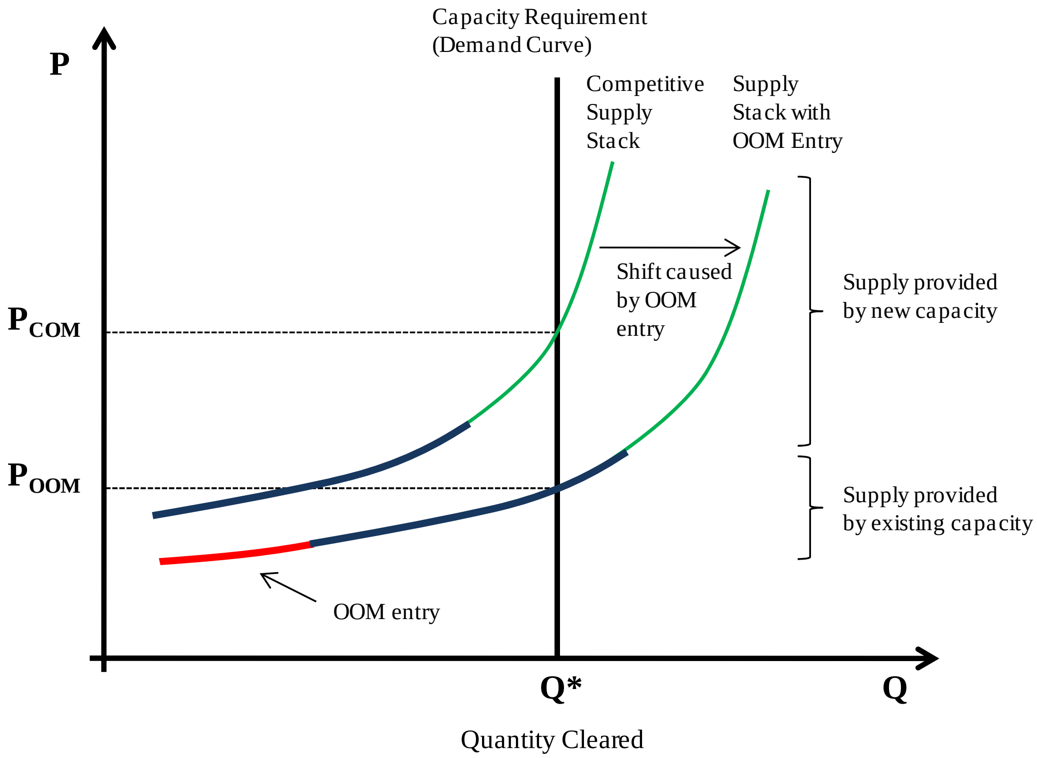
OOM capacity can lead to a clearing price in the FCA that is inefficiently low. This lowers capacity prices for all other new and existing resources, distorts price signals, and, therefore, adversely affects investment and retirement decisions in both the short-run and long-run.

- In the short-run, it can cause offers from new resources to not be accepted in the capacity market that would otherwise have been accepted and resulted in new capacity being built.
- In the long-run, the price effects of uneconomic entry create substantial risk related to future capacity revenues that would reduce the incentives of investors to enter the New England market.

126 We recognize that at sufficiently high prices, the demand curve is such that slightly less quantity may be demanded. However, there is no loss of the fundamental concepts by ignoring that subtlety.

These significant effects have been recognized by other ISOs as well -- both PJM and the New York ISO have capacity market rules to address OOM investment. The APR mechanism in New England is designed to address these effects by adjusting the forward capacity price to reduce the price effects of the OOM capacity. Figure 49 shows the effect of OOM entry on the supply curve and the FCA clearing price.

Figure 49: Clearing the Forward Capacity Auction (Case with OOM Entry)

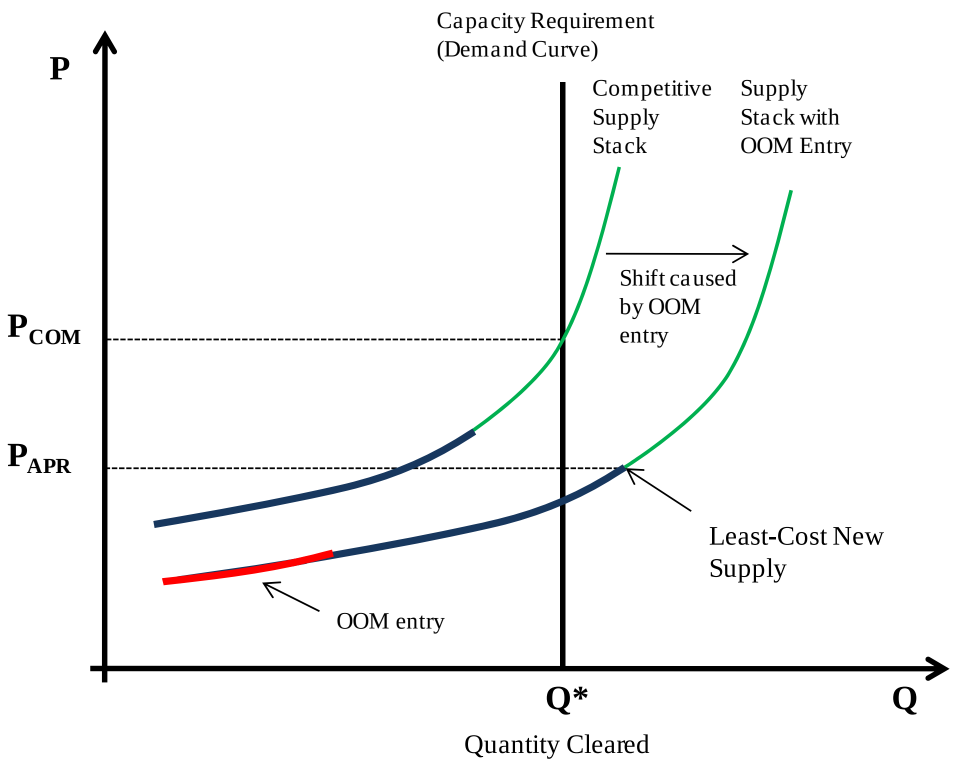


In the figure, the abbreviations are the same as in the previous figure, but we add P_{OOM} to represent the price after OOM entry. Because the OOM entry is likely to be at offers that are at the low end of the supply curve, the effect of OOM entry on the supply curve is represented as a shift to the right. From Figure 49 one can see that to negate the price effects of OOM entry, the auction price would need to be recalculated with only "in-market" resources (which would result in a price at P_{COM}).

The APR is generally designed to correct (at least partially) for the price impact of the shift in the supply curve shown in Figure 49. APR is invoked when: (1) *new capacity is needed in the FCM*, and (2) *OOM entry completely satisfies this new capacity need.*[127] In this case, a new resource would be expected to set the price in the FCA, but the OOM capacity would cause existing resources (or OOM resources) to set the price.

When APR is applicable, the clearing price is re-set to the lower of two values. The first value is the CONE, which is a value defined under the ISO-NE Tariff. CONE is derived from past FCA clearing prices when new capacity was needed. This methodology is intended to provide an accurate estimate of the net entry costs for new resources. The second value is the lowest offer among the group of offers by new entrants that was rejected in the auction. Figure 50 is a graphical representation to help illustrate the APR pricing rule.

Figure 50: Forward Capacity Auction Pricing Under the APR



127 This does not include the changes the Commission accepted in April 2010.

Figure 50 is the same as Figure 49, except that it shows the price that would be set under the current APR rules. As in Figure 49, the OOM entry is represented as a rightward shift in the supply curve because OOM resources are likely to be offered at the low end of the curve. Figure 50 also shows the "Least-Cost New Supply," which is the lowest-cost competitive entrant not selected in the auction. None of the competitive entrants cleared in the auction because the OOM entry is enough to fill the entire new capacity requirement. This means the lowest-cost competitive entrant is simply the lowest point on the top segment of the supply stack.[128] Because the APR chooses the lower of CONE and the offer of the lowest-cost competitive entrant not cleared in the auction, P_{APR} in the figure represents the maximum possible price under the current APR.

An important observation from Figure 50 is that the APR price (P_{APR}) will always be less than or equal to the price P_{COM}, which is the price that would prevail in the absence of OOM entry. Hence, APR does not restore competitive conditions (i.e., those that would prevail in the absence of OOM entry). Accordingly, it does not fully satisfy the efficiency objective introduced above. For this reason, we recommend the ISO revise the APR or replace the APR with provisions that would more completely mitigate the price effects of OOM entry.

In addition, the condition under which APR is triggered does not capture all the instances when OOM entry may inefficiently depress prices. Recall that these conditions are: (1) new capacity is needed in the FCM, and (2) OOM entry completely satisfies this new capacity need. However, OOM entry can inefficiently depress prices even when new capacity is not needed. It also can depress prices inefficiently when new capacity is needed and the OOM entry *cannot* satisfy the entire need. Accordingly, any revisions to the APR or subsequent provisions address the impacts of OOM entry (1) when new capacity is not needed and (2) when new capacity is needed but the OOM does not completely satisfy the need.

128 When all competitive entrants offer at costs above the highest-cost of existing capacity (as depicted in the figure), the lowest-cost competitive entrant not cleared will always be the lowest-cost competitive entrant in the auction. This is because APR is invoked when OOM capacity completely satisfies the new capacity need. If it happens that some existing capacity offers at higher costs than new competitive entrants, it may be that some competitive entrants clear in the market and the lowest-cost competitive entrant not cleared is not the lowest-cost competitive entrant in the FCM.

The revisions to the APR that were filed by ISO-NE and accepted by the Commission in April 2010 improved the APR by recognizing that OOM investments adversely affect the market for an extended period of time (rather than only the first FCA after they enter), and also by recognizing permanent de-list bids when determining whether new resources are needed. While these changes are improvements, they do not fully address the shortcomings of the current APR provisions that we discuss above.

The ISO also proposed to treat de-list bids that are rejected for reliability as OOM resources since they have a similar effect on the capacity clearing price. This approach would appear logical at first reading: The bid was not considered in the FCA pricing mechanism because the de-listing was initially accepted, meaning a lower offered supplier was marginal in the supply stack. When the ISO subsequently determined the unit is required for reliability and thereby "rejects" the de-list bid, it is reestablished in the stack for the purpose of setting the system-wide price. However, the bid was not rejected for the system-wide requirements, but only for a local reliability requirement that was not modeled accurately (or at all). If the local requirement had been modeled correctly, the de-list bid would not have been accepted by the market. Hence, the resource is not truly OOM and it is not correct to increase the system-wide price. Therefore, we have suggested that rejected de-list bids not be treated as OOM for purposes of the APR. Additionally, more complete modeling of the system's locational requirements would be beneficial as we discuss below.[129]

The ISO has acknowledged that additional work is needed on these provisions and the Commission has established a paper hearing to evaluate the FCM revisions proposed in February and to address the remaining FCM issues. We recommend that the ISO propose a substantially revised APR that would more completely mitigate the price effects of OOM investment. One

129 There is only one case we can identify where such a capacity price adjustment would be efficient. That is when the ISO uses discretion to reject a de-list bid that would not have been rejected, even if the reliability requirement had been fully modeled. This could happen if the ISO is conservative and rejects the de-list bid in response to a potential risk, such as the risk that a planned investment will not be completed on schedule. In that case, the well-functioning market would have allowed the resource to de-list and the ISO's rejection of the de-list bid artificially increase the supply and lower prices. It seems unlikely that a large share of the rejected de-list bids would be the result of this type of discretionary determination. We recommended that, if the ISO believes that this will be the case, APR-3 could be scaled-back so that it only applies under these circumstances.

way to do this is to establish an offer floor that would prevent resources from offering substantially below their true costs of entry. The New York ISO utilizes such an approach to mitigate uneconomic entry in New York City. This approach would cause the OOM resource to not clear in the capacity market and, therefore, to receive no capacity revenue. This is a more effective deterrent than the current APR in New England because the OOM resource would still be able to sell its capacity in the FCM. If ISO-New England continues to support an APR that would allow the OOM resource to sell its capacity, it may wish to consider provisions that would pay the OOM resource and other new entrants the "unadjusted" price. This would lower capacity costs and would prevent the higher APR-adjusted price from provoking investment in new resources that are not needed.

3. Zonal Capacity Modeling

As introduced above, the FCM allows capacity zones to form within New England to account for reliability needs that require a minimum level of capacity to be located within the zone. Capacity zones are established only when a projected shortage of capacity exists in one of the pre-defined capacity zones. This determination is made *prior* to the FCM auction. When a capacity zone is modeled, the zonal clearing price may exceed the clearing price in the rest of the market if higher-cost resources within the zone are needed to satisfy the zonal requirement. If a zone is not formed based on this *ex ante* accounting, and a local reliability need requires high-cost resources to provide capacity to the zone, the ISO may have to "reject" de-list bids from these resources.

The overall object of the FCM market is to facilitate efficient entry and exit of capacity to meet New England's capacity requirements. This efficiency objective includes market requirements that reflect the reliability needs of the system to the maximum extent that is feasible. This consistency between the market requirements and reliability requirements is important because it will cause the market to produce price signals that reflect the costs of satisfying these requirements and ensure that the requirements are satisfied over the long-term.

Issues arise in the current FCM when capacity zones are not always modeled when they may be warranted. This has been evidenced in the early FCAs by the fact that the ISO has rejected de-list bids in both the Connecticut and Boston areas when the zones were not modeled. The primary reason the capacity zones are not always modeled is related to market power. When a capacity zone is modeled, certain suppliers within the zone may be pivotal and have market power in the zone because they do not face competition from out-of-zone resources. Therefore, the supplier with market power may be able to raise prices, subject to the current mitigation measures. Therefore, one could conclude that the triggering criteria for capacity zones under the current FCM rules primarily serve a market power mitigation function.

In its February 2010 filing, the ISO's proposed amendments that broaden the range of circumstances when a capacity zone will be triggered and thereby improve the consistency between the capacity market and the reliability requirements. However, there will remain instances when a capacity zone may not form even when there is a need for local resources to meet reliability at a cost higher than the market-wide clearing price. This can occur when the FCA results in a price that clears de-list bids within the zone during the auction at a level sufficient to create a need to reject some of these bids in order to satisfy reliability. This will result in price distortions as discussed above.

To fully meet the efficiency objective, we recommend that the ISO permanently model capacity zones that have local sourcing requirements in the FCM. As discussed more below, this may raise market power concerns that are not fully addressed by the current mitigation measures. However, we would generally recommend improving the mitigation measures as necessary, rather than mitigating market power by adjusting the market design (i.e., by not always modeling the zones).

4. Enhanced Market Power Mitigation in Local Capacity Zones

Market power concerns underlie the reluctance to completely harmonize the FCM with the reliability requirements of the system. For example, because modeling a capacity zone can raise market power concerns, the FCA excludes de-list bids from pivotal suppliers in the revised *ex ante* test that triggers the modeling of a zone. Similarly, some of the APR provisions are influenced by market power concerns. The following are relevant questions in evaluating these market design choices:

- Are the existing mitigation measures adequate and effective?
- If not, is it preferable to modify the mitigation measures to make them effective or to adjust the design of the market to address the market power concerns?

To begin to answer these questions, we review the primary mitigation measures in the current FCM rules. The most important provision is the review of bids by the Internal MMU that begins at 80 percent of CONE. Dynamic de-list bids must be less than 80 percent of CONE and other de-list bids must be reviewed by the ISO if they exceed 80 percent of CONE.[130] This rule establishes a safe-harbor for de-list bids at or below 80 percent of CONE. To evaluate the effectiveness of the mitigation measures, a principal issue to consider is whether this 80 percent level allows suppliers with market power too much latitude to raise prices. We believe that this 80 percent threshold is too high to be fully effective in mitigating the substantial market power that likely exists in the local capacity zones. Further, the standard of review for de-list bids above 80 percent of CONE may not be fully effective in requiring that the de-list bids be competitive. It is not surprising, therefore, that the market design has been altered to more fully mitigate this market power.

We believe that wherever possible, market power concerns should be addressed by improving mitigation measures rather than by altering the market design. Altering the market design to address market power frequently will result in unnecessary economic inefficiencies.

The New York ISO experienced similar market power issues in its New York City capacity zone. In the NYISO capacity market, the New York City zone is always modeled in accordance with existing reliability requirements. Significant market power issues have arisen because supply is concentrated and there is more than one pivotal supplier in the zone. However, the NYISO implemented market power mitigation that has been fully effective in addressing the market power concerns and ensuring competitive outcomes. Hence, no compromises in the design of the market or in the market rules were necessary.

130 *Ibid.*, p.21.

Because the current mitigation measures are not likely to be fully effective in mitigating the market power, we recommend that the ISO and its stakeholders address these market power mitigation issues as they consider the long-term changes that are warranted in the APR and zonal modeling provisions. It is important that these issues be addressed together to ensure that the overall FCM framework will lead to competitive and economically efficient outcomes.

If the ISO accepts the invitation by the Commission to find a consensus with its stakeholders on these issues, it is important to introduce market power mitigation measures that allow suppliers to submit de-list bids that fully reflect their marginal costs of supplying capacity. For existing units, these marginal costs include:

- Going-forward costs that can be avoided by retiring or permanently de-listing the unit minus the expected net revenues from the energy and ancillary services markets;
- Going-forward costs that can be avoided by temporarily shutting down (“mothballing”) minus the expected net revenues from the energy and ancillary services markets; and
- Expected revenues of selling the de-listed capacity in the neighboring market;

However, it is important for the market power mitigation measures to recognize that these types of costs do not apply in the same manner to all existing units. For example, the going-forward costs associated with a retirement are only applicable for resources that are likely to be retired. For most units that are economic to remain in operation, the relevant costs would be limited to opportunity costs of exporting the capacity and the economic value of the Peak Energy Rent adjustment (and other FCM obligations). The ISO should consider these issues as it considers further changes to the APR and zonal modeling provisions.

5. Conclusion

This section addressed two main areas of concern relating to the FCM design: APR provisions addressing OOM capacity and the provisions governing the formation of capacity zones. A general efficiency objective based on economic principles guides the assessment of FCM. In general, the FCM design (including the Commission-accepted revisions) does not fully meet this objective. Therefore, we support the Commission’s mandate for ISO-NE to develop long-term solutions in these areas, as well as monitoring and mitigation provisions applicable to the FCM in this process and the resulting filing.