2010 Assessment of the Electricity Markets in New England

Prepared By:

POTOMAC ECONOMICS

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

External Market Monitor
ISO New England Inc.

June 2011

Table of Contents

List of Figures

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.
PER	Peak Energy Rent
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 2010 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 2010.

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."

Executive Summary

This report assesses the efficiency and competitiveness of New England's wholesale electricity markets in 2010. 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 2010, this report includes findings in two primary areas: the competitive performance of the markets and the operational efficiency of the markets. The broad findings in each of these areas are discussed below.

Competitive Performance of the Markets

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 2010. Although structural analyses indicate potential market power under some conditions in some areas, our assessment raised no competitive concerns associated with suppliers withholding resources to raise prices. Energy prices rose 20 percent from 2009 to 2010, due primarily to increases in fuel prices and load levels as well as reductions in surplus capacity. Natural gas prices increased 10 percent in 2010 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, higher fuel costs translate to higher offer prices and market clearing prices in a well-functioning, competitive market. The correspondence of fuel prices and offer prices in New England is an indication of the competitiveness of ISO New England's markets.

However, energy prices increased by more than fuel prices in 2010 due to higher load levels and lower levels of surplus capacity. Surplus capacity is the amount of online and quick-start resources available in the real-time market in excess of the energy and operating reserves demand. Surplus capacity fell because the ISO's supplemental commitments of resources for local reliability decreased significantly in 2009 and into 2010 due to transmission improvements. Average load increased by 3 percent and the peak load increased by 8 percent in 2010, primarily due to hotter summer weather and improved economic conditions. Accordingly, electricity

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 in a timeframe short enough to preclude expanding, retiring or converting the assets to another use. These costs include any foregone opportunity costs of producing such output. 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.

prices rose as higher demand required generating resources with higher marginal costs to operate more frequently, particularly during the summer months.

Finally, the system generally operated with substantially less surplus capacity in 2010 than in 2009. Surplus capacity is the amount of capacity that is online or capable of starting quickly in excess of the amount required to satisfy load and reserve requirements. The markets will naturally produce some surplus capacity, but this surplus will rise when the ISO must commit additional resources to satisfy local reliability requirements. Such "supplemental commitments" decreased markedly in late 2009 and 2010 due to transmission upgrades in New England, which resulted in a reduction in the daily minimum surplus capacity levels of more than 40 percent. Operating the system more tightly with lower levels of surplus capacity contributed to higher prices and increased price volatility, but substantially less Net Commitment Period Compensation ("NCPC"). These changes resulted in prices that more accurately reflected supply and demand conditions, which marks a significant improvement in the performance of the ISO New England markets.

Operational Efficiency of the Markets

Efficient real-time prices are a critical priority because they

- Provide incentives for market participants to operate in a manner that maintains reliability at the lowest overall cost;
- Facilitate efficient day-ahead scheduling and resource commitment; and
- Contribute to efficient investment in supply and demand response resources in the long term.

We find that both the day-ahead and real-time markets operated relatively efficiently in 2010 as prices appropriately reflected the effects of higher fuel prices and load levels. However, we also find that real-time prices often do not fully reflect the cost of satisfying demand and maintaining reliability during tight market conditions. Our recommendations in this report address this issue.

Upgrades to the transmission system in Connecticut and Southeast Massachusetts were completed in mid-2009, leading to significant changes in market operations. The upgrades sharply reduced the need for the ISO to commit generation for local reliability. Such commitments can often lead to significant surplus capacity in real time, which tends to depress energy and ancillary services prices in the real-time market. Accordingly, the reduced level of surplus capacity from 2009 to 2010 led real-time prices to be more consistent with underlying supply and demand conditions.

However, the effects of the decrease in supplemental commitments for local reliability were partly offset by an increase in supplemental commitments to satisfy system-wide (i.e., non-local) reserve requirements, which rose considerably after April 2010. The increase in these commitments was due to an extended outage of a large fast-start hydroelectric resource that substantially reduced the amount of reserves available in most hours (because it could provide a large quantity of reserves), as well as the reduction in commitments for local reliability that had previously contributed to higher levels of surplus capacity.

In addition to the changes in patterns of supplemental commitments made to maintain reliability in the short-term, there were significant changes in the longer-term reliability agreements that had been used to ensure that units needed for reliability remained in operation. Although such agreements had been necessary, they were poor substitutes for transparent market prices and did little to facilitate efficient investment. All remaining reliability agreements were terminated after the first Forward Capacity Commitment Period began on June 1, 2010. The Forward Capacity Market ("FCM") is designed to ensure that resources needed to maintain reliability remain available.

The reduction in supplemental commitments for local reliability and the expiration of the reliability agreements led to sharp reductions in uplift costs, while more frequent supplemental commitment for system-wide reliability partially offset these reductions. Taken together, these factors led to an overall decrease in uplift charges from $387 million in 2008 to $140 million in 2009 and $117 million in 2010.

Recommendations

Overall, we conclude that the markets performed competitively in 2010 and were operated well by the ISO. Based on the results of our assessment, however, we offer nine recommendations to further improve the performance of the New England markets. Seven of the nine were also recommended in our 2009 Annual Assessment. This overlap is expected since many of the recommendations require substantial resources and must, therefore, be prioritized with the ISO's other projects and initiatives. Most of these seven 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. A table of recommendations can be found at the end of this executive summary.

The following sections summarize our findings.

B. Energy Prices and Congestion

Average real-time energy prices increased 20 percent from approximately $44 per MWh in 2009 to $53 per MWh in 2010. This was due primarily to:

- Increases in fuel prices – Natural gas prices increased 10 percent from the prior year, which is important because natural gas-fired resources are most frequently on the margin in New England.
- Higher load levels -- Average load increased by 3 percent and peak load increased by 8 percent in 2010 due to improved economic conditions and hotter summer weather.[5]
- Lower surplus capacity levels -- The daily minimum surplus capacity level fell by more than 40 percent in late 2009 and 2010 as the ISO reduced its supplemental commitments for local reliability.

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

An outage of a large, fast-starting, flexible resource that lasted for nearly seven months (including the summer) contributed to tighter overall supply conditions and associated price increases.

Congestion and Financial Transmission Rights

In 2010, New England experienced very little congestion into historically-constrained areas, such as Boston, Connecticut, and Lower Southeast Massachusetts ("Lower SEMA"), as a result of transmission upgrades that have been made in recent years. Most of the price separation between net exporting regions and net importing regions was due to transmission losses, rather than to transmission congestion.

The 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 and $38 million in 2010. The modest increase in congestion revenue in 2010 was primarily due to: (i) higher fuel prices, which increased redispatch costs and associated congestion-related price differences during periods of congestion; and (ii) significantly higher load levels that generally increased the amount of power flows into high-load areas. Congestion revenue is used to fund the FTRs sold by ISO New England.

The ISO operates annual and monthly markets for FTRs, which allow participants to hedge the congestion and associated basis risk between any two locations on the network.[6] 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:

- In 2010, annual FTR prices generally over-estimated the congestion that prevailed in the energy market. Monthly FTR prices were more consistent with congestion patterns, which is to be expected due to additional information that becomes available regarding system conditions.

6 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).

- From 2009 to 2010, the consistency of FTR prices and congestion improved substantially overall.
  - ✓ These results suggest that market participants underestimated the decrease in congestion associated with the addition of transmission capability in 2009.
  - ✓ As market participants observed less congestion in 2009 and 2010, expectations changed accordingly, and the annual and monthly FTR prices converged more closely to day-ahead congestion levels in later periods.

The congestion revenue of $38 million collected by the ISO in 2010 was sufficient to fully fund the target value of the FTRs.

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 2010, indicating good overall convergence. However, average real-time prices were almost two percent higher than average day-ahead prices, which is unusual since electricity markets typically exhibit slightly higher day-ahead prices. The higher real-time prices are primarily attributable to:

- The reduction in surplus capacity during real-time operations discussed above. This led to more frequent high real-time price events than expected by the day-ahead market; and
- The allocation of NCPC charges to virtual load (which would otherwise have a strong incentive to buy at the lower day-ahead price and sell at the higher real-time price) likely inhibited the natural market response to the sustained real-time price premiums.

Uplift Allocation and Virtual Trading

Virtual trading plays an important role in overall market efficiency by improving price convergence between day-ahead and real-time markets, thereby promoting efficient commitment and scheduling of resources in the day-ahead market. Virtual trading in the day-ahead market consists of purchases or sales of energy that are not associated with physical load or physical resources. Since no physical energy will be supplied or consumed in real time, virtual transactions scheduled in the day-ahead market are settled against real-time energy prices and are only profitable when they contribute to price convergence between the two markets.

ISO New England allows virtual traders to schedule transactions at every pricing location. This includes individual nodes and more aggregated locations, such as the New England Hub and load zones. Virtual transaction quantities at individual nodes decreased sharply in May 2010. This was due primarily to the correction of a day-ahead modeling inconsistency that allowed virtual transactions to earn sustained profits at a small number of nodes. In contrast, virtual transactions at hubs and zones increased in 2010 in response to larger differences between day-ahead and real-time prices.

The reduction in nodal virtual trading volumes and the increase in supplemental commitments together led to significant changes in the allocation of uplift costs (i.e., NCPC costs) to virtual transactions. Under the current tariff, NCPC charges associated with supplemental commitments for system-wide needs (known as "Economic NCPC") are allocated to "real-time deviations" between day-ahead and real-time schedules.[7] This allocation assigns NCPC charges to transactions that actually tend to *reduce* the need for supplemental commitments, including virtual load. Therefore, as supplemental commitments for system-wide needs increased in 2010 and led to higher Economic NCPC, and the deviations associated with nodal virtual transactions decreased, the per MWh allocation of Economic NCPC to virtual transactions increased from an average of $0.68/MWh in 2009 to $2.10/MWh in 2010, reaching a high of $3.60/MWh from

7 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, virtual load schedules, and virtual supply schedules.

May to December. This increased allocation to virtual transactions likely hindered the day-ahead market's natural response to transitory price differences between the day-ahead and real-time market.

NCPC charges are caused by factors other than deviations, such as peaking resources not setting prices, congestion, system reliability needs, and outages. We find that the current allocation scheme over-allocates costs to deviations relative to the portion of the NCPC they likely cause. This is particularly true of virtual load transactions that tend to increase day-ahead commitments and, therefore, decrease the need for supplemental commitments. Hence, we recommend that the ISO modify the allocation of Economic NCPC charges to be in accord with a "cost causation" principle, which would generally involve not allocating NCPC costs to virtual load and other real-time deviations that cannot reasonably be argued to cause real-time economic NCPC. We are working with the ISO to develop a consensus approach to revising the NCPC allocation that would address this issue and improve the incentives for efficient day-ahead scheduling by market participants.

C. Reserve and Regulation Markets

The ISO operates a forward reserve market where reserves are procured in seasonal auctions, a real-time regulation market, and a real-time reserve market where reserves are scheduled with local requirements and co-optimized with the real-time energy market. These 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.

Real-Time Reserve Market Results

Reserve clearing prices rose, but remained relatively low in the real-time market in 2010. Clearing prices were highest for 10-minute spinning reserves ("TMSR"), averaging approximately $1.75 per MWh, up from $0.70 per MWh in 2009. The increases were related to the substantial reduction in average surplus capacity as supplemental commitments for local reliability decreased sharply since the second half of 2009. As expected, this led to higher operating reserve clearing prices associated with the tighter supply conditions. However, these prices remain well below the prices for comparable products in other markets.

The ISO has local reserve zones in Boston, Southwest Connecticut, and Connecticut, but real-time reserve prices did not vary substantially by location. This is largely due to transmission upgrades that have significantly reduced local requirements 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 which 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 operating reserves for All of New England, Boston, Connecticut, Southwest Connecticut, and Rest of System.[8]

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 2010, prices cleared below the $14 per kW-month price cap in Connecticut due to transmission upgrades that reduced the quantity of reserves procured from internal generation. Clearing prices outside Connecticut trended down as a result of lower offer prices in the forward reserve auctions. 95 percent of the resources assigned to satisfy forward reserve obligations in 2010 were fast-start resources capable of providing offline reserves.

The ISO may wish to consider the long-term viability of the forward reserve market for several reasons. First, it has not achieved one of its primary objectives, which was to lower NCPC by purchasing forward reserves from high-cost units frequently committed for reliability. Second, 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). Third, the Locational Forward Reserve Market is largely redundant with the locational requirement in the Forward Capacity Market.

8 The Rest of System 30-Minute Operating Reserves ("TMOR") purchase requirement has been eliminated beginning with the Summer 2011 Procurement Period.

Regulation Market

The regulation market performed competitively in 2010, with an average of approximately 740 MW of available supply competing to serve an average of 70 MW of regulation demand.[9] The significant excess supply generally limited 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 $23 million in 2009 to $14 million in 2010, largely due to the decrease in the regulation requirement, which fell roughly 25 percent on average from 2009 to 2010. Natural gas-fired combined cycle generators, which provide most of the regulation service in New England, were committed more frequently in 2010 due to increased load levels and reduced hydroelectric generation production. This generally increased the supply of low-priced online regulation and contributed to lower regulation expenses.

D. 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, with net import levels averaging 1,380 MW during peak hours and 910 MW during off-peak hours in 2010. This is characteristic of the efficient management of hydroelectric resources, whereby the largest imports are made in periods with the highest prices. However, the seasonal variations in net imports in 2010 were less consistent with typical patterns for hydro resources. Sales from Hydro Quebec were higher in the spring of 2010 when electricity prices were relatively low than in the summer of 2010

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

when prices were much higher. This pattern likely reflects that load levels were higher and reservoir levels were lower than expected in Quebec during the summer of 2010.

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 335 MW during peak hours in 2010), while participants schedule power flows in both directions on the larger interface depending on the relative prices. In 2010, New England exported an average of 255 MW to New York across the larger interface during peak hours and 105 MW 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. This has caused large real-time price differences to frequently occur between New York and New England, even when the interfaces are not fully utilized.

In July 2010, ISO New England and New York ISO commenced a joint effort known as the Inter-Regional Interchange System ("IRIS") initiative to address the problem of inefficient scheduling between the two markets. We employed simulations to estimate the benefits of the two proposed solution options in this initiative: 1) Tie Optimization; and 2) Coordinated Transaction Scheduling. The Tie Optimization proposal performed slightly better in our simulations than the Coordinated Transaction Scheduling proposal. The simulations indicated that consumers in both markets collectively would have saved $140 to $145 million per year using either proposal. These estimates of savings are likely to be conservative and would be larger under tighter supply and demand conditions over the long term.

ISO New England and the New York ISO are also planning to implement Market-to-Market Congestion Management Coordination, which would develop procedures for enabling one ISO to re-dispatch its internal resources to relieve congestion in the other control area (or to limit the effect of the largest generation contingencies on the other control area).[10,11] The benefits of this second initiative are substantially lower than the benefits of the IRIS initiative. Hence, we recommend that ISO-NE and the New York ISO place the highest priority on developing and implementing one of the two alternatives under the IRIS initiative.

E. 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-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. We evaluated five aspects of the real-time market related to pricing and dispatch in 2010 and have the following conclusions and recommendations:

1. *Real-Time Pricing of Fast-Start Resources:* 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. The significance of this issue has grown recently because the use of fast-start units has increased due to the decline in surplus capacity.

   ✓ 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.

2. *Real-Time Pricing during Forecasted and Actual Shortages of Operating Reserves*: The marginal cost of meeting system-level 30-minute reserve requirements can exceed the ISO’s $100 per MWh Reserve Constraint Penalty Factor (“RCPF”), requiring the ISO to curtail exports and take other manual actions outside the market. This can lead to

10 This will be presented to stakeholders in 2013.

11 When the Central-East Interface in New York is constrained, New England is required to reduce imports from Quebec, generation from the Seabrook nuclear unit, and/or generation from the Mystic generating plant. This is because a large sudden generation contingency in New England leads to a significant increase in flow across the Central-East Interface.

inefficiently low real-time prices that do not properly reflect the cost of maintaining reliability.

- ✓ We recommend that the ISO perform an evaluation to improve the consistency between its operating procedures and the 30-minute reserve RCPF, which should result in modified operating procedures or a higher RCPF for system-level 30-minute reserves. This will improve market participants' incentives to schedule resources that efficiently maintain reliability.

3. *Real-Time Pricing during Demand Response Activation:* Participation in real-time demand response programs has surged in New England, from 530 MW in January 2006 to 2,300 MW in January 2011. 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 when emergency 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 would have been a shortage without the activation of demand response resources.

4. *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: (i)creates a small upward bias in real-time prices in most areas; and (ii)sometimes distorts the value of congestion into constrained areas.

   - ✓ We recommend that the ISO consider modifying the inputs from UDS to the ex post pricing model to improve the consistency of the ex post and ex ante prices.

5. *Price Corrections*: We find that price corrections were very infrequent in 2010, which reduces uncertainty for market participants transacting 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.

F. 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 be undertaken only when necessary. In this section, we evaluate several aspects of the ISO's operations and processes for satisfying reliability requirements in 2010.

Accuracy of Load Forecasting

The day-ahead load forecast is important 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 relatively accurate and generally superior to comparable results in other RTO markets.

Supplemental Commitment for Local Reliability

Supplemental commitment for local reliability was low by historical standards. The average amount of committed capacity decreased from a daily average of approximately 1,000 MW in 2008 to 300 MW in 2009 and 180 MW in 2010. Such commitment declined sharply in Connecticut and Lower SEMA, primarily due to the effects of transmission upgrades into both areas completed by mid-2009.

Decreased commitment for local reliability has also contributed 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 prior to the transmission upgrades in mid-2009 to 760 MW from mid-2009 to the end of 2010. 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. Both types of commitments can depress real-time prices inefficiently, while supplemental commitments by the ISO also lead to increased uplift costs.

Recent transmission upgrades have substantially reduced the need for the ISO to commit generation to satisfy local reliability requirements since July 2009. However, the ISO must still make commitments to satisfy New England's system-wide reliability requirements. Our evaluation indicates that supplemental commitments to meet the system's overall capacity needs increased significantly after May 2010. This was largely due to the lower levels of surplus capacity discussed above, as well as the outage of a large, fast-start, hydroelectric resource. These two factors together increased the need for the ISO to make supplemental commitments to satisfy system-wide reliability requirements. After reviewing the supplemental commitments and the surplus capacity levels that resulted from real-time operating conditions, we found that roughly 65 percent of the supplemental resource commitments in 2010 were actually needed to maintain system level reserves in retrospect.[12] This is not surprising because resource commitments are "lumpy" (the market cannot commit exactly the quantity it needs and commitment decisions are often made well in advance when there is significant uncertainty regarding the necessity of the supplemental commitments.

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. This uncertainty was substantially increased in 2010 due to the unavailability of the largest fast start resource. Nonetheless, we have identified some areas for

12 This is a simple evaluation that treats any surplus capacity (online and available offline capacity less than needed to meet system load and reserve requirements) as "not needed" for the system. This simple evaluation tends to understate the necessity of supplemental commitments because: 1) the evaluation is based on hourly integrated peak rather than the higher instantaneous peak, and 2) the ISO cannot commit just a portion of a unit. For example, suppose the ISO needs an additional 200 MW of capacity to satisfy system reliability needs and commits the most economic unit with a capacity of 300 MW. In this evaluation, 100 MW of capacity would be deemed as "not needed".

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 local reliability and system-wide reserve requirements fell from 208 MW in 2008 to 60 MW in 2009, and then rose modestly to 74 MW in 2010. 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, and then rose to 210 MW in 2010. The decrease from 2008 to 2009 was consistent with the reduction in local reliability commitments during this period, while the increase from 2009 to 2010 was attributable to the increase in supplemental commitments for system-wide reserve requirements. The reductions in the amounts of surplus capacity and out-of-merit generation in 2009 and 2010 compared to prior years are positive developments because they indicate that the demand for energy and reserves are being satisfied more efficiently and producing more efficient price signals.

Uplift Charges

Overall, uplift charges fell significantly from $387 million in 2008, to $140 million in 2009 and $117 million in 2010 for two reasons. First, the transmission upgrades in Connecticut and Southeast Massachusetts have greatly reduced the need for supplemental commitment for local reliability, thereby substantially reducing the amount of capacity requiring NCPC payments. Second, several reliability agreements expired in 2009, and all remaining agreements expired on June 1, 2010 when the first Forward Capacity Commitment Period began. This reduced reliability agreement costs from $129 million in 2008 to $22 million in 2010.

Despite the overall reduction in uplift, NCPC payments to units committed for first contingency requirements and system-wide reserve requirements rose from $33 million in 2009 to $85 million in 2010. The majority of this uplift occurred after April 2010, consistent with the increase in supplemental commitment.

Conclusions

Our assessment of system operations indicates that the ISO has operated the system relatively well, and we found no major concerns. Additionally, the cost of satisfying the system's local reliability and system-wide reliability requirements continued to decrease in 2010.

However, we recommend one change, which is listed in the table of recommendations below. This change, 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.

G. 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 Forward Capacity Auction ("FCA1") was held in February 2008, facilitating the procurement of installed capacity from June 2010 to May 2011. Thus far, four auctions have been held, which had competitive results and satisfied New England's planning requirements through May 2014.

In June 2010, the first Capacity Commitment Period began, allowing for the expiration of the individual reliability agreements that had been used extensively to maintain the resource requirements in Connecticut, Boston, and Western Massachusetts. This has significantly improved the efficiency of the long-term incentives to capacity suppliers compared with relying on reliability agreements to retain existing capacity. Unlike markets, reliability agreements do not provide transparent prices indicating the marginal value of capacity in each area.

Each FCA has procured a significant amount of excess capacity. For example, FCA4 procured over 37 GW of resources, exceeding the Net Installed Capacity Requirement ("NICR") by almost 5.4 GW. The substantial excess is largely due to the effects of the price floor that prevents capacity prices from falling sufficiently to clear only the minimum requirement. When the floor is eliminated, the price is likely to fall close to zero due to the level of existing capacity and the vertical demand curve implicit in the FCM design. The latter may warrant reconsideration in the future.

In the first four FCAs, nearly 6.9 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 four auctions, large quantities of demand response resources have entered at prices well below the net entry costs for new generation. This outcome is efficient so 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.

Since 2010, the Commission has issued two orders on a set of market-design issues pertaining to the Forward Capacity Market.[13] The Orders' most significant determinations direct the ISO to model eight capacity zones corresponding to its eight Load Zones; to strengthen the supply-side market power mitigation rules; to extend the price floor in the auction through at least FCA6; and to develop buyer-side market power mitigation rules in order to address the shortcomings of the proposed Alternative Price Rule.[14] In the short term, the extension of the price floor will likely lead to a continuation of large quantities of excess capacity, as well as prices clearing at the floor. In the long term, the enhancement of buyer-side and seller-side market power mitigation measures and the enhancement of zonal pricing will likely lead to more efficient market outcomes. It will be important once the price floor is lifted, however, to evaluate whether the current design provides efficient economic signals to attract and retain supply and demand resources.

H. Competitive Assessment

The report evaluates the market concentration and competitive performance of the markets operated by ISO New England in 2010. 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 2010.

This competitive assessment has two main components. First, we utilize structural analyses to identify potential market power issues. Second, we evaluate the conduct of market participants in several areas. Although the structural analyses indicate that some suppliers may possess market power under certain conditions, our analyses do not indicate that suppliers withheld resources to raise prices in the New England markets.

---

13 See Order on Forward Capacity Market Revisions and Related Complaints, Docket ER10-787-000, *et al.* (Issued April 23, 2010) and Order on Paper Hearing and Order on Rehearing, Docket ER10-787-000, *et al.* (Issued April 13, 2011).

14 The Alternative Price Rule was a provision 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.

The structural component of our assessment evaluates each geographic market primarily 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"). This is most likely in constrained areas that can become separate geographic markets with a limited number of suppliers when congestion arises. Based on our pivotal supplier analysis, we found:

- The largest suppliers in Connecticut, Boston, and All of New England were pivotal in a large number of hours.
    - ✓ These increases were primarily driven by substantial declines in real-time surplus capacity that were discussed above.
    - ✓ However, transmission upgrades in Connecticut and Boston in recent years substantially increased transfer capability into these areas and reduced potential local market power concerns.
- Lower Southeast Massachusetts is no longer an import-constrained area following the transmission upgrades in July 2009, so no supplier was pivotal in that area.

The behavioral component 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 substantial decline in commitments for local reliability, this was not a significant concern in 2010. However, this was a significant issue in previous years. Revisions to the mitigation rules to better address this issue were made effective in late 2009.

I. Table of Recommendations

We make the following recommendations based on our assessment of the ISO New England's market performance in 2010. A number of these recommendations have been made previously and are now reflected in the ISO's Wholesale Market Plan.

Recommendation[15]	Wholesale Market Plan	High Benefit	Feasible/ Low Cost
Energy Markets			
1. Develop pricing changes to allow the costs of fast-start units and operator actions to maintain reliability (e.g., export curtailments) to be reflected in real-time prices.	4.12 & 4.13	✓	
2. Develop pricing changes to allow the costs of deployed demand response resources to be reflected in prices when they are needed to avoid a shortage.	4.12	✓	
3. Develop provisions to coordinate the physical interchange between New York and New England in real-time.	4.4	✓	
4. Modify allocation of "Economic" NCPC charges to make it more consistent with a "cost causation" principle.	4.8		✓
5. Modify inputs to the ex post pricing process to improve consistency with ex ante prices.			✓
6. Provide suppliers with the flexibility to modify their offers closer to real time to reflect changes in marginal costs.	4.15		
Ancillary Services Markets and System Operations			
7. Modify system-level Reserve Constraint Penalty Factor to be consistent with the costs incurred in the RAA process and in the Real-Time to protect operating reserves.	4.16.7		✓
8. Evaluate the RAA criteria and operating reserve requirements to identify improvements that could reduce supplemental commitments for reliability.	Relates to 4.16.16		
Capacity Market			
9. Modify demand response resources' capacity obligations to make them comparable to those of generation resources and imports.	4.6	✓	✓

15 *Feasible in Short Term*: indicated if the recommendation is likely to be feasible at a reasonable cost (limited in complexity and required software modifications).

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

I. Prices and Market Outcomes

In this section, we review wholesale market outcomes in New England during 2010. An analysis of overall price trends is provided in sub-section A, a discussion of transmission congestion patterns is in sub-section B, while the remaining sub-sections evaluate the convergence of prices between the day-ahead and real-time markets and virtual trading in the day-ahead market. The findings and conclusions from this section of the report are summarized below.

Average day-ahead and real-time energy prices increased 20 percent in 2010 due primarily to increases in fuel prices and load levels. Day-ahead energy prices rose from $43 per MWh in 2009 to $52 per MWh in 2010, while real-time prices were almost two percent higher in both years. Natural gas prices increased 10 percent from the prior year, and average load increased by 3 percent in 2010 due to the economic recovery and the hotter summer weather. A major generation outage lasting for nearly seven months (including the summer) and a reduction in the amount of capacity committed out-of-market for reliability reduced the supply available in real-time and contributed to the price increase.

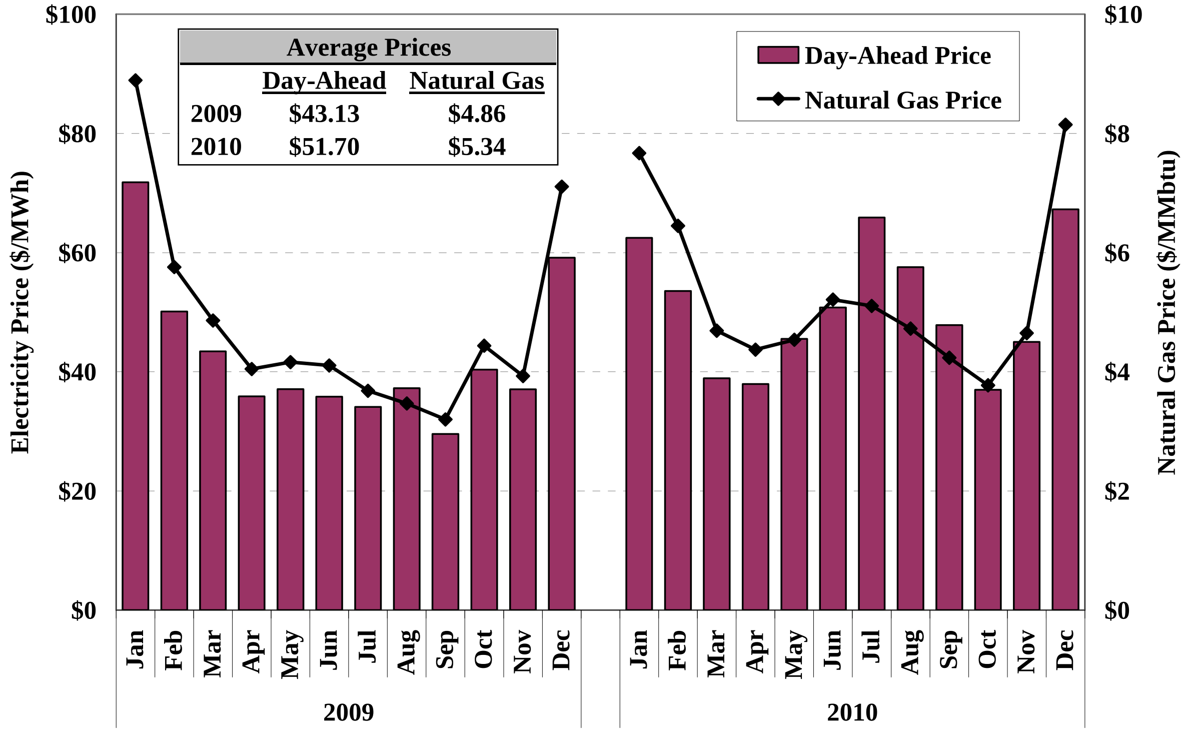
New England experienced very little congestion in 2010 into historically-constrained areas such as Boston, Connecticut, and Lower Southeast Massachusetts (“Lower SEMA”) as a result of transmission upgrades that have been made in recent years. Most of the price separation between net exporting regions and net importing regions was due to transmission losses, rather than transmission congestion.

Differences between day-ahead and real-time prices were moderate in 2010, indicating fair overall convergence. However, average real-time prices were almost two percent higher than average day-ahead prices, which is unusual since electricity markets typically exhibit slightly higher day-ahead prices. The higher real-time prices are primarily attributable to the following factors. First, the amount of surplus online capacity during real-time operations fell (following the transmission upgrades completed in mid-2009), leading to more frequent high real-time price events than expected by the day-ahead market. Second, the allocation of NCPC charges to virtual load has likely inhibited the market response to the sustained real-time price premiums.

A. Price Trends

Our first analysis examines day-ahead prices at the New England Hub in 2009 and 2010.[16] Figure 1 shows the load-weighted average price at the New England Hub in the day-ahead market for each month in 2009 and 2010. The figure also shows the average natural gas price,[17] which should be a key driver of electricity prices when the market is operating competitively

Figure 1: Monthly Average Day-Ahead Prices and Natural Gas Prices
New England Hub, 2009 – 2010



The figure shows that natural gas price fluctuations were a significant driver of variations in monthly average electricity prices in 2009 and 2010 as expected. In 2010, 41 percent of the

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

installed generating capacity in New England used natural gas as its primary fuel.[18] Low-cost coal and nuclear resources typically produce at full output, while natural gas-fired resources are on the margin and set the market clearing price in most hours. Therefore, electricity prices should be correlated with natural gas prices in a well-functioning competitive market. Natural gas prices are typically higher during the winter months when demand increases during colder weather. Accordingly, natural gas prices decreased from January to March and rose from October to December in both 2009 and 2010, leading to concomitant changes in electricity prices over the same period.

Electricity prices usually increase during high summer and winter load periods when the demand for cooling and heating are highest. The effects of seasonal changes in demand were more significant in 2010 than in 2009, particularly during the summer months. For example, average natural gas prices decreased 2 percent in July 2010 from the prior month, while the average electricity prices rose 30 percent in July 2010 from the prior month due to increased demand levels.

The average New England Hub price in the day-ahead market increased 20 percent from 2009 to 2010 primarily due to increases in fuel prices and load levels. The most significant driver of the higher prices was the increase in the average natural gas price of 10 percent from 2009 to 2010. In addition, #6 fuel oil prices increased 29 percent, #2 fuel oil prices increased 30 percent, and the coal prices increased 27 percent from 2009 to 2010. The average load increased 3 percent from 14.6 GW in 2009 to 15.0 GW in 2010. Peak load rose even more substantially -- up 8 percent to 27.0 GW in 2010. The reductions in hydroelectric generation from internal units and imports from Quebec also contributed to the increase in electricity prices. Hydroelectric production fell 11 percent and imports from Quebec fell 15 percent from 2009 to 2010.

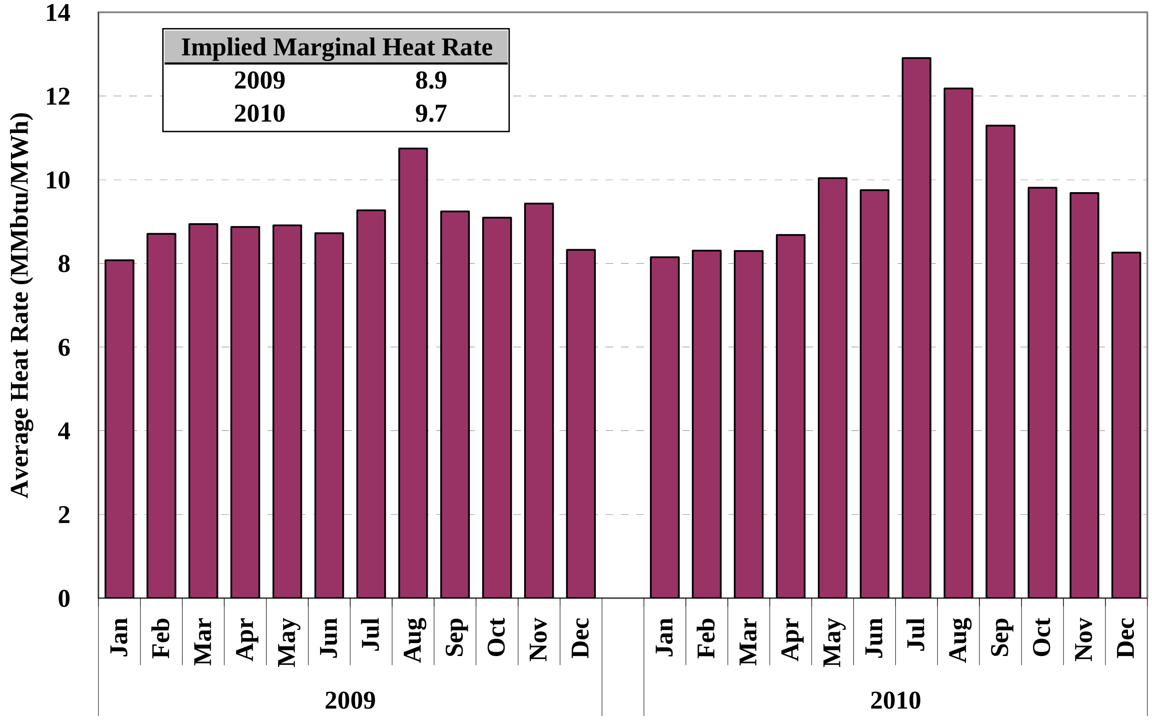
To better identify changes in energy prices that are not related to the fluctuations in natural gas prices, Figure 2 shows the marginal heat rate that would be implied if natural gas resources were

---

18 ISO New England, “2010-2019 Forecast Report of Capacity, Energy, Loads, and Transmission (CELT) Report,” April 2010.

always on the margin. The *implied marginal heat rate* is equal to the energy 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. A higher marginal heat rate indicates that factors other than higher natural gas prices have contributed to the increase in energy prices. Figure 2 shows the load-weighted average implied marginal heat rate for the New England Hub in each month during 2009 and 2010.

Figure 2: Monthly Average Implied Marginal Heat Rate Based on Day-Ahead Prices at New England Hub, 2009 – 2010



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, particularly in 2010. This was 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 and August 2010, which were also the months with the hottest temperatures and highest average loads.

The average implied marginal heat rate rose approximately 9 percent from 2009 to 2010. During the summer months (June to August), the implied marginal heat rate rose 22 percent from an average of 9.6 MMbtu per MWh in 2009 to 11.7 MMbtu per MWh in 2010. Outside the summer, the average implied marginal heat rate rose only slightly from an 8.7 MMbtu per MWh in 2009 to 8.9 MMbtu per MWh in 2010. The increase in the implied heat rate was primarily due to the following factors:

- Load rose significantly from 2009 to 2010, particularly during the summer. Average load increased 3 percent and peak load increased 8 percent. The number of hours when load exceeded 20 GW increased from 250 hours in 2009 to 531 hours in 2010 due to the hotter weather and increased economic activity in 2010.
- Hydroelectric generation from internal units fell 11 percent in 2010 compared to 2009 due to the extended outage of a large resource from May to December and led to more frequent dispatch of high-cost generation.
- Real-time price spikes became more frequent in 2010. The implied heat rate exceeded 20 MMbtu per MWh in 273 hours in 2010 compared to only 104 hours in 2009. This increase was partly due to a reduction in the average quantity of surplus capacity, which is the amount of dispatchable capacity (i.e., online or available offline quick-start) in excess of the energy and reserve needs of the system. The amount of surplus capacity declined partly because generator commitments for local reliability needs decreased substantially after July 2009. This is discussed in greater detail in Section VI.D.

B. Prices in Transmission Constrained Areas

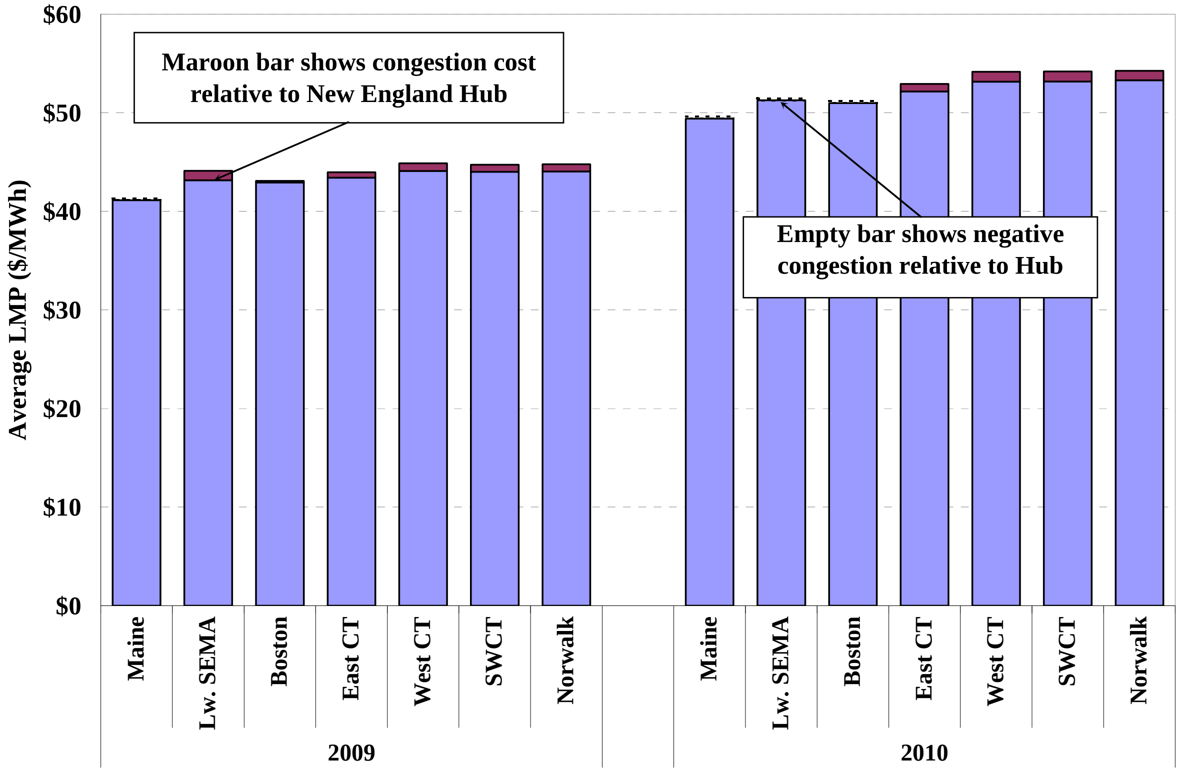
Historically, there have been significant transmission limitations between net-exporting and net-importing regions in New England. In particular, exports from Maine to the rest of New England have been limited by transmission constraints at times, while Connecticut and Boston were often unable to import 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 flows increase 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 in 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 III.

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 2009 and 2010 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., areas within Connecticut) exhibit positive congestion from the Hub.

Figure 3: Average Day-Ahead Prices by Location
2009 – 2010



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

The overall pattern of congestion did not change significantly from 2009 to 2010, with all areas exhibiting very low levels of congestion in both years. The most recent change of significance in congestion patterns occurred in Lower SEMA where 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, and to *negative* $0.20 per MWh in 2010. The reduction in congestion was due to transmission upgrades into Lower SEMA that were placed in-service in early July of 2009. This significantly increased the thermal transfer capability of the Lower SEMA transmission interface and virtually eliminated congestion into Lower SEMA in the last six months of 2009 and the entirety of 2010.

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. This 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 accept their offers only 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.

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.

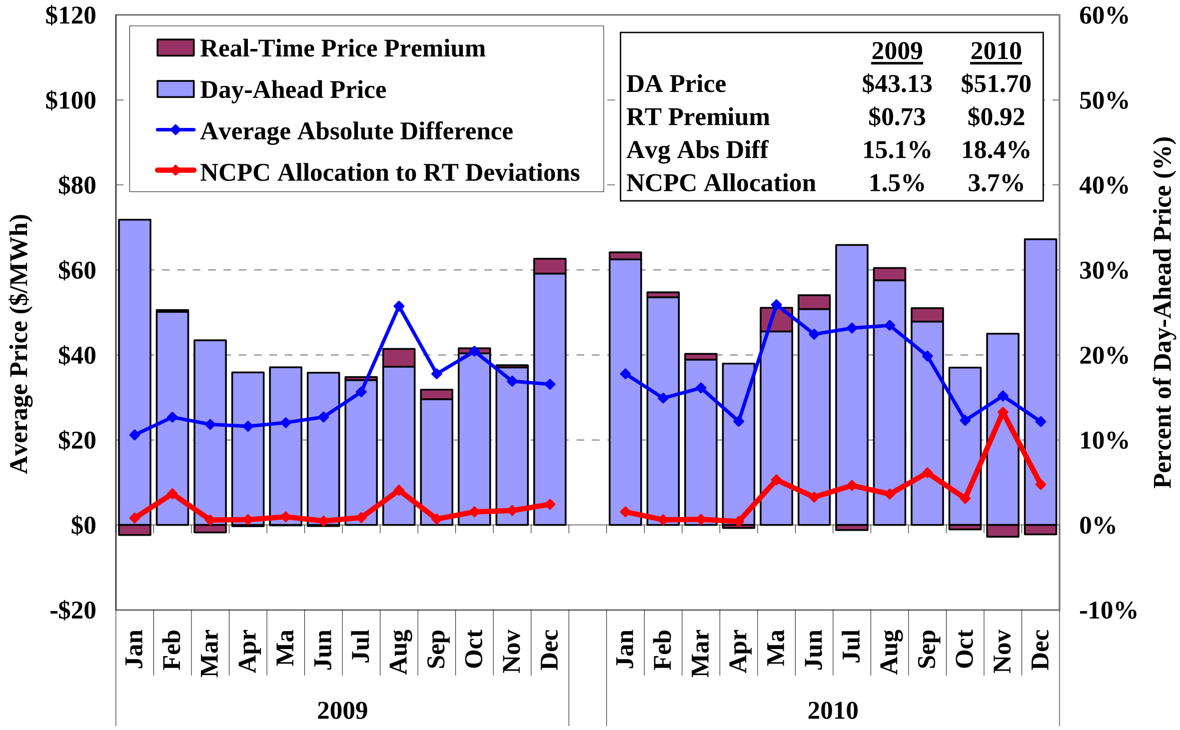
Good convergence between day-ahead and real-time prices is important. Since the day-ahead market facilitates most of the energy settlements and generator commitments in New England, good price convergence with the real-time market helps ensure efficient day-ahead generator commitments and external schedules that are consistent with 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 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 2009 and 2010.[19]

Figure 4: Convergence of Day-Ahead and Real-Time Prices at New England Hub 2009 – 2010



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-

19 Day-ahead and real-time prices are averaged on a load-weighted basis.

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 is reported as a percentage of the average day-ahead price in the month. The figure also shows the monthly average rate of Net Commitment Period Compensation ("NCPC") that is charged to real-time deviations, which is shown by the red line and is also reported as a percentage of the average day-ahead price in each month.

In electricity markets, average day-ahead prices tend to be slightly higher than average real-time prices. This is 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. Although day-ahead premiums occurred in some months in 2009 and 2010, Figure 4 shows that the majority of months exhibited a real-time premium. From July 2009 to December 2010, 13 out of 18 months exhibited real-time premiums. One reason for the more frequent real-time premiums may be the decline in surplus capacity (i.e., available online and offline quick-start capacity in excess of the energy and reserve needs of the system) that occurred in this timeframe because the ISO has been committing less generation after the day-ahead market for reliability purposes. The lower operating capacity margins generally lead to increased price volatility and higher average real-time prices. To the extent that participants in the day-ahead market did not immediately adjust their real-time price expectations to fully account for the lower operating capacity margins, their day-ahead schedules would lead to day-ahead market prices that were lower than real-time prices beginning in July 2009.[20]

Another reason for the more frequent real-time premiums is that the average allocation of NCPC charges increased significantly in May 2010, which is discussed in detail in the next subsection. These charges have likely inhibited the natural market response to the predictable real-time premiums. The increased allocation of NCPC charges (per MWh) to virtual load in particular has likely inhibited the natural market response to the sustained real-time price premiums.

---

20 The underlying causes of the reduction in surplus capacity are evaluated in detail in Section VI.D.

Hence, one of our recommendations is to discontinue the allocation of real-time NCPC charges to virtual load and other deviations that generally do not cause real-time NCPC. Virtual transaction patterns and the NCPC allocated to these transactions are evaluated in the next subsection.

The second measure of price convergence evaluated in the figure above is the average absolute difference between day-ahead and real-time prices. This measure is calculated by averaging the absolute value of the hourly differences between day-ahead and real-time prices. As a percentage of the average day-ahead price in each year, the average absolute difference increased from 15.1 percent in 2009 to 18.4 percent in 2010, which is consistent with the increase in price volatility that accompanied the reduction in surplus in 2010. The average absolute difference was particularly elevated during the peak summer months as one would expect due to the hotter conditions and higher associated load levels.

E. Uplift Allocation and Virtual Trading

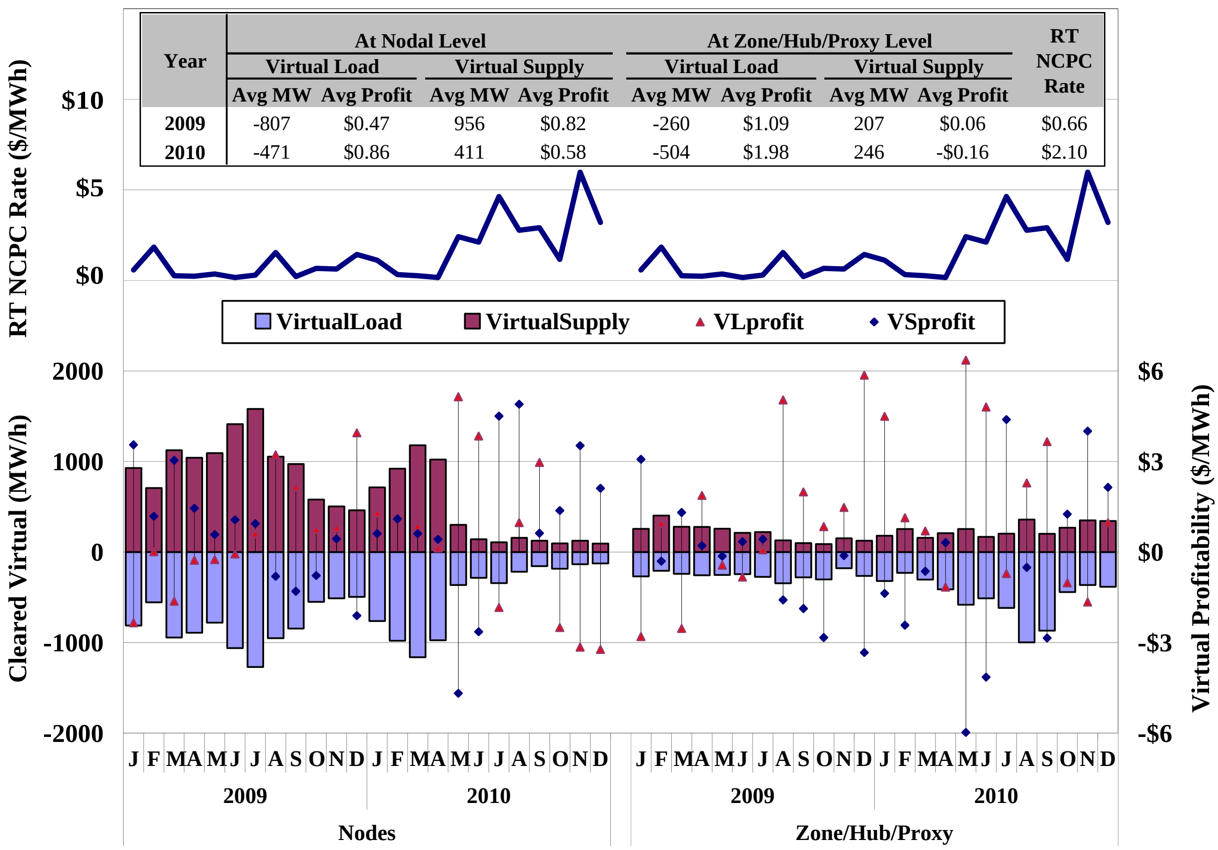
Virtual trading plays an important role in overall market efficiency by improving price convergence between day-ahead and real-time markets, thereby promoting efficient commitment and scheduling of resources in the day-ahead market. Virtual trading in the day-ahead market consists of purchases or sales of energy that are not associated with physical load or physical resources. Virtual bids and offers provide liquidity to the day-ahead market because they constitute a substantial share of the price-sensitive supply and demand that help establish efficient day-ahead prices.

Virtual transactions that are scheduled in the day-ahead market are settled against real-time energy prices. Virtual demand bids are profitable when the real-time energy price is higher than the day-ahead price; likewise, virtual supply offers are profitable when the day-ahead energy price is higher than the real-time price. Accordingly, if prices are lower in the day-ahead market than in the real-time market, a virtual trader may purchase energy in the day-ahead market and sell it back in the real-time market. This will tend to increase day-ahead prices and improve price convergence with the real-time market. Hence, profitable virtual transactions improve the performance of the day-ahead market.

ISO-NE allows virtual traders to schedule transactions at every pricing location, including individual nodes and more aggregated locations, such as the New England Hub and load zones. This provides flexibility for traders to arbitrage the price differences at various locations between day-ahead and real-time, leading day-ahead prices to converge with real-time prices.

Figure 5 shows the average volume of virtual supply and demand that cleared the market in each month of 2009 and 2010 by location, as well as the monthly average gross profitability of virtual purchases and sales.

Figure 5: Virtual Transaction Volumes and Profitability
2009 – 2010



Year	At Nodal Level				At Zone/Hub/Proxy Level				RT NCPC Rate
	Virtual Load		Virtual Supply		Virtual Load		Virtual Supply		
	Avg MW	Avg Profit	Avg MW	Avg Profit	Avg MW	Avg Profit	Avg MW	Avg Profit	
2009	-807	$0.47	956	$0.82	-260	$1.09	207	$0.06	$0.66
2010	-471	$0.86	411	$0.58	-504	$1.98	246	-$0.16	$2.10

Gross profitability is the difference between the price at which virtual traders bought and sold positions in the day-ahead market compared to the price at which these positions were covered in the real-time market. These quantities are shown separately for transactions at individual nodes and transactions at aggregated locations (i.e., the New England Hub, load zones, and external proxy buses). The gross profitability shown here does not account for NCPC cost allocations. The upper portion of the figure shows the average real-time NCPC rate for each month.[21]

The figure shows that scheduled virtual transactions decreased from 2009 to 2010. Overall, scheduled virtual load fell 9 percent and scheduled virtual supply fell 44 percent from 2009 to 2010. The decrease was due primarily to the decline in the volumes of virtual trading at the nodal level where scheduled virtual load and virtual supply fell 42 percent and 57 percent, respectively. The decrease, however, was offset by the increase in virtual trading activity at aggregated locations, particularly for virtual load.

The substantial drop in virtual transactions at the nodal level occurred in May 2010 when the ISO deployed a software solution to address an inconsistency in loss modeling at certain locations. This modeling inconsistency, which we first detected in late 2008, motivated a significant quantity of virtual trading at the affected locations because they produced low levels of consistent virtual profits (due to predictable differences between day-ahead and real-time LMPs). Hence, when this inconsistency was remedied, the associated virtual trading at those nodes ceased.

Despite this source of profits, Figure 5 shows that profits to virtual traders were relatively low in 2009 and 2010. Virtual traders netted approximately $12.7 million of gross profits in 2009 and $14.1 million of gross profits in 2010. Virtual supply was generally more profitable at the nodal level because larger price differences occur at individual nodes that are less liquid than the New England Hub or other aggregated locations. The average profit of cleared virtual supply at individual nodes was $0.58 per MWh in 2010, compared to a loss of $0.16 per MWh at more

21 The monthly real-time NCPC rate is defined as the total NCPC charges allocated system wide divided by the total real-time deviations for each month.

aggregated locations. Virtual load was considerably more profitable than virtual supply in 2010, due largely to the prevailing real-time price premiums. Virtual load accounted for 87 percent (or $12.3 million) of total virtual gross profit in 2010.

New England currently allocates nearly all real-time "Economic" NCPC charges to deviations between the day-ahead and real-time markets, including virtual supply and virtual load. Figure 5 shows that virtual transactions would net a loss on average after paying NCPC charges, particularly in 2010. Gross profitability of all cleared virtual transactions was $0.65 per MWh in 2009 and $0.98 per MWh in 2010, less than the average real-time NCPC rates of $0.66 per MWh in 2009 and $2.10 per MWh in 2010.

It is notable that the real-time NCPC rate increased substantially in May 2010. The real-time NCPC rate averaged nearly $3.60 per MWh in the months of May to December 2010, up significantly from $0.66 per MWh in 2009 and $0.46 per MWh in the first four months of 2010. The increase in the average NCPC charge rate was due to:

- The increase in overall NCPC charges, which was driven by the substantial increase in supplemental commitment for system reliability after a large flexible resource was unavailable beginning in May 2010; and

- The substantial decline in nodal virtual trading after the modeling inconsistency was remedied, which caused the NCPC costs to be allocated to a smaller quantity of real-time deviations (which includes virtual transactions).

The reduction in virtual trading activity raises potential concerns regarding the efficiency of the day-ahead market because active virtual trading in the day-ahead market promotes price convergence with the real-time market. Good price convergence, in turn, facilitates an efficient commitment of generating resources. Active virtual supply also protects the day-ahead market against market manipulation and market power abuses.

NCPC charges are caused by factors other than deviations, such as peaking resources not setting prices, congestion, system reliability needs, and outages. Hence, we find that the current allocation scheme over-allocates costs to deviations relative to the portion of the NCPC they likely cause. This is particularly true of virtual load transactions that tend to increase day-ahead commitments and, therefore, decrease the need for supplemental commitments. Given that real-time price premiums prevailed for much of 2010, allocating substantial NCPC costs to virtual load that does not cause these costs likely degraded the performance of the day-ahead market.

Hence, we recommend that the ISO modify allocation of Economic NCPC charges to be more consistent with a "cost causation" principle, which would generally involve not allocating NCPC costs to virtual load and other real-time deviations that cannot reasonably be argued to cause real-time economic NCPC. We are working with the ISO to develop a consensus approach to revising the NCPC allocation that would address this issue and improve the incentives for efficient day-ahead scheduling by market participants.

F. Conclusion

Energy prices increased by 20 percent in 2010, driven primarily by higher fuel prices and higher load that prevailed as the economy recovered and hotter conditions emerged in the summer months. Very little transmission congestion occurred as the transmission investments made in recent years continued to provide substantial capability into historically constrained areas.

Differences between day-ahead and real-time prices were relatively small in 2010, but the sustained real-time premiums raised a potential concern that the market was unable to quickly adjust to the higher real-time prices. These market outcomes are consistent with the inefficient allocation of real-time NCPC costs to virtual load and other real-time deviations. Therefore, we recommend that the ISO revisit the allocation methodology for Economic NCPC, revising it to be more consistent with cost causation principles.

II. Transmission Congestion and Financial Transmission Rights

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

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

Financial Transmission Rights ("FTRs") can be used to hedge the congestion costs of serving load in congested areas or as speculative investments for purchasers who forecast higher congestion revenues between two points than the cost of the associated FTR. An FTR entitles a participant to payments corresponding to the congestion-related difference in prices between two locations in a defined direction. For example, a participant that holds 150 MW of FTRs from point A to point B is entitled to 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

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

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

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, which increased approximately 38 percent from 2009 to 2010 because congestion in the day-ahead market increased in 2010. This increase was largely due to higher fuel prices that increase the cost of redispatching resources to manage constraints. Payments to FTR holders are funded by the congestion revenue collected by the ISO. In 2010, the congestion revenue collected by the ISO was sufficient to satisfy 100 percent of the obligations to FTR holders (referred to as the "target payment amount"), which was up from 95 percent in 2009.

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 2010, FTR prices in the monthly auctions were more consistent with congestion values in the day-ahead and real-time markets than FTR prices in the annual auction. The improvement in consistency of FTR prices and congestion values from the annual auction to the monthly auctions is expected because market participants gain more accurate information about market conditions as the lead time for the auction decreases.

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

24 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 day-ahead 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.
    - ✓ When a real-time constraint binds at a limit that is less than the scheduled flows in the day-ahead market, it results in *negative* congestion revenue.[25]
    - ✓ These costs are generally recovered as a form of uplift.

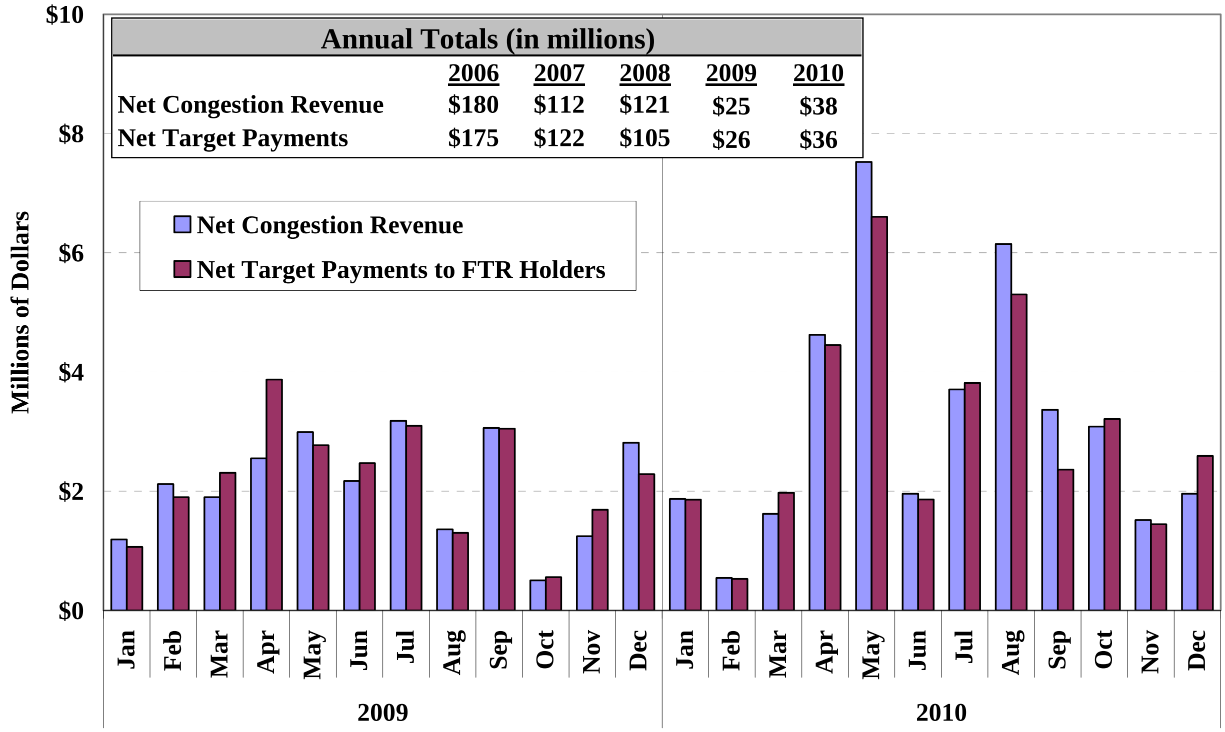
When the total congestion revenue collected by the ISO is not sufficient to satisfy the targeted payments to FTR holders, it implies that the quantities sold in the FTR auctions exceeded the capability of the transmission system. In months when this occurs, the unpaid FTR amounts are accrued until the end of the year when any excess congestion revenues remaining from months with a surplus are used to pay amounts accrued from months with a shortage plus interest. 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 are returned to transmission customers per the tariff.

Figure 6 compares the net congestion revenue collected by the ISO with the net target payments to FTR holders in each month of 2009 and 2010. The inset table compares the two quantities in

25 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.

the past five years. 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 6: Congestion Revenue and Target Payments to FTR Holders 2009 – 2010



The net congestion revenue rose from $25 million in 2009 to $38 million in 2010, a 52 percent increase. Likewise, the net target payments to FTR holders increased from $26 million in 2009 to $36 million in 2010. The increase in congestion in 2010 was primarily due to:

- Significantly higher load levels due to warmer weather and improved economic conditions, and

- Higher fuel prices, which raise redispatch costs and associated congestion-related price differences.

Nonetheless, as shown in the inset table, the net congestion revenues in 2009 and 2010 were substantially less than in prior years due to transmission upgrades in Boston, Connecticut and

Southeast Massachusetts that were completed from 2007 to 2009. The patterns of congestion are evaluated in greater detail in subsection B below.

The figure also shows that net congestion revenues were more than the net target payments to FTR holders in most months during 2009 (7 months) and 2010 (8 months). However, the total net congestion revenues for the 12 months in 2009 were only sufficient to fund 95 percent of the net target payments, while net congestion revenues in 2010 were sufficient to fund 100 percent of the net target payments

Congestion revenues exceed the net target payments to FTR holders when the amount of FTRs purchased along a congested transmission path is lower 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 increased to 1100 MW in the day-ahead market and the interface is congested, the ISO will collect 110 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 7 shows day-ahead and real-time congestion prices and FTR prices for each of the eight New England load zones and five sub-areas of interest in 2010. The congestion prices shown are calculated for peak hours relative to the New England Hub. Hence, if the congestion price in the figure indicates $4 per MWh, this is interpreted to mean the cost of congestion to transfer a megawatt-hour of power from the New England Hub to the location averaged $4 per MWh during 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 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 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. The table compares the average day-ahead and real-time congestion prices and FTR prices from the New England Hub to Connecticut in 2009 and 2010.

Figure 7: FTR Auction Prices vs. Day-Ahead and Real-Time Congestion Average Difference from New England Hub in Peak Hours, 2010

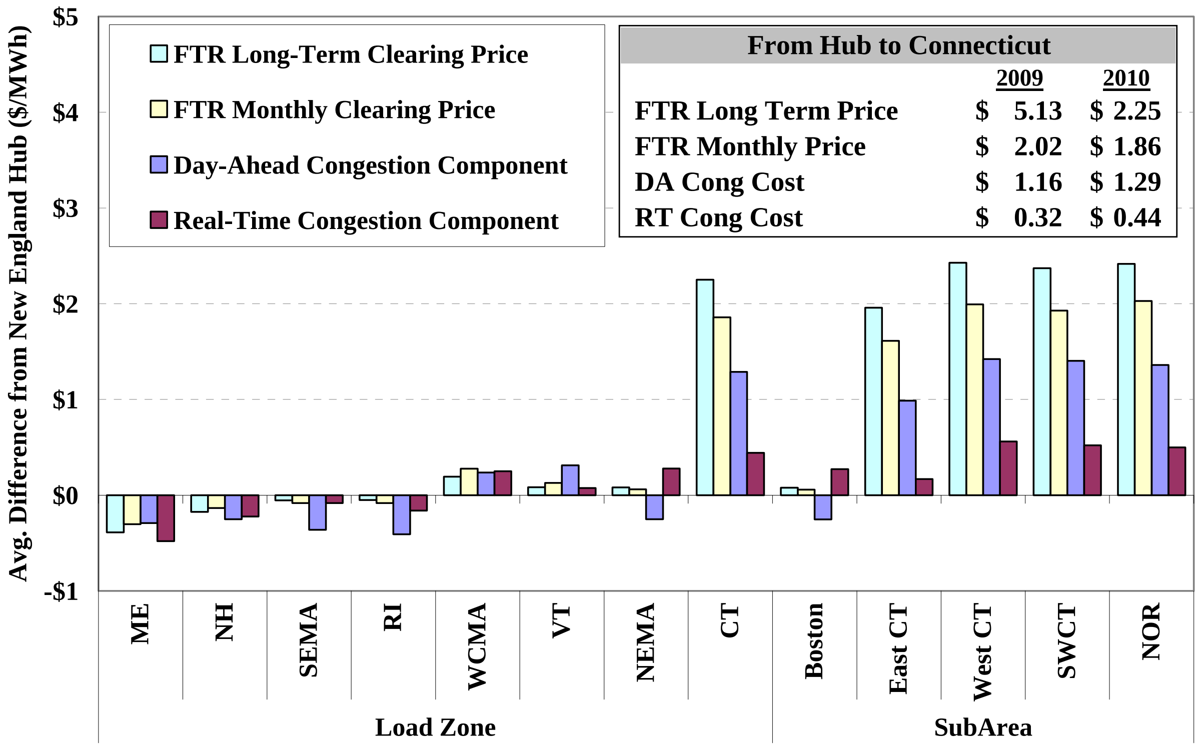


Figure 7 shows that in most areas during 2010, monthly FTR prices were more consistent with congestion prices in the day-ahead market than were annual FTR prices. For example, the annual FTR prices from the New England Hub to the various areas in Connecticut were 69 to 98 percent higher than the corresponding day-ahead congestion values, while the monthly FTR prices were only 40 to 63 percent higher. 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.

Monthly FTR auction prices exceeded the day-ahead congestion prices from the New England Hub to Connecticut. 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 2010. The over-estimates of congestion in the monthly FTR auctions were likely influenced by the congestion that occurred in prior periods. Congestion decreased substantially in the past two years from prior years.

However, the consistency of FTR prices and congestion improved substantially from 2009 to 2010. The table shows that the annual FTR price from the New England Hub to Connecticut was 342 percent higher than the day-ahead congestion value in 2009, but only 74 percent higher in 2010. Likewise, the average monthly FTR prices from the New England Hub to Connecticut were 74 percent higher than the day-ahead congestion value in 2009, but only 44 percent higher in 2010. 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 and 2010, they updated their expectations and the annual and monthly FTR prices converged more closely to the day-ahead congestion levels in later periods. We expect that this improvement will continue.

Given the volatile nature of congestion patterns and the variations in congestion patterns resulting from transmission upgrades in recent years, we found that FTRs were reasonably valued in the FTR auctions. The FTR market responded to changes in patterns of day-ahead congestion during 2010. Although this response has been slow, we conclude that the FTR markets performed reasonably well in 2010.

III. Reserve and Regulation Markets

This section evaluates the operation of the markets for operating reserves and regulation. 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 of 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.

The ISO runs a market for regulation service, which is the capability of specially equipped generators to increase or decrease their output every few seconds in response to signals from the ISO. The ISO uses regulation to balance actual generation with load in New England. The regulation market provides a market-based system for meeting the ISO's regulation needs.

This section evaluates market results in the following areas: the real-time reserve market, the forward reserve market, and the regulation market. The final part of this section provides a summary of our conclusions relating to the markets for reserves and regulation.

A. Real-Time Reserve Market Results

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 1,379 MW in 2010. 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 46 percent of the 10-minute reserve requirement in the form of spinning reserves during intervals with binding TMSR constraints in 2010.[26]

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 2010. Since higher quality reserves may always be used to satisfy requirements for lower quality products, the entire 30-minute reserve requirement can be satisfied with TMSR or TMNSR.

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,

26 The TMSR requirement is binding when a non-zero cost is incurred by the market to satisfy the requirement. This occurred in 3.9 percent of the intervals in 2010.

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 significant share of its local area reserve requirements with imported reserves. For example, imported reserves satisfied 31 percent of the Connecticut requirement during constrained intervals in 2010.

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
- $250 per MWh for the local 30-minute reserve constraints.[27]

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

27 The RCPF for local 30-minute reserve constraints was changed from $50 per MWh to $250 per MWh, effective January 1, 2010.

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.[28] Tight operating conditions can result in a shortage of 30-minute reserves, which leads to reserve clearing prices of $100 per MWh or more and a contribution to the energy prices of $100 per MWh. Alternatively, more severe conditions that result in shortages of both 30-minute and 10-minute reserves would produce 10-minute reserve clearing prices of $950 per MWh or more ($100 plus $850 per MWh) and energy prices exceeding $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.[29]

When available reserves are not sufficient to meet a requirement or when the marginal system cost of maintaining a particular reserve requirement exceeds the applicable RCPF, the real-time model will be short of reserves and set clearing prices based on the RCPF. For example, if the marginal system cost of meeting the system-level 30-minute reserve requirement were $150 per MWh, the real-time market would not schedule sufficient reserves to meet the requirement and the reserve clearing price would be set to $100 per MWh. This is efficient as long as 30-minute reserves truly have a reliability value of $100 per MWh. However, if the the operator intervenes at this point to maintain the required level of reserves at a cost greater than $100 per MWh, this out-of-market action will undermine the efficiency of the market because: a) it artificially lowers energy and reserve prices, and b) it is more costly than the value of the reserves being maintained. The RCPFs are analyzed in greater detail later in Section V.B.

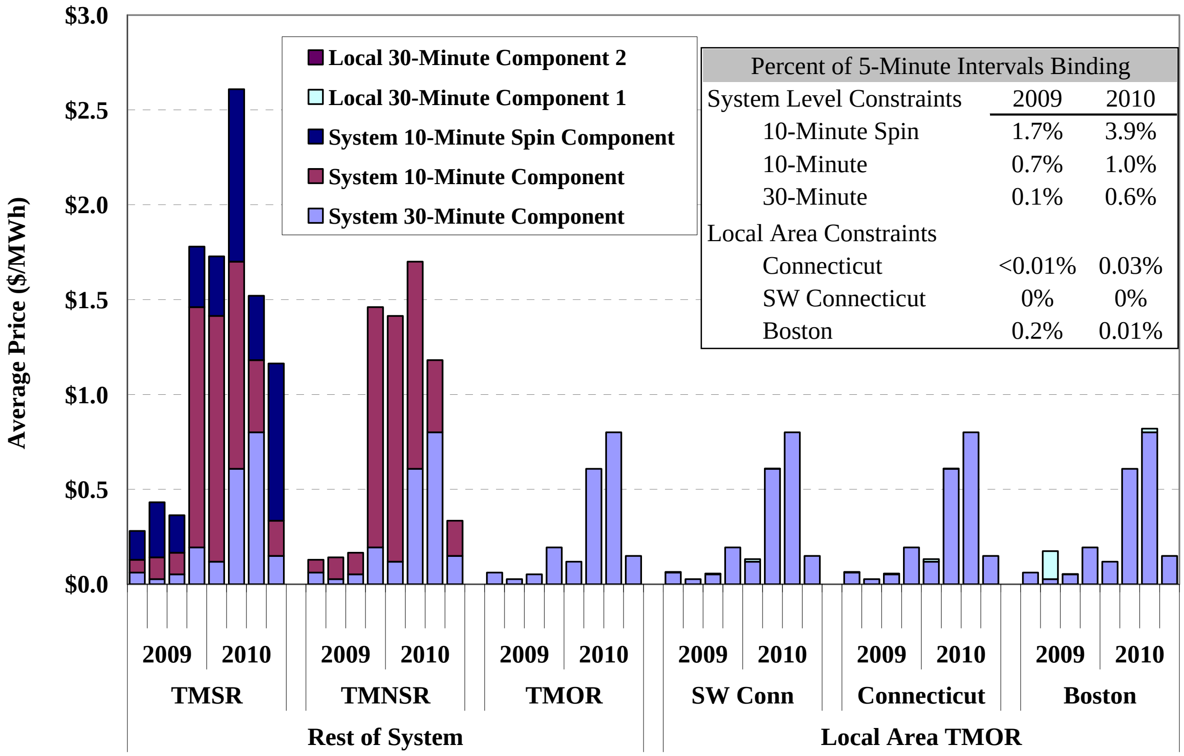
28 This assumes the operating reserve shortage results from a general deficiency of generating capacity.

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

3. Market Outcomes

Figure 8 summarizes average reserve clearing prices in each quarter of 2009 and 2010. 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 10-minute reserve price is based on the costs of meeting three requirements: the 10-minute spinning reserve requirement, the 10-minute non-spinning reserve requirement, and 30-minute reserve requirement (because system-level 10-minute reserves can satisfy all of these requirements).

Figure 8: Quarterly Average Reserve Clearing Prices by Product and Location 2009 – 2010



Outside the local constrained areas, the average TMSR clearing price rose from $0.36 per MWh in the first three quarters of 2009 to $1.76 per MWh in the five quarters including the fourth quarter of 2009 and all of 2010. The TMSR clearing price rose beginning 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 1,250 MW in the first nine months of 2009 to 720 MW from the fourth quarter of 2009 through 2010.[30] Surplus capacity fell because generator commitments for local reliability needs decreased substantially in 2009 as a result of changes that are discussed in greater detail in Section VI.D.

In the local areas, TMOR clearing prices were virtually identical to those in other areas because the local TMOR requirements were rarely binding in the real-time market.[31] Binding local reserve constraints have been very infrequent since transmission upgrades were made in Boston in the spring of 2007 and in Connecticut between 2007 and 2009.

Average reserve clearing prices are relatively low because reserve clearing prices are $0 in the vast majority of real-time intervals. This reflects that there is surplus capacity in most hours sufficient to meet system-level and local reserve requirements without redispatching generation. Figure 8 indicates that the system-level 10-minute reserve requirement was binding in just one percent of intervals in 2010, indicating that the requirement can be met at no cost with surplus capacity in 99 percent of intervals. However, when the system-level TMSR requirement is binding, the clearing price of TMNSR can rise quickly. In 2010, the average TMNSR clearing price was $78 per MWh in intervals when the system-level TMSR requirement was binding.

B. 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

---

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

31 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 $1.76 per MWh. This is composed of $1.76 per MWh for TMSR outside the local areas and the Boston 30-minute reserve component of $0.00 per MWh.

auction satisfy their obligations by providing reserves in real time from online resources or offline fast-start resources (i.e., peaking resources). This section evaluates the forward reserve auction results and examines how suppliers satisfied their obligations in real time.

1. Background on Forward Reserve Market

The ISO purchases several reserve products 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, SW 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 reserves in the real-time market. First, the forward reserve market has no TMSR requirement. Second, the forward reserve market has a minimum requirement for reserves in Rest of System, while the real-time market has no corresponding requirement. However, the minimum requirement for reserves in Rest of System was eliminated beginning in the Summer 2011 Procurement Period.

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

Forward reserve obligations may be satisfied in real time with reserves of equivalent or higher quality. When obligations are met with reserves of equivalent quality, the reserve provider receives the forward reserve payment instead of real-time market revenue based on the reserve clearing price. When obligations are met with reserves of higher quality, the reserve provider receives the forward reserve payment in addition to real-time market revenue based on the difference in clearing prices between the higher and lower quality products.[32]

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 2010/11 procurement period (October 2010 to May 2011) was held in August 2010. 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.[33]
- For Rest of System (i.e., areas outside Connecticut and Boston), the minimum purchase requirement is 600 MW. This is multiplied by a factor that accounts for the typical availability of resources deployed for TMOR to determine the TMOR requirement for each auction.[34] However, this requirement was eliminated beginning in the Summer 2011 Procurement Period.
- 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. The TMOR requirement is also adjusted for major changes in the topology of the system or the status of supply resources.

In the Forward Reserve Market auction, an offer of a high quality reserve product is capable of satisfying multiple requirements in the auction. In such cases, the higher quality product is priced according to the sum of the values of the underlying products, although this is limited by the $14 per kW-month price cap. For instance, 1 MW of TMNSR sold in Boston contributes to meeting three requirements: system-level TMNSR, system-level TMOR, and Boston TMOR.

32 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 real-time price difference between Boston TMSR and Boston TMOR.

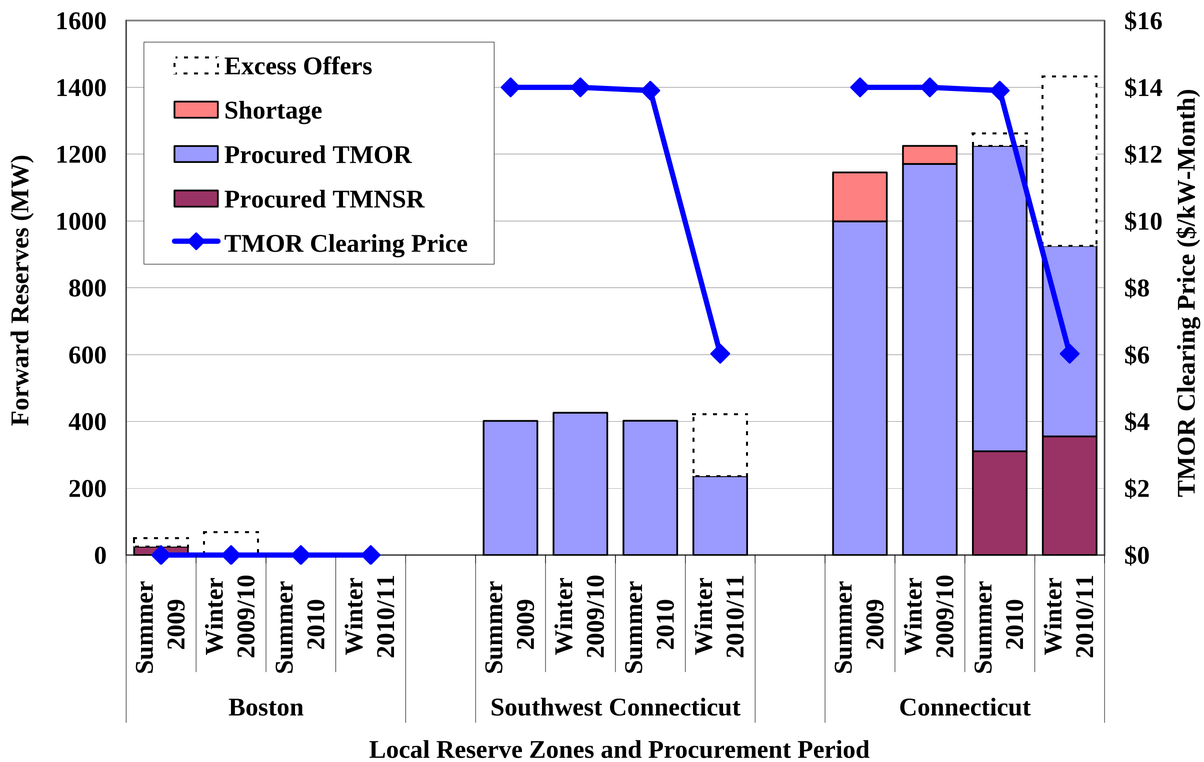
33 Usually, the forecasted largest contingency is the HQ Phase I/II Interconnection and the forecasted second largest contingency is the combination of the Mystic 8 and Mystic 9 generating units.

34 This factor is referred to as the "R-factor" which reflects the estimated availability of resources deployed to provide system-wide contingency reserves. The R-factor has been 1.33 for years and changed to 1.2 for the Winter 2010/11 auction, which corresponds to the availability discount factor for peaking resources that was filed by the ISO in February 2010.

Hence, the Boston TMNSR clearing price equals the system-level TMNSR clearing price (which incorporates the price of system-level TMOR) plus any premium for TMOR in Boston.

The following two figures summarize the quantities purchased in the last four forward reserve auctions towards each requirement. Figure 9 shows auction outcomes for the three local reserve zones, and Figure 10 shows auction outcomes for the system-level and Rest of System requirements. For each local reserve zone in each procurement period, Figure 9 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 9: Summary of Forward Reserve Auction for Local Areas Procurement for June 2009 to May 2011



The figure shows that very few resources were offered and procured in the Boston reserve zone in the last four auctions because there were no reserve requirements for internal Boston resources in these auctions. A substantial amount of transmission capability was added into the Boston area in 2007, leading the ISO to assume between 830 and 1,610 MW of External Reserve

Support in the recent 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 0 MW (i.e., no need for local resources due to enough External Reserve Support) in the last four auctions. Accordingly, the TMOR prices in Boston cleared at the same levels with the system TMOR prices ($0 per kW-month in all four auctions).

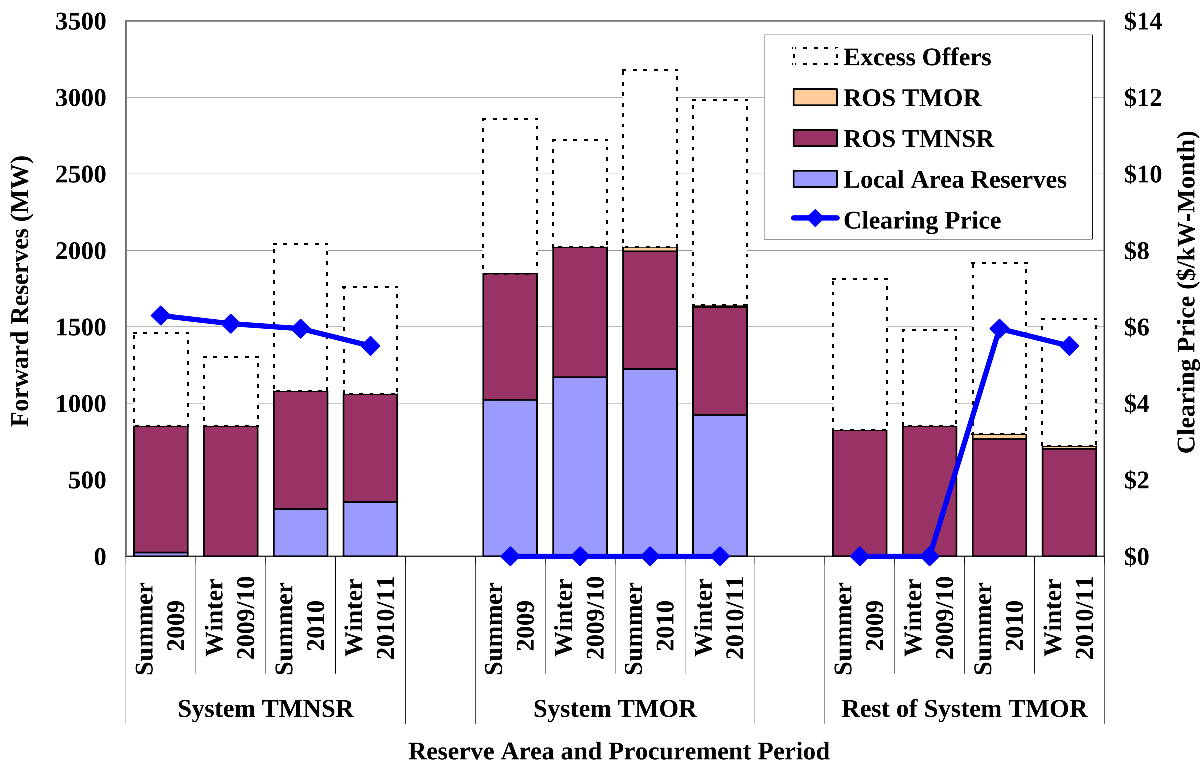
In Connecticut, the local reserve zone requirement was not met in the first two of the four auctions shown in Figure 9, but was satisfied in the last two auctions. Hence, the TMOR prices cleared at the $14 per kW-month price cap in the first two auctions and at lower prices ($13.9 and $6.0 per kW-month) in the last two auctions. The offer quantities have increased steadily over the past four procurement periods, from nearly 1,000 MW in the Summer 2009 period to 1,430 MW in the Winter 2010/11 period. This increase is due to the 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. The reserve requirement for internal resources in Southwest Connecticut was merely 22 MW in the Summer 2009 auction and was 0 MW in the subsequent three auctions. This is due primarily to transmission upgrades into Southwest Connecticut that were brought into service in early 2009. Starting in the Summer 2009 auction, nearly all of the requirement for Southwest Connecticut was satisfied by External Reserve Support. As a result, the local requirement was satisfied in all four auctions and TMOR prices cleared at the same levels with the TMOR prices in Connecticut.

Figure 9 shows that suppliers started to sell TMNSR in Connecticut in the Summer 2010 auction. The low level of TMNSR sales in the prior auctions was 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 had a strong incentive to sell TMOR rather than TMNSR in the Forward Reserve Auction when they received the price cap of $14 per kW-month in either case. In the last two auctions, however, it was likely that suppliers anticipated falling TMOR prices. This increased their incentive to sell TMNSR, which resulted in increased TMNSR sales in Connecticut.

Figure 10 shows the same analysis for the system-level and Rest of System requirements. For each procurement period, Figure 10 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 10: Summary of Forward Reserve Auction for Outside Local Areas Procurement for June 2009 to May 2011



Outside of the local reserve areas, the forward reserve requirements were satisfied in each auction. In all four 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 per kW-month in the Summer 2009 and the Winter 2009/10 auctions because the requirements were met by procurement for the TMNSR requirements. In the subsequent two auctions, 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.

Figure 10 also shows that a large share of the TMNSR requirement was procured outside of the local areas in the first two of the four auctions. 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 nearly 300 MW of TMNSR-capable fast-start capacity existed in the local areas. The low level of TMNSR sales in the local areas was 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 resulted in relatively higher clearing prices for TMNSR system-wide in those two auctions. In the subsequent two auctions, however, an average of roughly 330 MW of TMNSR was procured in the Connecticut reserve zone. This was likely a response to the falling TMOR prices in Connecticut in these auctions, which increased the incentive to sell the higher quality TMNSR product.

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.[35] Resources assigned to provide forward reserves forfeit any NCPC payments that they would otherwise receive. Forward reserve providers can arrange bilaterally for other suppliers to meet their obligations, although bilateral trading of obligations between non-affiliated firms was very limited in 2010. Suppliers that do not meet their forward reserve obligations incur a Failure to Reserve Penalty.[36]

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.[37] 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

---

35 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 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 monthly fuel index price is based on the lower of the natural gas or diesel fuel index prices in dollars per MMbtu.

36 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.

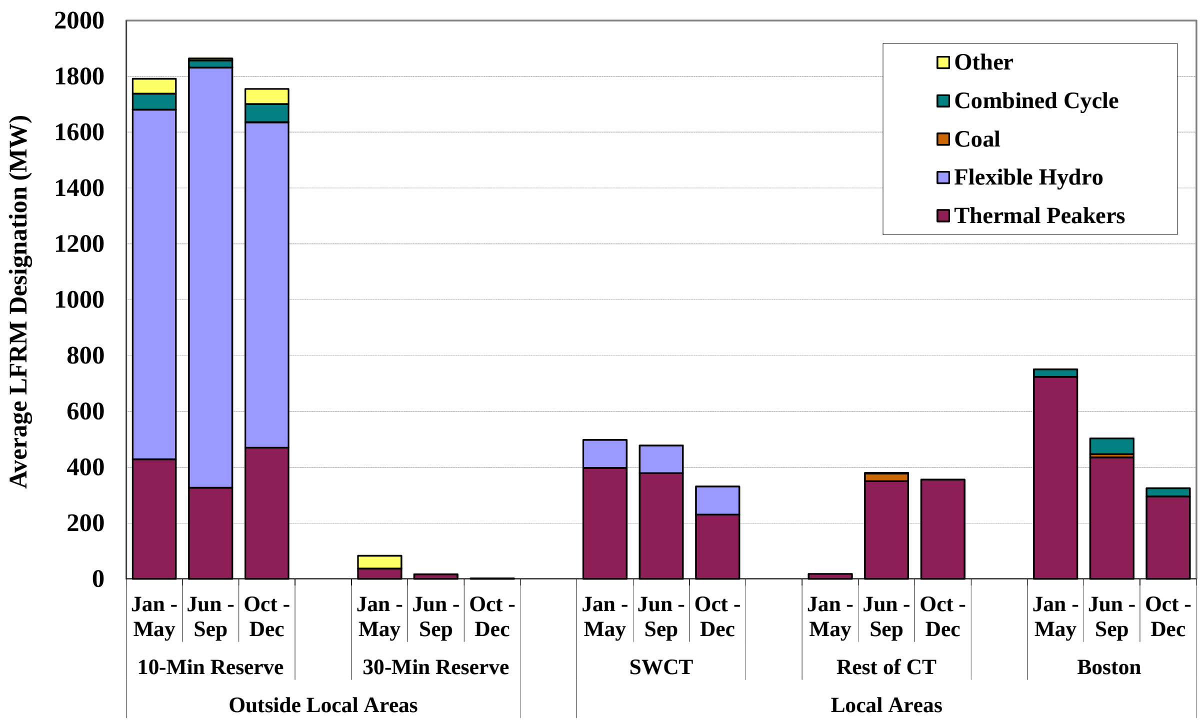
37 The Failure to Activate penalty is equal to the number of megawatts that does not respond times the sum of the Forward Reserve Payment Rate and the Failure to Activate Penalty Rate, which is 2.25 times the higher of the LMP at the generator's location or the Forward Reserve Payment Rate.

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 kinds of costs may be incurred by all units that provide forward reserves, but there are additional costs that are faced only 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 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 2010 by procurement period. The figure shows the average amount of reserves assigned in each region by type of resource.

Figure 11: Forward Reserve Assignments by Resource Type
2010



Approximately 95 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 2010. 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 2010 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 was roughly 12 MW in 2010. 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.

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 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.

C. 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 six years: from 143 MW in 2005 to 134 MW in 2006, 129 MW in 2007, 121 MW in 2008, 89 MW in 2009, and 68 MW in 2010.

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

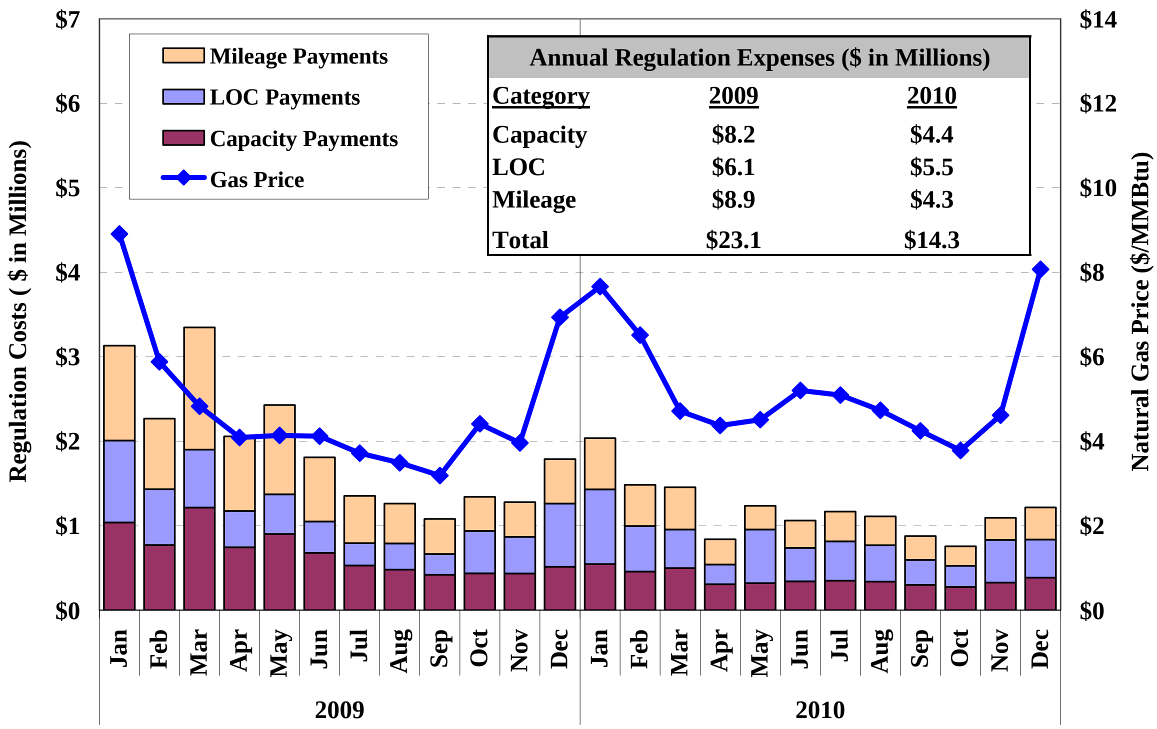
1. Regulation Market Expenses

Resources providing regulation service receive the following payments:[38]

- *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 12 by month for 2009 and 2010. The figure also shows the monthly average natural gas prices.

Figure 12: Regulation Market Expenses
2009 – 2010



38 In ISO-NE Manual M-11, Capacity Payment is the "Time-on-Regulation Credit," Mileage Payment is the "Regulation Service Credit," and the Lost Opportunity Cost Payment is 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 38 percent from $23.1 million in 2009 to $14.3 million in 2010, due largely to the reduction in regulation requirement in 2010, which was down roughly 25 percent from 2009.

The figure also 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 general decrease in regulation opportunity cost expenses in the summer months compared to the winter months.

Changes in natural gas prices and commitment patterns led to changes in offer patterns that explain some of the fluctuations in regulation market expenses in 2009 and 2010. 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.[39]

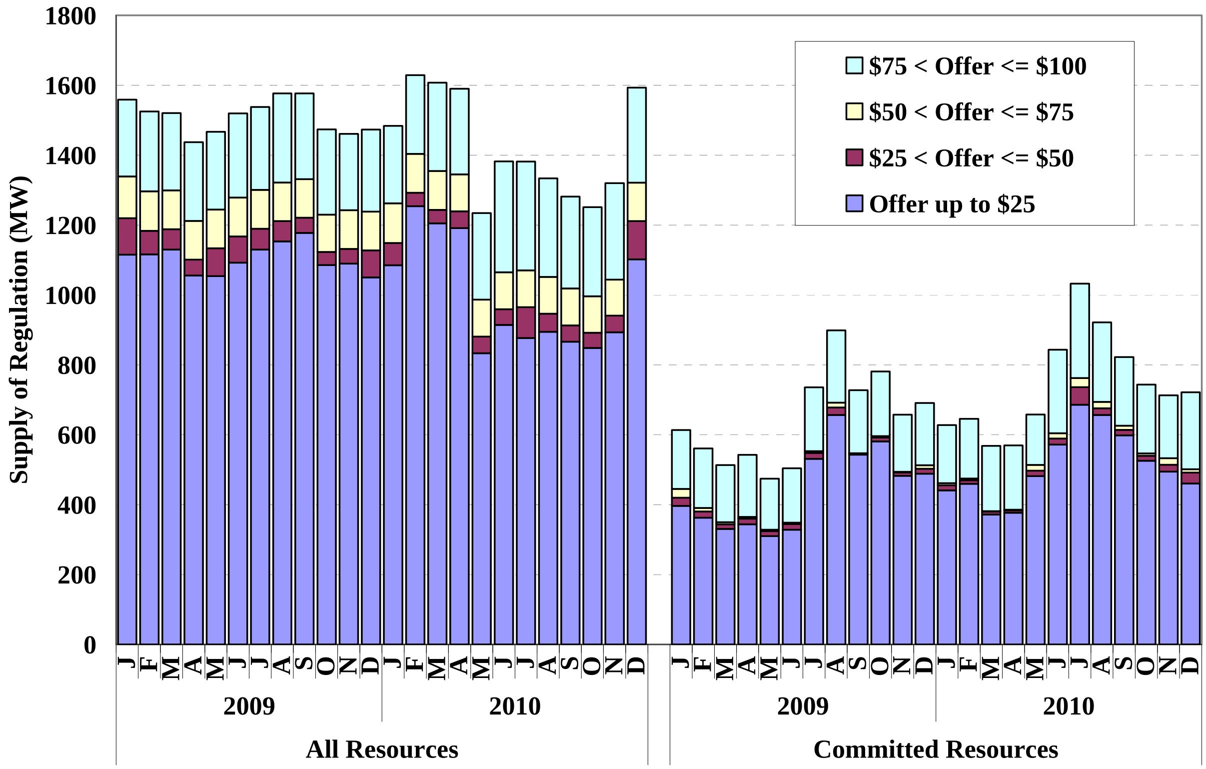
This section evaluates the offer patterns of regulation suppliers. Offline units cannot provide regulation service so selection of units is limited to units that are online at the time the service is needed. For this reason, we separately examine regulation offers from all resources and from online resources. Figure 13 shows monthly averages of the quantity of regulation offered into the market in 2009 and 2010 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 different colors on the bars in the chart show the average quantities offered by offer price range.

In Figure 13, the left panel shows that the regulation offer prices and quantities over the past two years were relatively consistent during most of the period. The quantities of total regulation offers varied typically between 1,400 MW and 1,600 MW in most months of 2009 and 2010. In 2010, however, the average quantity of total regulation offers was approximately 270 MW lower in the months of May to November than in other months. This was primarily because a large flexible generator was unavailable during the period.

39 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.

The portion of regulation offers in each price range was also relatively consistent over the past two years, except that the low-cost portion of regulation offers decreased from May to November 2010 due to the unavailability of a large flexible resource. On average, 70 percent of the total regulation offers were below $25 per MWh, 12 percent were between $25 and $75 per MWh, and 18 percent were more than $75 per MWh in 2010.

Figure 13: Monthly Average Supply of Regulation
2009 – 2010



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 52 percent of the regulation offered in the day-ahead market was available to the hourly real-time selection process in 2010. 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 from 2009 to 2010. This was largely driven by increased load levels and reduced hydroelectric production in 2010. These factors led to more commitment of combined cycle units, which provide most of the regulation capability in New England.

During 2009 and 2010, 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

Although the current market performed relatively well in 2010, the ISO may wish to consider co-optimizing the regulation market with the energy and operating reserve markets in the future. 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.

D. Conclusions

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. Outside the local constrained areas, the average TMSR clearing price rose from roughly $0.70 per MWh in 2009 to $1.75 per MWh in 2010. This increase was related to the reduction in the average surplus capacity, which is the amount of online and quick start generation in excess of the energy and reserve needs of the system. This reduction is evaluated and explained in Section VI.D of this report.

In the forward reserve market, clearing prices in Connecticut fell in 2010 due to transmission upgrades that reduced the quantity of reserves procured from internal generation, while clearing prices outside Connecticut trended down as a result of lower offer prices in the forward reserve auctions. We find that 95 percent of the resources assigned to satisfy forward reserve obligations in 2010 were fast-start resources capable of providing offline reserves.

The ISO may wish to consider the long-term viability of the forward reserve market for several reasons. First, it has not achieved one of its primary objectives (to lower NCPC by purchasing forward reserves from high-cost units frequently committed for reliability). Second, 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). Third, the Locational Forward Reserve Market is largely redundant with the locational requirement in the Forward Capacity Market.

Overall, the regulation market performed competitively in 2010. On average, approximately 740 MW of available supply competes to provide 70 MW of regulation service. The significant excess supply generally limits competitive concerns in the regulation market.

IV. 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 in either direction depending on market conditions, although New England exported more power to New York than it imported in 2010. 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.

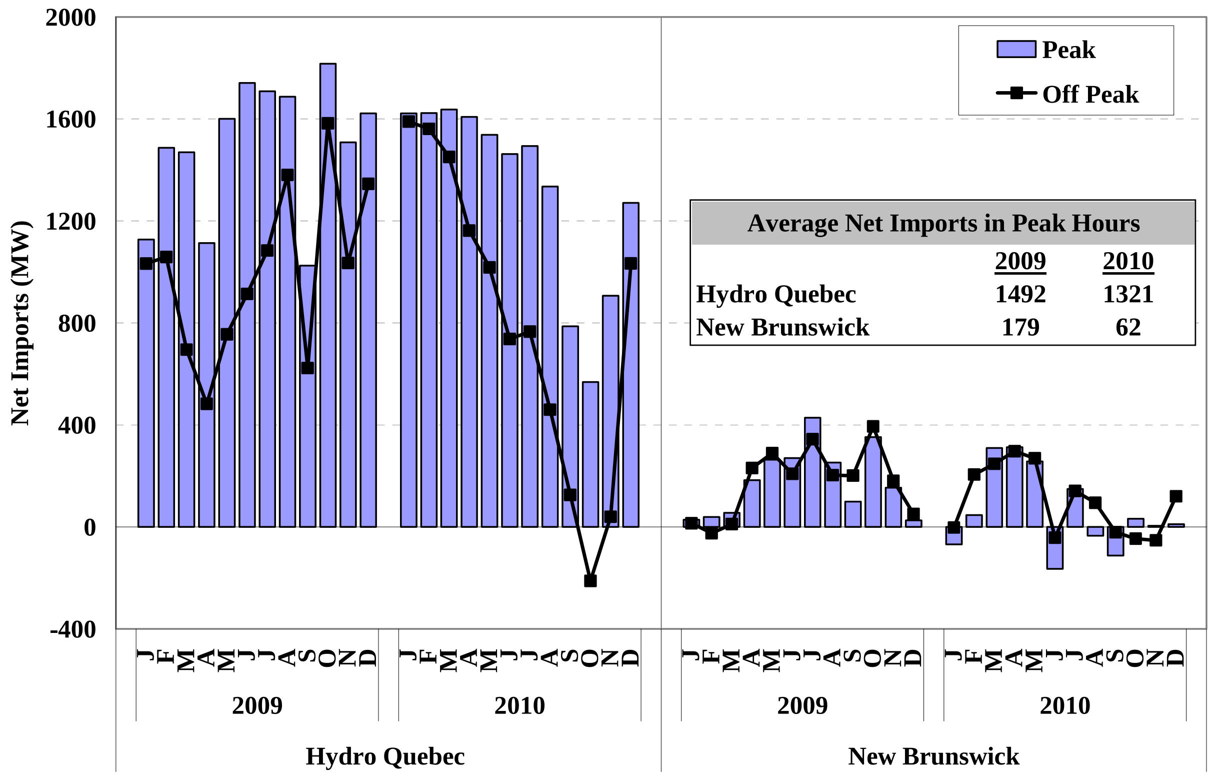
ISO-NE is interconnected with three neighboring control areas: the New York ISO, TransEnergie (Quebec), and the New Brunswick System Operator. New England and New York are interconnected by three interfaces: (i) the Roseton Interface, which is the primary interface and includes several AC tie lines connecting upstate New York to Connecticut, Massachusetts, and Vermont; (ii) the 1385 Line, a controllable AC interconnection between Norwalk and Long Island; and (iii) the Cross-Sound Cable, a DC interconnection between Connecticut and Long Island. New England and Quebec are interconnected by two interfaces: Phase I/II (a large DC interconnection), and the Highgate Interface (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 2009 and 2010. Section B evaluates the efficiency of scheduling by market participants between New York and New England. Section C presents estimates of the benefits that would result from two specific proposals under consideration for coordinating the interchange on the primary interface between New York and New England, and Section D provides a summary of our conclusions and recommendations.

A. Summary of Imports and Exports

The following two figures provide an overview of imports and exports by month for 2009 and 2010. Figure 14 shows the average net imports across the three interfaces with Quebec and New Brunswick by month, for peak and off-peak periods.[40] The net imports across the two interfaces linking Quebec to New England are combined.

Figure 14: Average Net Imports from Canadian Interfaces
2009 – 2010



40 Peak hours include hours ending 8 to 23, Monday through Friday (not including NERC holidays), and the remaining hours are included in Off-Peak.

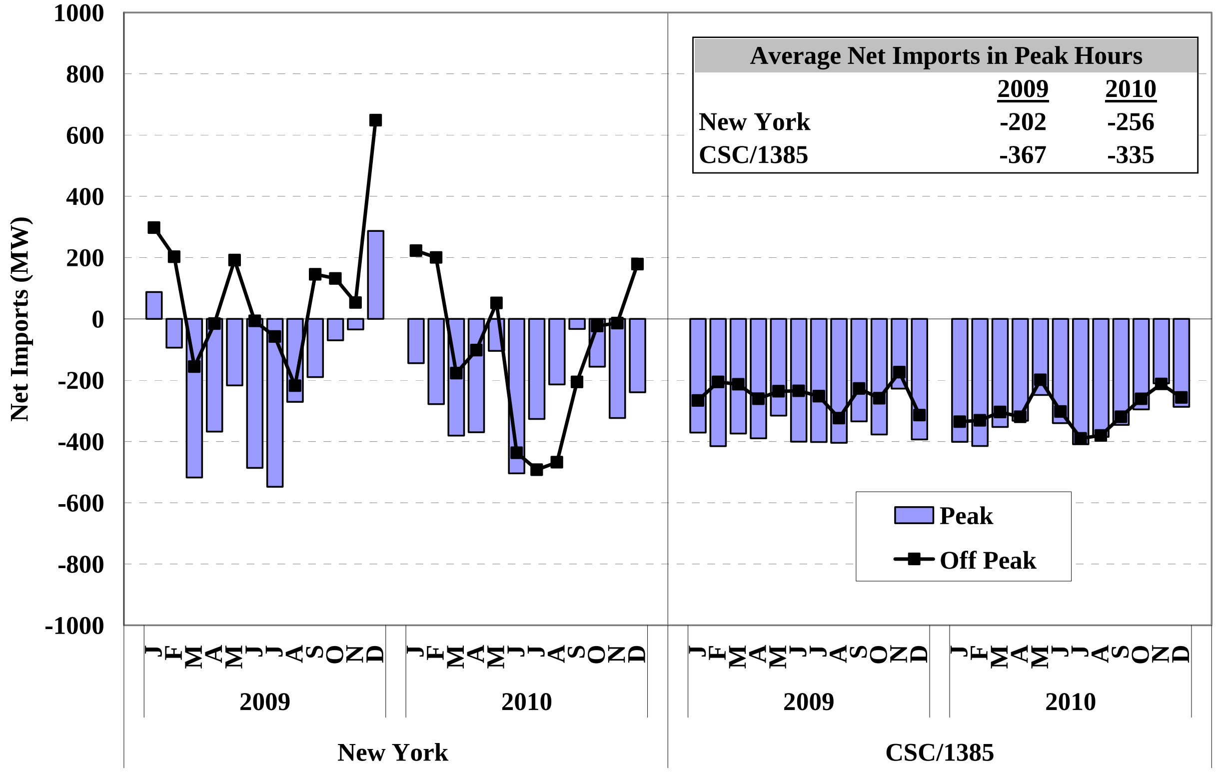
Figure 14 shows that power is generally imported from Quebec and New Brunswick. Across the two interfaces with Quebec, average net imports were higher during peak hours than during off-peak hours by roughly 500 MW in both 2009 and 2010. This reflects the tendency for hydro resources in Quebec to store water during low demand periods in order to make more power available during high demand periods. This pattern is beneficial to New England because it tends to smooth the residual demand on New England internal resources. 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 has historically sold more power to New England during periods when electricity prices were high. However, net imports from Hydro Quebec fell throughout 2010 in a pattern that did not result in the largest imports being made in periods with high prices. For example, sales were higher in the spring of 2010 (March to May) than in the summer of 2010 (June to August) even though prices were much higher in the summer. This pattern likely reflects that internal load levels in Quebec were higher and/or reservoir levels were lower than expected in the summer of 2010.

Figure 15 shows average net imports across the three interfaces with New York by month in 2009 and 2010 for peak and off-peak periods. The net imports across the Cross-Sound Cable and the 1385 Line are combined.

Figure 15 shows that the direction and the level of flows varied considerably across the primary interface with New York (the Roseton interface) during the two years. New England was a net exporter to New York across the primary interface during peak hours in 2009 and 2010. On average, power flowed into New York during peak periods in almost every month except in the winter months when New England is usually more affected than New York by cold temperatures and tight natural gas supplies.

Figure 15: Average Net Imports from New York Interfaces
2009 – 2010



The figure also shows that flows were relatively consistent from New England to Long Island across the Cross-Sound Cable and the 1385 Line, averaging approximately 307 MW in 2009 and 332 MW in 2010. The Cross-Sound Cable and the 1385 Line have a transfer capability of 330 MW and 100 MW, respectively. They were usually 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. In this sub-section, we evaluate the extent to which the interface is scheduled efficiently.

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 at the time 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. Nevertheless, we expect trading to improve the efficiency of power flows between regions.

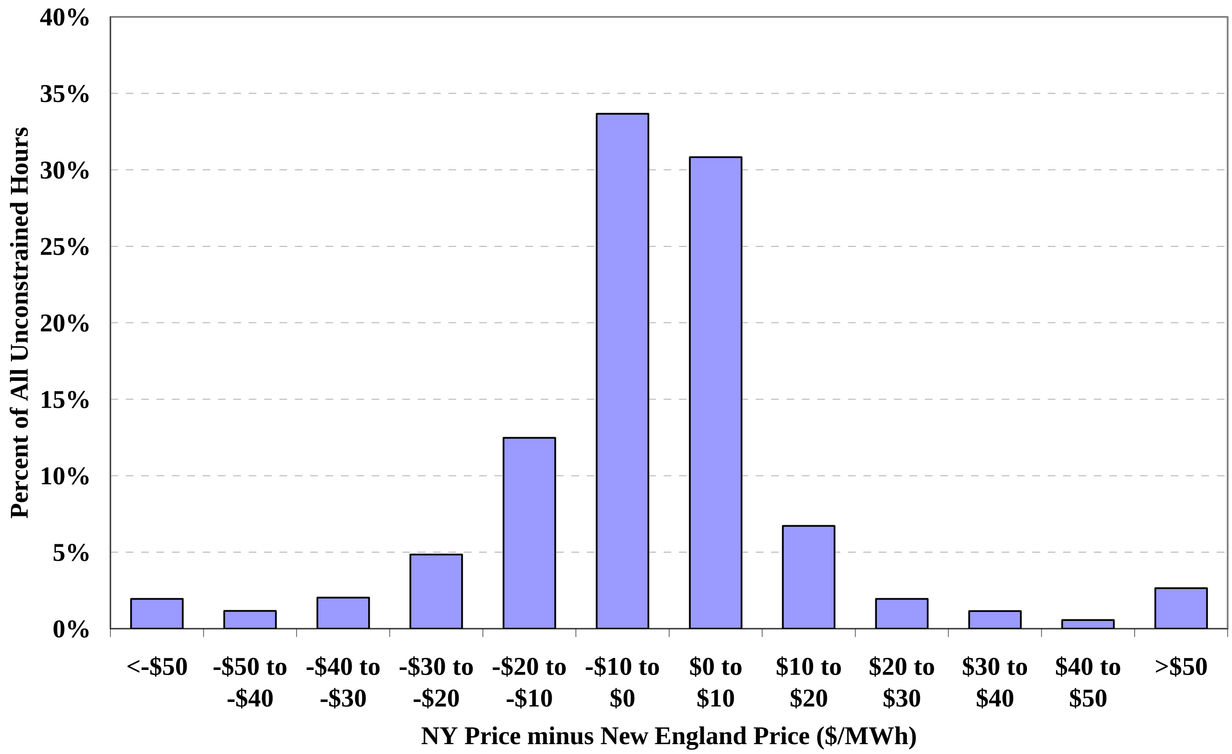
The following figures focus on the efficiency of scheduling across the primary interface between New England and New York. The Cross-Sound Cable is not evaluated in the following figures because it is scheduled under separate rules.[41] The 1385 Line is also not included because it was

---

41 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.

usually fully scheduled from Connecticut to Long Island in 2010 (i.e., the schedules are not responsive to prices). Figure 16 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.[42]

Figure 16: Real-Time Price Difference Between New England and Upstate New York Unconstrained Hours, 2010



While the factors described above prevent complete arbitrage of price differences between regions, trading should help keep prices in the neighboring regions from diverging excessively. Nonetheless, Figure 16 shows that approximately 35 percent of the unconstrained hours have real-time price differences of greater than $10 per MWh. In 6 percent of the hours, the price difference is greater than $40/MWh.

42 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).

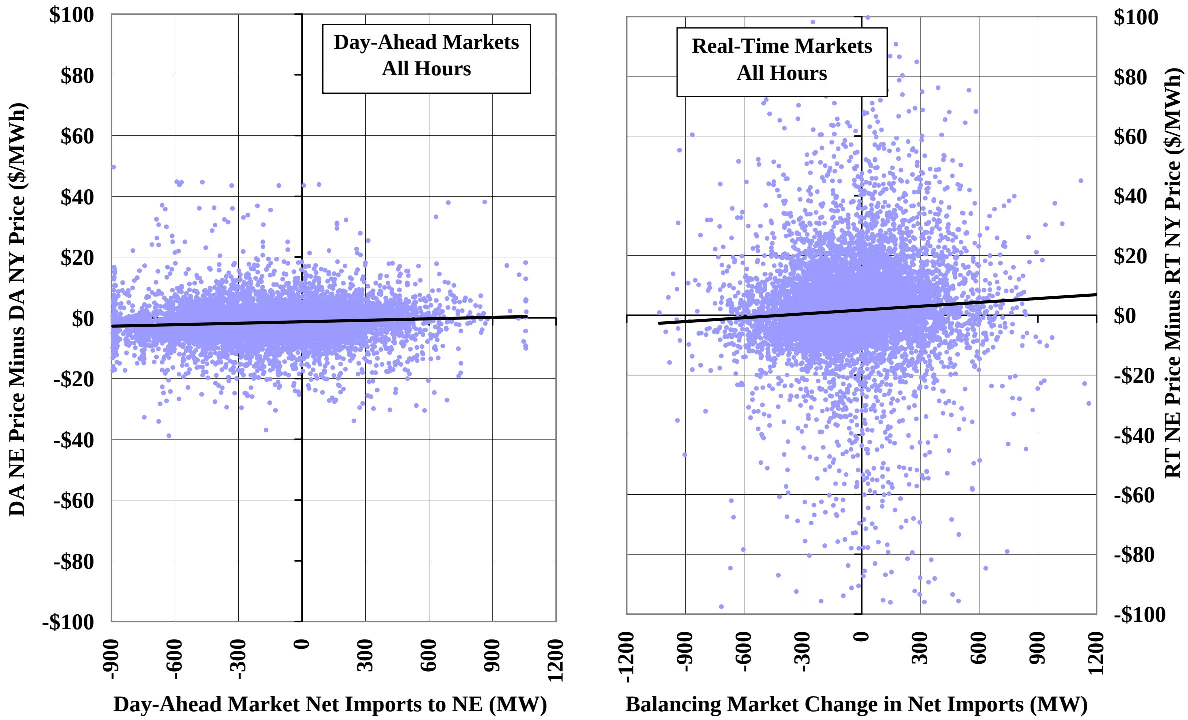
These results indicate that the current process does not fully utilize the interface. Given the pattern of price differences shown, there are many hours when increasing flows from the lower priced region to the higher priced region would have significantly improved the efficiency of clearing prices and 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.

Although market participants have not fully arbitraged the interface 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 17 shows the net scheduled flow across the interface versus the difference in prices between New England and upstate New York for each hour in 2010. 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 (500 – 300).

The trend lines in the left and right panels show statistically significant positive correlations between the price difference and the direction of scheduled flows in the day-ahead market and in the real-time market. However, the correlation in the day-ahead market is extremely weak, which indicates the difficulty participants have in scheduling transactions efficiently. The correlation is only slightly stronger in the real-time market. These positive relationships indicate that the scheduling of market participants generally respond to price differences by increasing net flows scheduled into the higher priced region. However, this response is highly variable.

Figure 17: Efficiency of Scheduling in the Day-Ahead and Real-Time Interface Between New England and New York, 2010



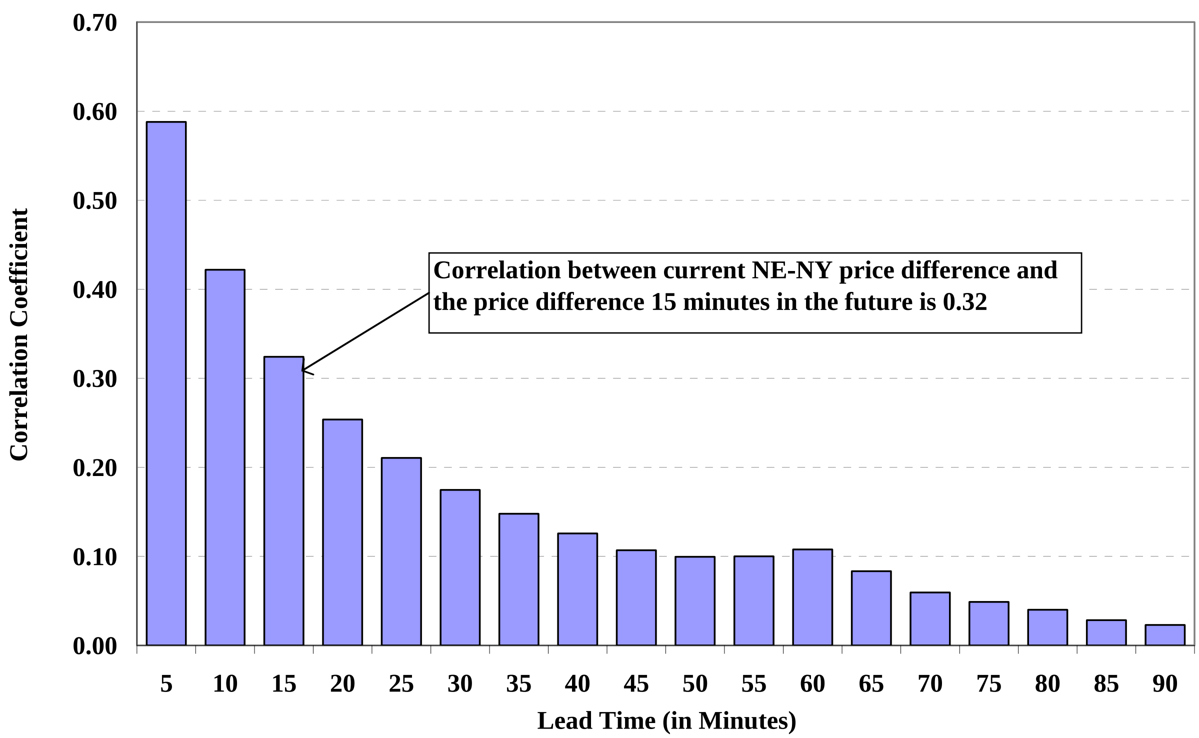
Although the arbitrage is not complete, the positive correlation between the price differences and the schedule changes indicate that participants generally respond rationally to the price differences in the real-time market. Additionally, total net revenues from cross-border scheduling in 2010 were $3.3 million in the day-ahead market and $2.1 million in the real-time market (not accounting for transaction costs).[43] 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 overall by facilitating the convergence of prices between regions.

43 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.

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 17. Forty-six percent of the points in the real-time market panel are in unprofitable quadrants – upper left and lower right – indicating hours when the net real-time adjustment by market participants shifted scheduled flows in the unprofitable direction (*from* the high-cost market *to* the low-cost market). Although market participant scheduling has helped converge prices between adjacent markets, Figure 17 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 18 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 0.32 in 2010.

Figure 18: Correlation Between Price Differences and Lead Time
Interface between Upstate NY and New England, 2010



Not surprisingly, Figure 18 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. Hence, the further in advance a participant schedules transactions, the less likely the transactions will be profitable (i.e., efficient). 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 beginning 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, the correlation remains relatively low at lead times of 15 minutes or more. The correlation was 0.17 at 30 minutes ahead of real time, which is the shortest scheduling lead time currently used by any RTO. Hence, the likely benefits of reducing scheduling lead-times are modest relative to the benefits from more direct coordination of the interchange. 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 the interchanges by the ISOs.

In July 2010, ISO New England and New York ISO commenced a joint effort known as the Inter-Regional Interchange Scheduling project to address the issue of inefficient scheduling between the two markets. The RTOs proposed two solution options: 1) Tie Optimization; and 2)

Coordinated Transaction Scheduling. The remainder of this section summarizes our assessment of the benefits of two initiatives to improve the efficiency of the interchange between the New York and New England.

1. Benefits Study Approach

The efficiency benefits of these proposals are best measured by their effect on production costs in the two regions, which reveals the true net economic savings of the proposals. Price differences between regions are reduced by scheduling power from the lower-priced market to the higher-priced market. As lower-cost resources in one market displace higher-cost resources in the adjacent market, total production costs are reduced and the savings are equal to the cost difference between the resources in the two markets. In this way, aggregate production costs fall as price convergence improves.

The estimated production cost savings naturally tend to be smaller than estimated consumer net savings resulting from energy price changes in each market. 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 modest production cost savings. Since the consumption of energy far exceeds the quantity of high-cost generation that is displaced by lower-cost generation, the estimated consumer net savings associated with the energy price changes tend to be much larger than the production cost savings.

Our previous assessments have consistently found that coordinating the interchange between ISOs would lead to significant reductions in both production costs and consumer costs.[44] The previous assessments estimated the benefits that would result from *optimal* scheduling of the interfaces between the markets. However, the share of the potential benefits that are ultimately realized depends on the effectiveness of the market solutions that are implemented by the ISOs. The assessment described in this sub-section builds on prior assessments by estimating the benefits of specific proposals for coordinating the interchange between the ISOs.

44 See "2009 Assessment."

The ISOs are currently evaluating specific proposals that will improve, but not perfectly optimize the interchange due to uncertainties present at the time the interchange is determined. The results discussed in this sub-section compare the benefits from optimal scheduling to the benefits that would result from the two proposals put forward by the RTOs as Inter-Regional Interchange Scheduling project:

- Tie Optimization – The ISOs exchange information 15 minutes in advance and optimize the interchange based on a prediction of market conditions. The interchange would be adjusted every 15 minutes.
- Coordinated Transaction Scheduling ("CTS") – Identical to Tie Optimization, except the interchange schedule is only adjusted to the extent that market participants have submitted intra-hour Interface Bids priced below the predicted price difference between the markets.

2. Modeling Assumptions

To quantify the share of potential benefits that would be captured by each proposal, we performed the simulations using three sets of assumptions:

- Ideal Interchange Case – Assumes the interchange is adjusted to the optimal level based on perfect information. The interchange is adjusted toward the higher-priced market until: (i) the interface is fully loaded, (ii) internal constraints prevent additional re-dispatch, or (iii) the adjustment reaches 500 MW.
- Tie Optimization Case – Assumes the interchange is adjusted to the *forecasted* optimal level. The ISOs' forecast may differ from actual conditions, so the resulting interchange may not be optimal. The following forecasts are used for this study:
  - ✓ On the NYISO side of the border, the forecast is based on the latest available advisory prices that are produced by its dispatch model ("RTD").
  - ✓ On the ISO-NE side of the border, the forecast is based on its hour-ahead forecast model. The forecast errors on the ISO-NE side are larger, which is understandable given that the ISO-NE forecast is performed further in advance than the NYISO forecast. Hence, we assumed the ISO-NE errors would fall by 50 percent to account for the effects of shortening the timeframe.
- Coordinated Transaction Scheduling Cases – This is the same as the Tie Optimization Case, except an assumed interface "bid stack" limits re-dispatch when the marginal bid exceeds the forecasted price difference. We assumed a interface bid stack beginning at zero and rising linearly up to $10 at 500 MW in the first case ("Int Bid1"), and rising linearly to $40 at 500 MW in the second case ("Int Bid2").

Comparing the results of these simulations allows us to evaluate the efficiency of specific proposals compared to ideal interchange scheduling. The simulations discussed in this section differ from the simulations used in previous assessments in the following respects:

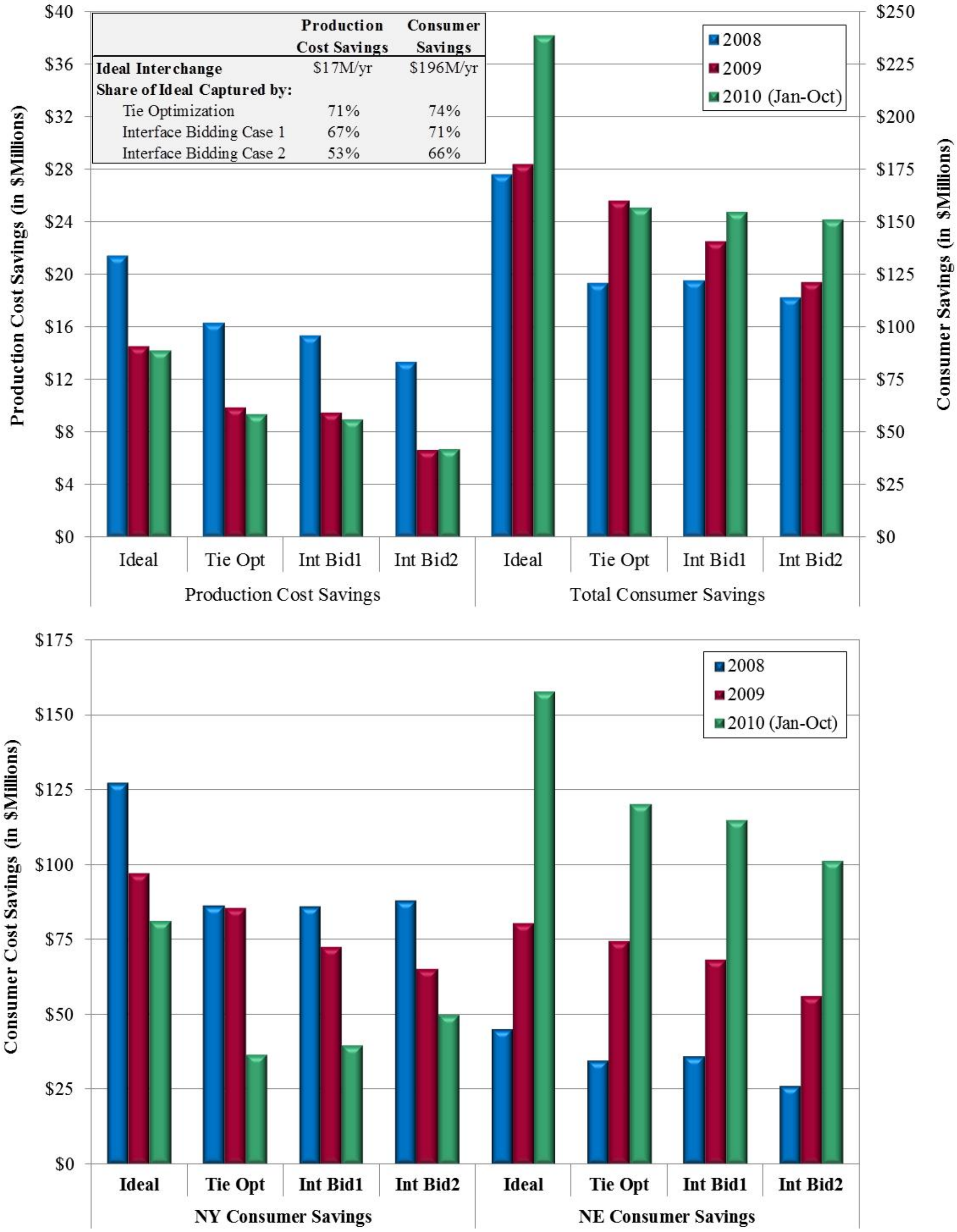
- These simulations use 15-minute interval level data, while previous assessments estimated the optimal interchange using historic hourly-integrated real-time data. Hourly data resulted in conservative estimates because it assumed one interchange value for the hour. It is usually efficient to adjust the interchange throughout the hour.
- These simulations impose a 500 MW limit on the size of the adjustment in the interchange in any interval relative to the actual interchange, while the previous simulations had no limit. We use the 500 MW limit to be conservative because the simulation model does not "see" internal transmission constraints that would bind due to the interchange adjustment.
- We exclude intervals when the New York border price is negative, since these are likely to become far less prevalent after the NYISO implements several market design changes in 2011.
- These simulations exclude intervals at the top of each hour. These intervals are frequently affected by ramp constraints and other conditions that lead to transient price spikes which our simulation model is not designed to model accurately. Hence, we conservatively estimate $0 production cost savings from these intervals, although it is likely that the interchange would be improved in these intervals.

Both the simulations presented in previous reports and this report use simplified network models that assume interchange adjustments are not possible if they would exacerbate congestion on active transmission constraints. This is conservative because re-dispatch would be possible in some such cases, and so there may be significant additional savings that we do not capture.

3. Simulation Results

The following figures and table summarize the estimated effects of the two proposals to optimize the interchange between New York and New England. Figure 19 summarizes the production cost savings and consumer savings to both markets that would result under each proposal, as well as under optimal interchange. The figure also summarizes the distribution of consumer savings between the two ISOs in each of the cases that we analyzed. Lastly, Table 1 provides statistics summarizing the estimated changes in market outcomes in each of the cases that we analyzed.

Figure 19: Estimated Benefits from Coordinating the Interchange with New York
January 2008 – October 2010



The average production cost savings in the Ideal Interchange Case is roughly $17 million per year, although this is likely conservative as a long-run expectation because: (i) one quarter of the hour is not included; (ii) the supply and demand conditions in both areas were not as tight as they are likely to be in the long run, causing shortages to be relatively infrequent; and (iii) natural gas prices were relatively low for much of this period.

The results indicate that a large share of the potential benefits would be captured by the two proposals for implementation. The Tie Optimization Case captures 71 percent of the efficiency benefits. The CTS Case with lower-priced bids performed nearly as well as the Tie Optimization Case, capturing 67 percent of the potential efficiency benefits. The second CTS case with higher-priced bids captured 53 percent of the efficiency benefits from the Ideal Interchange Case. It is difficult to predict how market participants would bid under CTS, but the bid assumptions used in the simulations are most likely far above the bids that market participants would actually submit, so it is likely that CTS would perform nearly as well as Tie Optimization.

Figure 19 shows that consumer savings in the ideal case average almost $200 million per year, which is conservative for the same reasons as listed above. Nearly three-quarters of the savings are captured by Tie Optimization, which falls only slightly to 71 percent with low-priced interface bids. The figure shows that the consumer savings accrue to both areas, although the relative savings has shifted year-to-year as congestion patterns and supply conditions have changed.

The following table provides some additional detail regarding the results of the simulations. It summarizes how frequently flow would be adjusted towards New York and towards New England. It summarizes the average size and the average price impact on each side of the border from these adjustments. This information is provided for the Ideal Interchange, Tie Optimization, and Coordinated Transaction Scheduling Cases.

Table 1: Simulated Effects from Coordinating the Interchange with New York
January 2008 – October 2010

	Ideal Interchange 2008	Ideal Interchange 2009	Ideal Interchange 2010	Ideal Interchange 2008-10	Tie Opt 2008-10	Int Bid 1 2008-10	Int Bid 2 2008-10
Flow Adjusted Into NY (% of intervals)	42%	46%	44%	44%	43%	43%	43%
Flow Adjusted Into NE (% of intervals)	41%	44%	45%	43%	42%	42%	42%
When Flow Adjusted Into NY:							
Avg. Adjustment (MW)	266	259	265	264	262	186	101
Avg. System LBMP Change in NY ($/MWh)	-$10.63	-$7.19	-$7.07	-$8.30	-$8.24	-$7.11	-$5.64
Avg. System LMP Change in NE ($/MWh)	$7.00	$2.96	$3.39	$4.45	$4.84	$3.95	$2.75
When Flow Adjusted Into NE:							
Avg. Adjusted Interchange (MW)	-226	-220	-237	-228	-210	-153	-89
Avg. System LBMP Change in NY ($/MWh)	$7.96	$4.36	$4.83	$5.72	$6.73	$5.62	$4.14
Avg. System LMP Change in NE ($/MWh)	-$8.21	-$4.93	-$7.43	-$6.86	-$6.88	-$5.87	-$4.39

The table shows that in each year, the adjustments toward New York and toward New England occur in nearly equal frequency. Because the average real-time price effect is larger when flow is adjusted into a market than out of a market, the simulations presented in previous reports and in this report have consistently found net consumer benefits. In other words, 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 levels than at lower prices levels. Because power flows are adjusted into each area with equal frequency, the average prices would fall in both markets and cause consumers in both areas to benefit from the improved coordination.

4. Conclusions of Benefits Study

The results of the simulations show sizable efficiency and consumer savings in all cases analyzed, which supports the ISOs' initiative to pursue the Inter-Regional Interchange Scheduling project. For the reasons we have discussed, these savings are likely to be conservative and would be larger under tighter supply/demand conditions over the long-run. The estimated savings are large, so this initiative should be a high priority. While the Tie Optimization proposal performed better in our simulations than the Coordinated Transaction Scheduling proposal, the benefits are very similar if participants submit relatively low-cost interface bids. Therefore, we would support either alternative.

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 and the New York ISO are planning two initiatives that are intended to improve the efficiency of scheduling between the two control areas.[45] First, the Inter-Regional Interface Scheduling initiative would create a system for coordinating the interchange between control areas based on expected near-term price differences. Second, market-to-market congestion management coordination, which will be presented to stakeholders in 2013, would 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. The estimated benefits of the second initiative are substantially lower than the benefits of the coordinated interchange initiative. We continue to recommend that ISO-NE and the New York ISO place a high priority on developing and implementing one of the coordinated interchange alternatives.

45 See the 2011 ISO-NE Wholesale Markets Plan.

V. 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 2010. This section examines the following areas:

- Prices during the deployment of fast-start generators;
- Prices during forecasted and actual shortages of operating reserves;
- Prices during the activation of real-time demand response;
- Frequency of price corrections; and
- Efficiency of real-time ex post prices.

At the end of this section, we provide a list of our conclusions and recommendations regarding the efficiency of real-time prices.

A. 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 by the real-time market in 2010 (because fast-start peaking units are relatively inflexible, they frequently do not set the real-time price even when they are the marginal source of supply).

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

This subsection describes how fast-start peaking units are committed by the real-time market 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 instructions to each generator for the next dispatch interval based on a short-term forecast of load and other operating conditions.[46] 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.[47] It is more efficient to allow UDS to start fast-start generators than to rely exclusively on operators to manually commit such units because UDS performs an economic optimization.[48]

When determining dispatch instructions for most online generators, UDS considers only incremental offers. However, for fast-start generators, UDS also considers commitment costs (since they must be committed from an offline state) and uses various assumptions regarding the

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

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

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

dispatchable range of the generator. The treatment of commitment costs and the dispatchable range have important implications for price setting by the real-time software (i.e., how real-time LMPs are determined). UDS schedules fast-start generators using the following criteria:

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

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

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

The following example illustrates the challenges for efficient pricing when fast-start generators are economically deployed by the real-time market. 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 2010

The following two analyses assess the efficiency of real-time pricing during periods when fast-start units were deployed in merit order in 2010. The first analysis summarizes 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 analysis evaluates how real-time prices would be affected if the average total offers were fully reflected in real-time LMPs.

The first analysis shown in Figure 20 summarizes the consistency of the average total offer (including no-load and start-up costs amortized for 1 hour) of fast-start generators that were deployed economically by UDS with the average real-time LMP over the initial 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. Figure 20 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 units in each month.

Figure 20: Comparison of Real-Time LMPs to Offers of Fast-Start Generators First Hour Following Start-Up by UDS, 2010

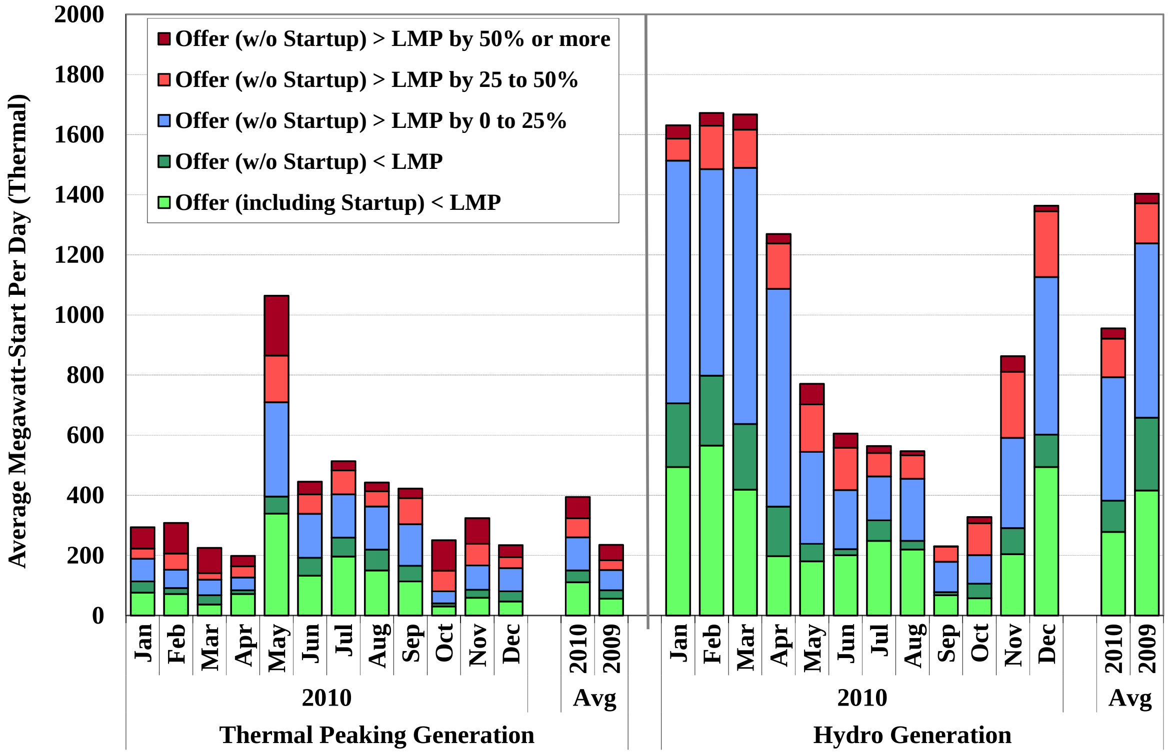


Figure 20 shows that flexible hydro generation accounted for the majority of fast-start generation that was started by UDS in 2009 and 2010. However, the operation of flexible hydro generation fell considerably from 2009 to 2010 because the largest flexible hydro generator was not operating from May through October.[50] Accordingly, the operation of thermal peaking resources rose substantially during this period.

The overall efficiency of real-time prices when fast-start resources are committed by UDS did not change significantly from 2009 to 2010. In roughly 70 percent of starts in both years, the average total offer was higher than the real-time LMP. This ratio is slightly higher for thermal peaking resources as discussed below, which may be due to the fact that the thermal peaking generators are generally less flexible than hydro generators. Hence, the full costs of the thermal peaking units frequently exceed the real-time LMPs.

50 The Energy Information Agency's EIA-923 database reports the monthly electricity output of individual generators at http://www.eia.doe.gov/cneaf/electricity/page/eia906_920.html.

Nearly 400 MW of thermal peaking generation were started each day in 2010, 72 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, 62 percent of the thermal peaking generation started exhibited offers that exceeded the average LMP over the minimum run time. Although thermal peaking generators are deployed in a relatively limited number of hours, they are frequently the marginal source of supply to the system in the hours that they run. This makes it particularly important to reflect the full cost of their deployment in real-time LMPs when they are deployed efficiently in merit order.

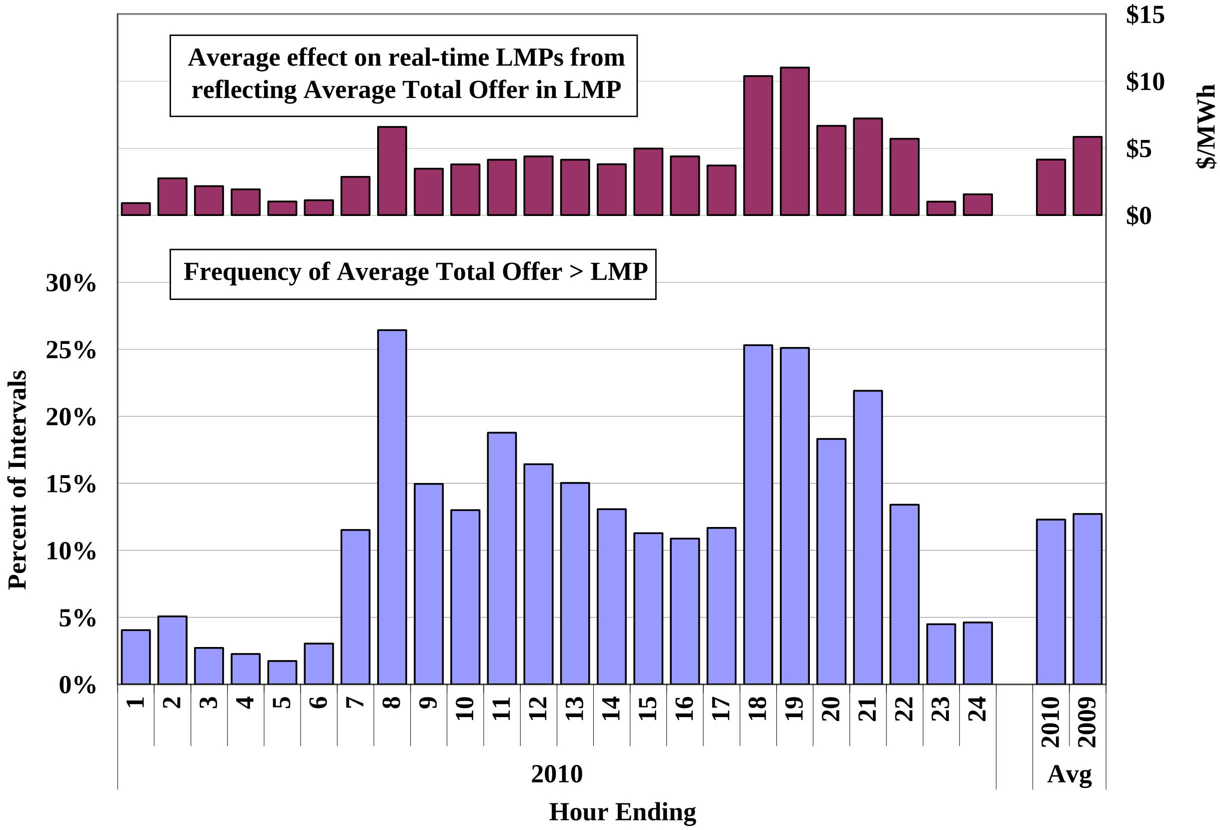
The average amount of thermal peaking capacity that was started by UDS increased by 68 percent from 2009 to 2010. The increase in commitment of thermal peaking units was partly due to the decline in surplus capacity in real-time from 2009 to 2010. As detailed in Section VI of the report, average surplus capacity fell notably in the second half of 2009 and remained below historic levels during 2010 because fewer resources were committed after the day-ahead market by the ISO for local reliability after July 2009 following several major transmission upgrades in Southeast Massachusetts and Connecticut. In addition, the outage of a large flexible hydro resource in New England increased the need to deploy thermal peaking units. If the amount of surplus capacity does not increase in the future, fast-start units will continue to play an increased role in satisfying real-time demand.

The prior analysis in Figure 20 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.[51]

[51] 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.

The analysis summarizes the portion of the fast-start units' costs that were not fully reflected in real-time LMPs in 2010. The lower portion of Figure 21 shows how frequently thermal and hydro fast-start units were started economically by UDS when their average total offers were greater than the LMP during the minimum run time in 2010.[52] 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.[53] 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 21: Difference Between Real-Time LMPs and Offers of Fast-Start Generators First Hour Following Start-Up by UDS, 2010



52 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.

53 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 21 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.

Fast-start units were started economically by UDS when their average total offer exceeded the real-time LMP over the minimum run time in 12 percent of all hours in 2010 and 13 percent of all hours in 2009. If the average total offers were fully reflected in the energy price in these hours, the average real-time LMP would increase approximately $4.15 per MWh in 2010 and $5.80 per MWh in 2009. The price effect would be largest in hour-ending 19 when the average LMP would rise by $11 per MWh.

However, it is important to note that these values are likely overstated 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 21, 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 net 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 over the long term by facilitating more efficient contracting and investment.

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.[54]

B. Real-Time Operation and Pricing during Forecasted and Actual Shortages

In the real-time market, the Reserve Constraint Penalty Factors ("RCPFs") limit the costs that the model may incur to meet the reserve requirements (i.e., marginal dispatch actions that would exceed the relevant RCPF are foregone). Consequently, if the cost of maintaining the required level of a particular reserve exceeds the applicable RCPF, the real-time market model will allow a reserve shortage and set the reserve clearing price based on the level of the RCPF.[55] For example, suppose an online generator with a $60 per MWh incremental offer could be backed down to provide reserves when the LMP is $160 per MWh. In this case, the marginal cost to the system of providing reserves from this unit is the opportunity cost of the unit not providing energy at the LMP. This opportunity cost is equal to the difference between the LMP and the incremental offer of the unit or $100 per MWh in this example ($160 per MWh LMP minus $60 per MWh incremental cost). If the RCPF is $50 per MWh, the market will not back the unit down to provide reserves and the system would be short of reserves since the marginal system cost of doing so ($100 per MWh) exceeds the RCPF ($50 per MWh).

The RCPF levels are important because they determine how the real-time market responds under tight operating conditions. When it is not possible to meet the reserve requirements, the RCPFs prevent the model from incurring extraordinary costs for little or no reliability benefit. However, if RCPFs are not sufficiently high, the model may not schedule all available resources to meet the reliability requirements and real-time clearing prices may not adequately reflect the market conditions when this occurs. In such cases, like the example above, the operator will likely intervene to maintain reserves and significantly affect market clearing prices in the process.

---

54 Extended LMP ("ELMP"): Real Time Market and Settlements, Stakeholder Workshop, July 1, 2010

55 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.

Hence, it is important to evaluate the RCPF levels periodically to determine whether modifications are warranted.[56] In this section, we evaluate the $100 per MWh RCPF for system-level 30-minute reserves to determine whether it is consistent with the costs of the operating actions the ISO takes to maintain the required level of 30-minute reserves.

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.

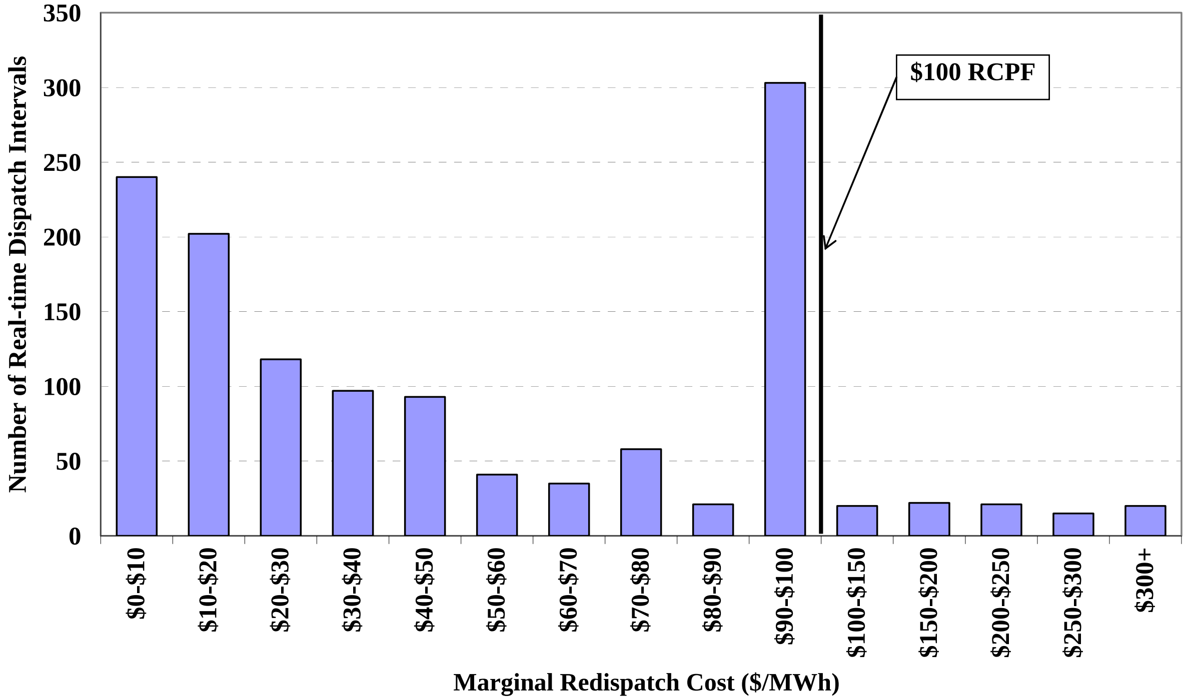
There are at least three ways for the ISO to maintain the required level of system 30-minute reserves when the real-time model does not schedule all available reserves. The ISO can:

- Curtail exports to neighboring areas, which allows internal units to reduce output in order to provide more reserves;
- Manually dispatch online generators that have some available capacity that cannot otherwise be ramped-up quickly enough to provide 30-minute reserves; or
- Manually commit slow-start generators to bring additional capacity online, although the operator's ability to do this is limited by the lead times necessary to start available offline generators.

The following analysis compares the $100 per MWh RCPF for system-level 30-minute reserves to *some* of the redispatch costs incurred to meet the requirement during 2010. Specifically, the analysis shows the maximum of: (i) the shadow price of system-level 30-minute reserve constraint, which is limited by the RCPF, and (ii) the marginal cost of export curtailments that is implied by the price difference between New York and New England at the border. However, the figure does not include the redispatch costs that are incurred when the ISO manually adjusts the output of online generators or manually starts offline generators. Each bar shows how frequently the marginal redispatch costs were in each range shown on the x-axis for 2010.

56 Accordingly, the local RCPF levels were increased from $50 per MWh to $250 per MWh in January 2010 after it was concluded that the previous local RCPF was not sufficiently high to schedule available resources to satisfy the local reserve requirements under some circumstances, since it led the operators to intervene to maintain adequate reserves.

Figure 22: Redispatch Costs to Meet System 30-Minute Reserve Requirements 2010



The figure shows 1,963 intervals in 2010 when redispatch was done to protect 30-minute reserves at the system level. The results in Figure 22 indicate that the marginal cost of meeting system-level 30-minute reserve requirements exceeded the $100 per MWh RCPF in approximately 100 real-time pricing intervals. In other words, the value of the exports curtailment in the instances was higher than the value of the reserves protected.

This analysis understates the inconsistency between the operator actions and the RCPF for the following reasons:

- It does not include the costs of manual redispatch or supplemental commitments made to address potential 30-minute reserve shortages; and
- The OP-4 procedures call for operators to begin taking actions when a 30-minute reserve shortage is forecasted, so a number of these export curtailments occur when the system was not in shortage and may not have been in shortage without the curtailment.

These results indicate an inconsistency between the ISO's operating procedures and the RCPF for 30-minute reserves, which are more common than any other type of reserve shortage in New England. This is important because understated RCPFs can undermine the prices that prevail during shortage conditions. Pricing real-time shortages inefficiently has a number of undesirable effects on the market outcomes. It inefficiently affects imports and exports in both the real time and day ahead market, reduces the efficiency of generation commitments by reducing commitments through the day-ahead market, and understates the long-term price signals in New England that can affect investment and retirement decisions.

Hence, we recommend that the ISO perform an evaluation to improve the consistency between its operating procedures and the 30-minute reserve RCPS. If the RCP F is determined to be appropriate, the ISO should seek to revise its operating procedures to avoid actions substantially more costly than $100 per MW to protect its 30-minute reserves. However, if the operating procedures cannot be changed or are deemed to be appropriate, the RCPF for system-level 30-minute reserves be raised to a level that is more consistent with the costs of the operating actions it takes to maintain the 30-minutes reserve. Additionally, if the ISO must continue to take actions in response to a forecasted shortage rather than an actual shortage, it should modify its real-time pricing to ensure that it reflects the costs of these actions. Today, if costly actions are taken by the ISO in response to a forecasted shortage and these actions prevent the shortage, the real-time energy market prices will not reflect the costs of these actions or the value of the reserves that were maintained.

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

Price-responsive demand has the potential to enhance wholesale market efficiency in theory. 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 four 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

Prior to the beginning of the first Forward Capacity Commitment Period on June 1, 2010, the ISO was operating the following four *active* real-time demand response programs:

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

The first three programs were reliability-based programs that activated emergency demand response resources according to the OP-4 protocol during a capacity deficiency, and the resources received capacity payments for being available to do so.[57] 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, and it was the only one of the four that was originally extended beyond the start of the first Capacity Commitment Period under FCM.[58]

Many resources transitioned from one of the above programs to one of the following programs under the FCM:

57 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.

58 Resources in the Real-Time Price Response Program do not receive capacity payment. This program is 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 Demand Response – Demand resources comprising installed measures (e.g., products, equipment, system, services, practices, and/or strategies) at end-use customer facilities. These resources may be deployed by the ISO with 30 minutes notice.
- Real-Time Emergency Generation – Distributed generation whose federal, state and/or local air quality permit(s) limit their operation to hours when the ISO dispatches Real-Time Emergency Generation Resources. These resources may be dispatched by the ISO with 30 minutes notice.
- On-Peak Demand Resource – These typically consist of non-dispatchable measures that are not weather sensitive and reduce load across the per-defined hours. On-Peak Demand Resources measure their load reduction during: 1) Summer on-peak hours (1:00pm – 5:00pm on non-holiday weekdays from June to August), and 2) Winter on-peak hours (5:00pm – 7:00pm on non-holiday weekdays in December and January).
- Seasonal Peak Demand Resource – This is designed for non-dispatchable, weather sensitive measures (e.g., energy efficient HVAC measures). These resources must reduce load during non-holiday weekdays when the real-time system hourly load is equal to or greater than 90 percent of the most recent "50/50" system peak load forecast for the applicable Summer or Winter season.

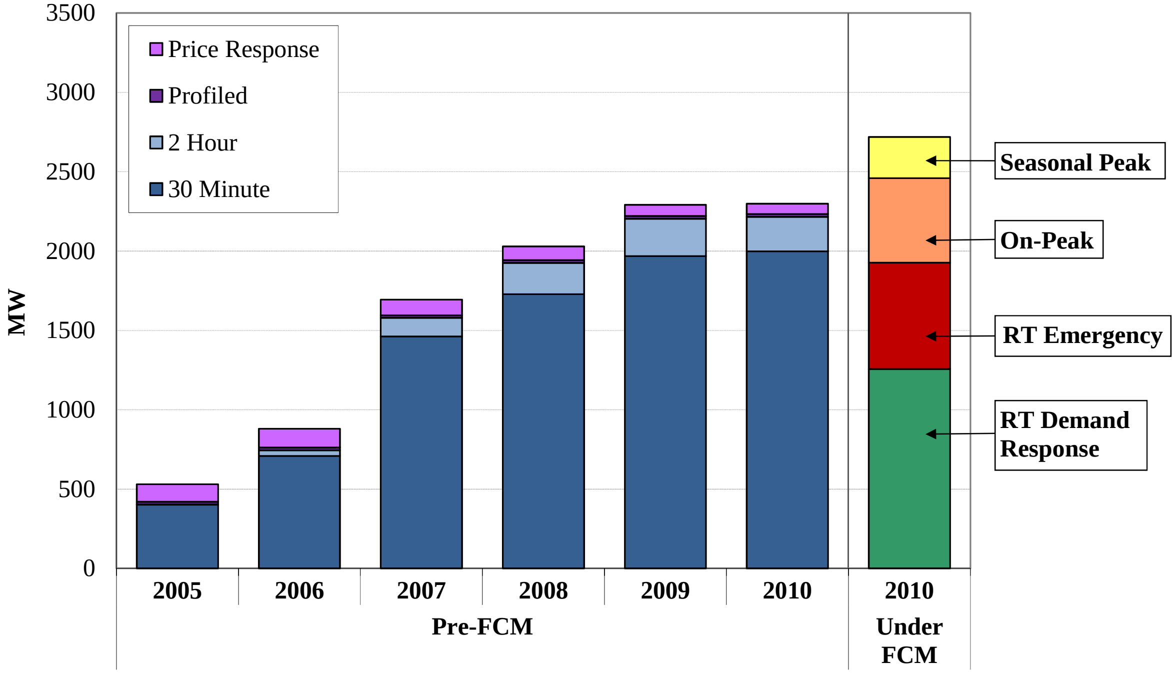
The first two are *active* (i.e., dispatchable) demand resources that operate based on real-time system conditions via dispatch by the ISO. They are defined at the Dispatch Zone level and reduce energy demand during OP-4 conditions.[59] The last two are *passive* (i.e., non-dispatchable) demand resources that are defined at the Load Zone level and reduce energy demand during peak hours.[60] Demand response participation has surged in New England in recent years. Figure 23 shows the quantity of resources enrolled in each of the real-time demand response programs from 2005 to 2010.[61]

59 There are 19 dispatch zones defined in New England: Northwest Vermont, Vermont, New Hampshire, Seacoast, Maine, Bangor Hydro, Portland ME, Western MA, Springfield MA, Central MA, North Shore, Boston, SEMA, Lower SEMA, Norwark-Stamford, Western CT, Northern CT, Eastern CT, and Rhode Island. Real-time demand response resources can be called under OP-4 Action 2, and real-time emergency generation resources can be called under OP-4 Action 6.

60 There are eight load zones defined in New England: Vermont, New Hampshire, Maine, Southeast Massachusetts, West Central Massachusetts, North East Massachusetts, Connecticut, and Rhode Island.

61 The quantities reported in this figure represent enrollments at the end of each year, except the quantities reported for pre-FCM periods during 2010 represent enrollments on May 31, 2010.

Figure 23: Real-Time Demand Response Program Enrollments 2005 – 2010



During the periods before the first FCM Capacity Commitment Period commenced, the quantity of enrolled resources increased from 530 MW in 2005 to 2,298 MW in 2010. Most demand response capacity was enrolled in the Real-Time 30-Minute Demand Response Program (87 percent at the end of May 2010). The enrollment in the Real-Time Price Response Program decreased over the period, from 110 MW in 2005 to 65 MW in 2010.

The FCM has attracted more passive demand response resources and less active demand response than the previously existing programs. Nonetheless, a total of 2,719 MW of demand resources were enrolled by the end of 2010, with 71 percent (or 1,927 MW) being active resources. Hence, capacity payments before and under FCM have encouraged the development of demand response resources, which is discussed in detail in Section VIII.

2. Real-Time Pricing During Demand Response Activation

The rise in demand response participation is beneficial in many ways, but it also presents significant challenges for efficient real-time pricing. Demand resources in the current real-time price response programs are not dispatchable and must be activated in advance based on forecasted conditions at least two hours ahead.[62,63] Active demand resources (real-time demand response and real-time emergency generation) are currently not dispatchable within the real-time dispatch software and cannot, therefore, set real-time energy prices. Instead, they are dispatched as part of OP-4 proceudres under Action 2 and 6. These inflexibilities can lead to inefficient real-time pricing.

The activation of demand response can inefficiently depress real-time prices substantially below the marginal cost of the foregone consumption by the demand response resources, particularly during shortages or near-shortage conditions. In 2009 and 2010, this problem 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. Nevertheless, these problems are likely to be more significant in the future as capacity margins fall and demand response resources must be deployed more , 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 2010. 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 2010, the average real-time clearing price was substantially lower than the average cost of activating these resources. Of the 502 hours when these resources were activated, the clearing price at the New England Hub was less than $100 per MWh in 72 percent of the hours and less than $70 per MWh in roughly half of the hours.

62 "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.

63 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. The current real-time price response program 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.

3. Demand Response 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 can provide an alternative to costly new generation investment. Demand response levels are likely to continue to grow in New England under the Forward Capacity Market discussed in Section VIII and due to the implicit subsidy contained in the Commission's Order 745 that addresses the compensation of demand response resources. This Order directs the ISO to pay the full LMP to real-time demand response resources that are activated in real-time when doing so satisfies the "net benefits test."

However, since the majority of demand response resources are not dispatchable in the real-time market, it can be challenging to set prices that efficiently reflect shortage or near-shortage conditions 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 would have been a shortage without the activation of demand response resources.

D. Ex Ante and Ex Post Pricing

Ex ante prices are produced by the real-time dispatch model (UDS) when it determines dispatch instructions, although the ISO uses ex post prices to settle with market participants in the real-time market. In this section, we examine inconsistencies between the ex ante and ex post prices, and we identify several factors that can undermine the efficiency of the ex post prices.

Ex ante prices are produced by the real-time dispatch model (UDS) that are consistent with the cost-minimizing set of dispatch instructions. They are consistent in the sense that the offer prices of dispatched resources are less than or equal to the LMP and the offer prices of un-dispatched resources are greater than or equal to the LMP. Hence, ex ante prices are set to levels that give generators an incentive to follow their dispatch instructions (assuming they are offered at marginal cost). Because they are consistent with the optimized dispatch, they are an efficient reflection of the prevailing market conditions.

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[64] 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

64 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 identifies three inconsistencies between ex ante and ex post prices in 2010:

- The current implementation of ex post pricing results in a small (0.3 percent) but persistent upward bias in real-time prices.
- Inconsistencies between ex ante and ex post prices do not improve the incentives of generators to follow dispatch instructions.
- Occasional distortions in the ex post prices 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 24 summarizes differences between ex ante and ex post prices in 2010 at a location close to the New England Hub.[65] 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 2010. 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 2010.

65 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 24: Average Difference Between Five-Minute Ex Post and Ex Ante Prices 2010

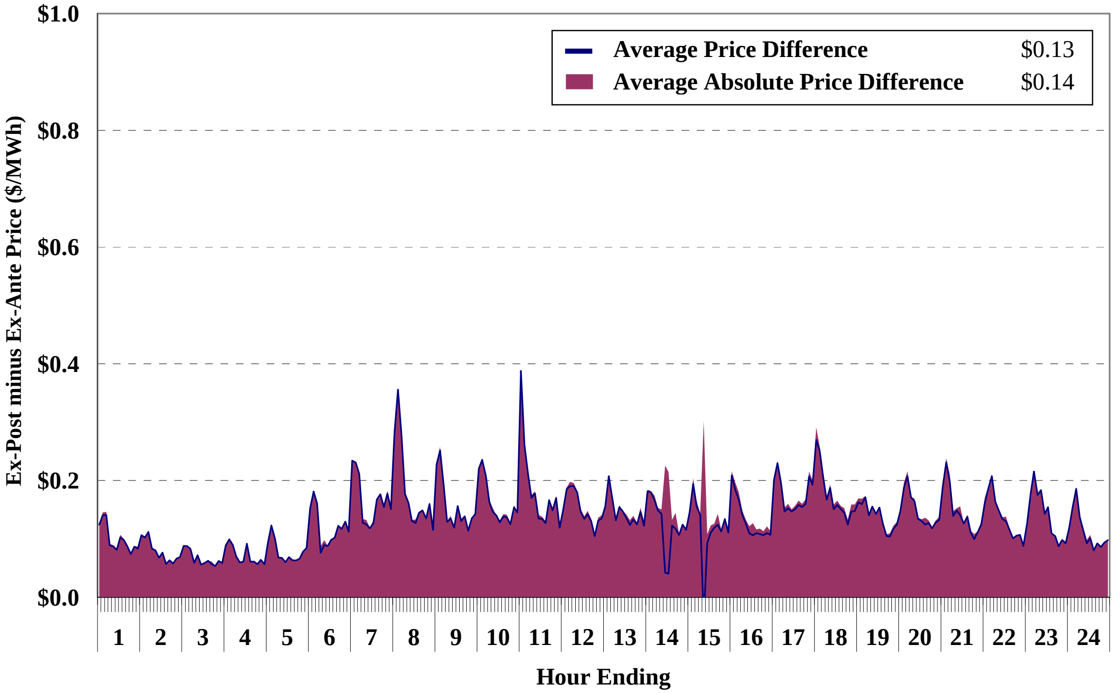


Figure 24 shows that the average ex post price is greater than the average ex ante price for every interval. This persistent bias is the result of the interaction between the following two factors. First, loss factors change slightly due to the time lag between the calculation of the ex ante and ex post prices. 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, which causes the ex post pricing model to move resources. Second, 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 a typical interval without congestion, four or five units are ramped down and 100 or so units are 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

Proponents have justified ex post pricing partly 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 because very little of their output is settled at real-time prices. 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.[66]

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

---

66 Uninstructed deviation penalties are penalties applied to suppliers that are not within a specified range of the dispatch instruction sent by ISO New England.

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, there are large differences between ex ante prices and ex post 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 in the ex post pricing.[67]

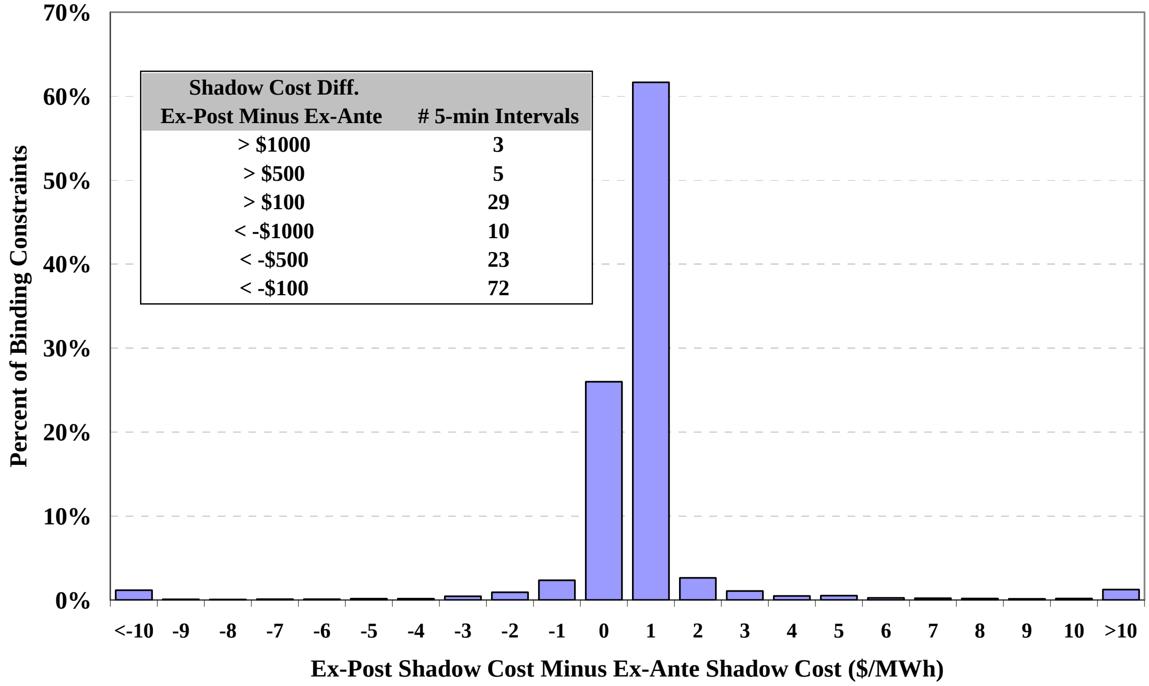
For example, suppose a combustion turbine with an incremental offer of $150 per MWh and an amortized start-up and no-load cost of $100 per MWh is started in order to resolve a load pocket constraint. Suppose that there is also a $50 per MWh unit in the load pocket that is dispatched at its maximum level. 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 25 summarizes differences in constraint shadow prices between ex post and ex ante calculations in 2010. A positive value indicates a higher shadow cost in the ex post calculation. For example, the value “2” on the x-axis means the ex post shadow cost is higher than the ex ante cost by $1 to $2.

---

67 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 25: Difference in Constraint Shadow Costs Between Ex Post and Ex Ante All Binding Constraints, 2010



The average difference was not significant in 2010. Nearly 98 percent of all differences were within $10 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 37 intervals during which ex post shadow prices were at least $100 per MWh higher than ex ante prices, and 105 intervals during which ex post shadow prices were at least $100 per MWh lower than ex ante prices. These results can be attributed partly to the very low levels of congestion that currently prevail in the ISO New England markets. However, as load grows and transmission congestion increases, we expect that these instances will also increase.

4. Conclusions regarding Ex-Post Pricing

Our evaluation of the ex post pricing results indicates that the real-time ex post prices:

- Are slightly biased in the upward direction in uncongested areas;
- Introduce small potential inefficiencies when they are not consistent with dispatch instructions; and

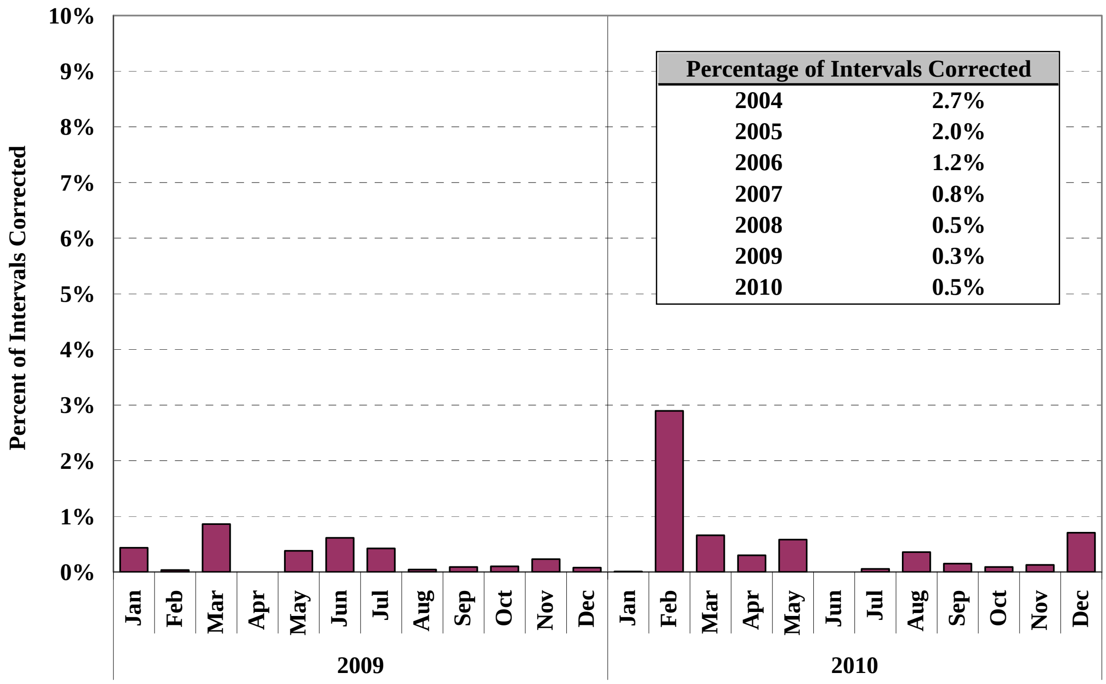
- Sometimes distort the value of congestion into constrained areas.

The primary 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 that ex post prices are sometimes set at inefficient levels, we recommend that the ISO consider modifying the inputs from UDS to the ex post pricing model to improve the consistency of the ex post and ex ante prices.

E. Real-Time Price Corrections

This subsection evaluates the rate of real-time price corrections during 2010. 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 26 shows the rate of real-time price corrections in New England in each month of 2009 and 2010. The inset table shows the annual rate of price corrections in the past seven years.

Figure 26: Rate of Real-Time Price Corrections
2009 – 2010



The figure shows that real-time price corrections were infrequent in both 2009 and 2010. The rate was less than one percent in nearly every month during 2009 and 2010. The lone exception was February 2010 when the rate of price correction was nearly 3 percent. This was caused by price corrections for 20 hours on one day due to software errors, which affected the LMPs at a very limited set of pricing nodes. The annual rate of price corrections has declined since 2004 and has been at or below 0.5 percent in recent years. It is also notable that nearly 40 percent of the intervals that experienced price corrections in 2009 and 2010 were due to issues with the real-time software's Dead Bus Logic, which affects the LMPs at very few pricing nodes.[68] 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 four years, supporting the conclusion that the real-time market software for the New England wholesale market has functioned well.

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 2010. Our evaluation leads to the following conclusions and recommendations:

---

68 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.

- 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 marginal cost of meeting system-level 30-minute reserve requirements can exceed the $100 per MWh RCPF, requiring the ISO to curtail exports and take other manual actions outside the market. This can lead to inefficiently low real-time prices that do not properly reflect the cost of maintaining reliability.

  ✓ We recommend that the ISO perform an evaluation to improve the consistency between its operating procedures and the 30-minute reserve RCPS, which should result in modified operating procedures or a higher RCPF for system-level 30-minute reserves. This will provide market participants efficient incentives to schedule resources that efficiently maintain reliability.

- 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 that ex post prices are sometimes set at inefficient levels, we recommend that the ISO consider modifying the inputs from UDS to the ex post pricing model to improve the consistency of the ex post and ex ante prices.

- Price corrections were very infrequent in 2010, 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.

VI. 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 satisfy these requirements in a manner that is economically efficient. 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.
- Uplift Allocation and Virtual Trading – This evaluates the allocation of uplift to virtual transactions and how the allocation affects the incentives of virtual traders.

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).

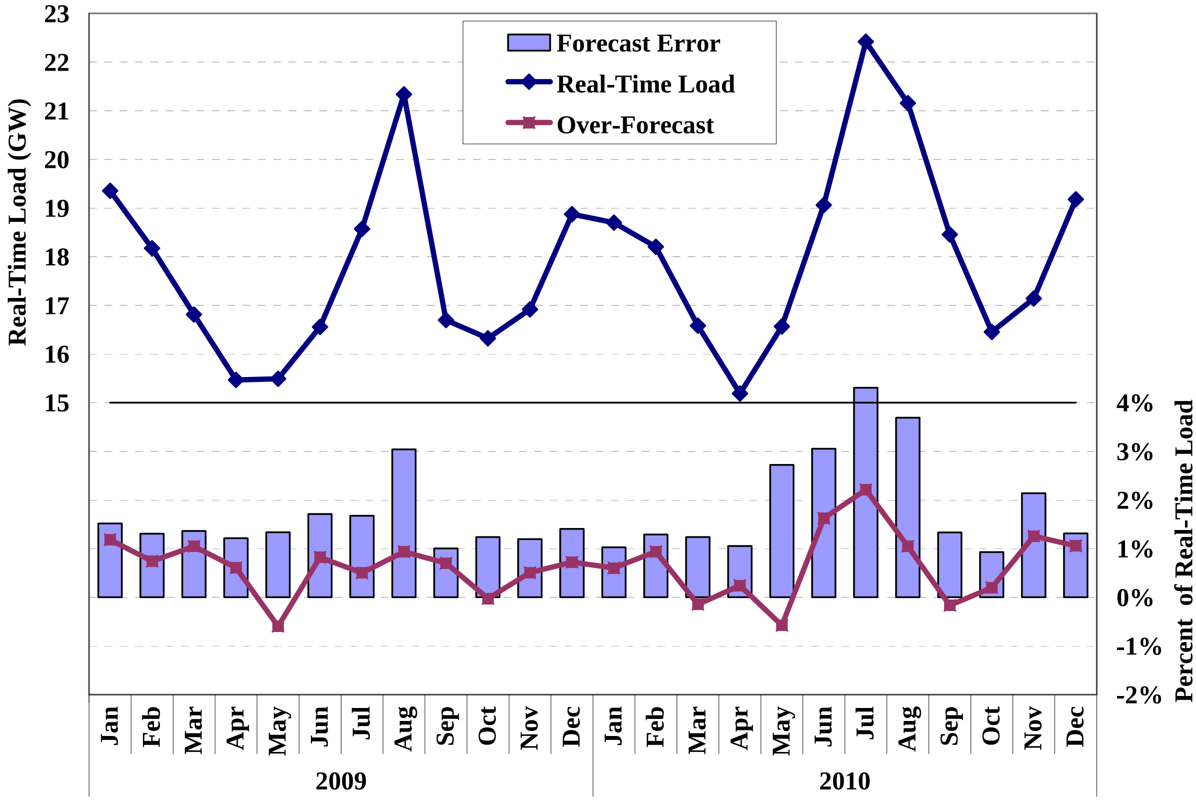
Accurate load forecasts promote efficient scheduling and unit commitment. Inaccurate load forecasts can cause the day-ahead market and/or the ISO to commit too much or too little capacity, which can affect prices and uplift. Therefore, it is desirable for the day-ahead forecast to accurately predict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

Figure 27 summarizes daily peak loads and two measures of forecast error on a monthly basis during 2009 and 2010. The *Over-Forecast* is the percentage by which the average day-ahead forecasted daily peak load exceeded the average real-time daily peak load[69] in each month.

69 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 27.

Positive values indicate over-forecasting on average and negative values indicate under-forecasting on average. The *Forecast Error* is the average of the absolute difference between the day-ahead forecasted daily peak load and the actual daily peak load, expressed as a percentage of the average actual daily peak load.

Figure 27: Average Daily Peak Forecast Load and Actual Load Weekdays, 2009 – 2010



The figure shows a characteristic pattern of high loads during the winter and summer and mild loads during the spring and fall. Overall, load increased notably from 2009 to 2010. The annual peak load of 27.1 GW occurred on July 6, 2010, up 8 percent from the peak load of 25.1 GW in 2009. The average load also rose 3 percent, from 14.6 GW in 2009 to 15.0 GW in 2010. In addition, the frequency of peak load conditions exceeding 20 GW increased from 250 hours in 2009 to 531 hours in 2010. The increased load levels in 2010 were primarily due to hotter summer weather and by improved economic conditions.

The ISO's day-ahead load forecasts are very consistent with actual load, although the ISO tends to slightly over-forecast load on average. The average over-forecast was comparable in the two years: 0.6 percent in 2009 and 0.7 percent in 2010. The ISO regularly evaluates the performance of its load forecasting models to ensure there are no factors that bias the forecast unjustifiably.[70]

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 2009 to 2010, the average forecast error rose modestly 1.5 percent to 2.0 percent. The increase in forecast errors was more evident during the summer months when the forecast error rose from 2.1 percent in 2009 to 3.7 percent in 2010. This is, however, expected as forecast errors tend to increase at higher load levels. These levels of forecast error are still relatively small and we find that the load forecasting performance of the ISO remains good overall.

B. Commitment for Local and System Reliability

In New England, sufficient resources must be available to satisfy local and system reliability requirements. To ensure reliability at the system level, sufficient online and quick start resources are needed to satisfy forecasted load, to recover from the largest single contingency, and to recover from 50 percent of the second-largest single contingency. To ensure that local areas can be served reliably, a minimum amount of capacity must be committed in each load pocket (i.e., import-constrained area). Specifically, sufficient online capacity is required to: (i) 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); (ii) ensure that

70 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.

reserves are sufficient in local constrained areas to respond to the two largest contingencies; (iii) support voltage in specific locations of the transmission system; and (iv) manage constraints on the distribution system that are not modeled in the market software (known as Special Constraint Resources ("SCRs")).

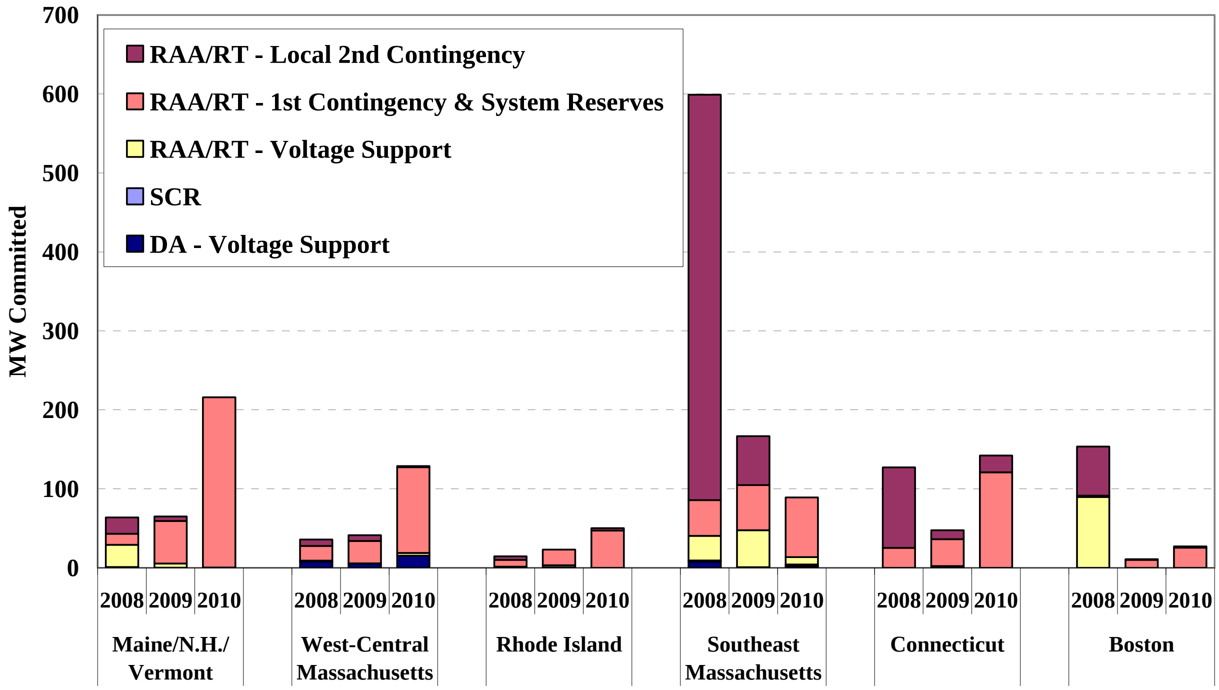
In the day-ahead market, generators are scheduled based on the bids and offers submitted by buyers and sellers. A generator is committed when demand bids from load serving entities and virtual traders are high enough for the unit to be economic given its start-up, no-load, and incremental offer components. The willingness of load serving entities and virtual traders to buy (or sell) power in the day-ahead market is partly based on their expectations of LMPs in the real-time market on the following day. Thus, the resulting 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 and system-level reliability requirements. Once the commitment costs have been incurred, these generators may be inexpensive providers of energy and reserves. 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 when generators are committed for reliability. 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, which reinforces the tendency of the day-ahead market-based commitment to not satisfy 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 reliability requirements in the real-time market, reducing the need for supplemental commitment. Nonetheless, supplemental commitments are still needed to meet reliability requirements, so it is important to continue evaluating potential market improvements. This section summarizes the pattern of supplemental commitment for reliability in the past three years and discusses several initiatives by the ISO to reduce the frequency and effects of supplemental commitment.

Figure 28 shows the average amount of capacity committed to satisfy local and system-level requirements in the daily peak load hour in each zone from 2008 to 2010.[71] The category *RAA/RT – First Contingency & System Reserves* shows capacity committed for local first contingency protection and for system-level reserve requirements together since the ISO does not maintain data that distinguishes between these two reasons for commitment. 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 28: Commitment for Reliability by Zone
Daily Peak Hour, 2008 – 2010



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

71 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.

Supplemental commitment for local reliability (i.e., local second contingency) has fallen significantly since 2008. The amount of supplemental commitment for local reliability decreased 70 percent from 2008 to an average of 300 MW in 2009, and continued to decrease in 2010, falling 40 percent to an average of 180 MW. These reductions were offset by increases in supplemental commitments to satisfy first contingency requirements and system-wide reserve requirements, which increased from an average of 203 MW in 2009 to 593 MW in 2010. The increased need for these types of requirements was partly due to the reduction in commitments for local reliability (that also provided reserves to the system) and partly due to the reduced system ramp capability associated with the extended unavailability of a large hydroelectric resource.

Most of the decrease in supplemental commitment for local reliability occurred in Southeast Massachusetts where the average quantities of these commitments decreased from 600 MW in the peak hour in 2008 to 165 MW in 2009 and to only 30 MW in 2010. This reduction in Southeast Massachusetts accounts for 70 percent of total decrease from 2008 to 2010. The substantial reduction in supplemental commitment in Southeast Massachusetts 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 was usually required to commit at least one large unit. As a result, the units were frequently committed for local reliability, incurring substantial NCPC payments. After the transmission upgrades were completed in mid-2009, the ISO no longer needs to commit additional generation for local reliability in this area after the day-ahead market, which has substantially reduced the amount of supplemental commitment and associated uplift costs.

In Boston, the average amount of supplemental commitment fell from 150 MW in 2008 to less than 10 MW in 2009 and 2010. This decrease was primarily due to revisions that the ISO made in early April 2008 in its operating guide for Boston-area reliability.[72] The revisions recognized

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

the reduced need to commit generation for voltage support following the transmission upgrades into Boston that were made in 2007.

In Connecticut, the average amount of supplemental commitment fell from 130 MW in 2008 to less than 50 MW in 2009 and 2010. This reduction 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 transfer capability into and within Southwest Connecticut, reducing the need to commit local capacity for reliability.

The analysis in this section highlights changes in the supplemental commitment patterns. Overall commitment for local reliability declined in New England by more than 90 percent from 2008 to 2010 due to substantially reduced commitment in Lower Southeast Massachusetts, Connecticut, and Boston. These reductions were primarily due to transmission upgrades in these areas that were completed between 2008 and early 2009, which have reduced the need to commit generation for reliability in these areas. These reductions were partially offset by increased levels of supplemental commitments to satisfy system-wide requirements, which on average rose from under 100 MW in 2008 to over 500 MW in 2010.

Variations in the pattern of supplemental commitments have substantially affected operations in several ways that are discussed later in this section. Subsection C illustrates how the quantities of out-of-merit dispatch (i.e., capacity producing output at a cost greater than the LMP) have changed. Subsections D and E show that the amount of surplus online capacity has decreased, which has affected real-time prices. Subsections F and E report the uplift charges resulting from reliability-committed units and how the allocation of uplift has affected virtual trading.

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.

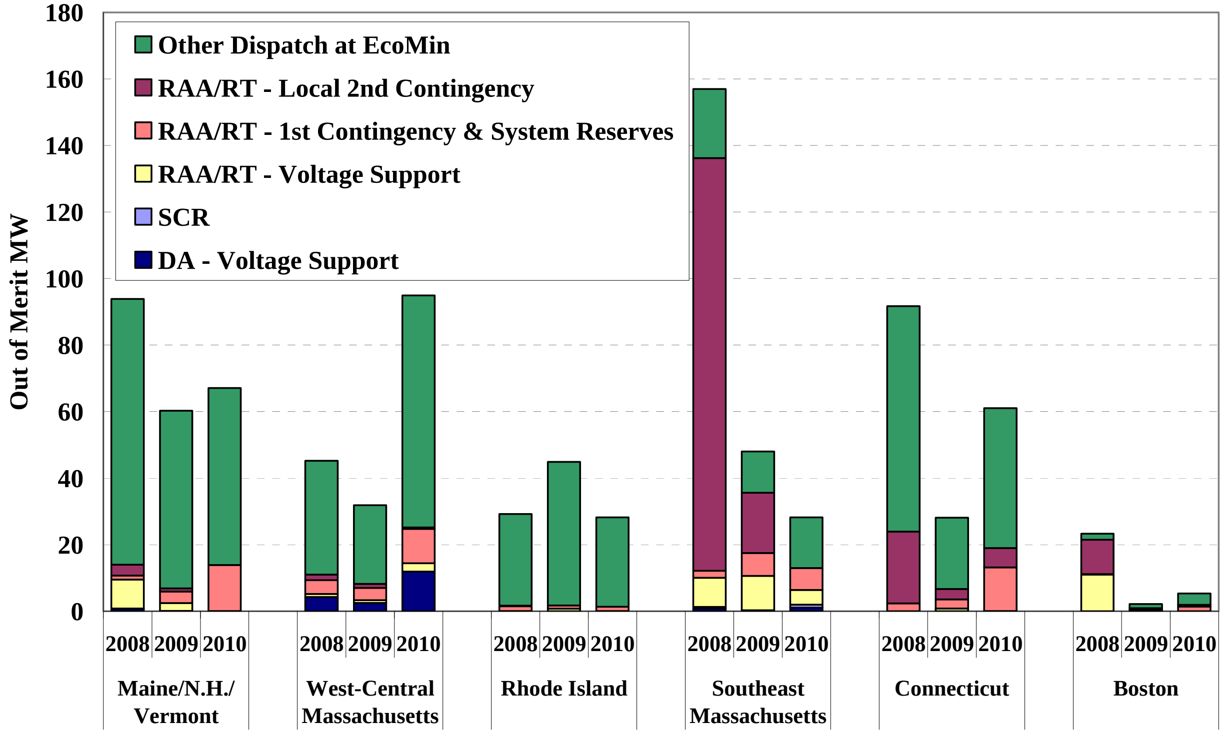
Out-of-merit generation occurs for several reasons. First, a unit may run at its EcoMin to satisfy its minimum run time after having run in-merit in previous hours or in anticipation of running in an upcoming hour. This is efficient because the software is minimizing cost over the total run-time of the 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.[73]

Figure 29 summarizes the average out-of-merit generation by location during peak hours (weekdays 6 AM to 10 PM, excluding holidays) in 2008, 2009, and 2010. The figure shows five categories of out-of-merit generation on units that are committed (and occasionally dispatched) for reliability reasons.[74] The figure also shows an "other dispatch" category that includes generation from units that were economically committed but are running at their EcoMin.

73 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 facilities that are not represented in the dispatch model; or (c) providing local reserves to protect against second contingencies.

74 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 29: Average Hourly Out-of-Merit Generation
Weekdays 6 AM to 10 PM, 2008 – 2010



In most regions, Figure 29 shows that the vast majority of the out-of-merit generation was attributable to non-local reliability units being dispatched at EcoMin in 2010. However, this was not the case historically in Boston and Southeast Massachusetts where most of the out-of-merit dispatch was from units committed through the RAA process for local second contingency protection or voltage support. For example, 91 percent of out-of-merit dispatch in 2008 and 65 percent of out-of-merit dispatch in 2009 in Southeast Massachusetts were from units committed through the RAA process for local second contingency protection or voltage support.

The average quantity of out-of-merit generation from units committed for local reliability (including second contingency, voltage support, and SCR) declined 85 percent from 2008 to 2010, from an average of roughly 200 MW in 2008 to 30 MW in 2010. The decline in out-of-market generation from units committed for local reliability tracked the changes in supplemental commitments and was caused by the same underlying factors. The reduced commitment for local reliability in Southeast Massachusetts, Connecticut, and Boston led to proportionate reductions in out-of-merit energy in those zones.

The amount of out-of-merit energy from units that were committed economically (i.e., Other Dispatch at EcoMin) declined from an average of 232 MW in 2008 to 155 MW in 2009, but rose in 2010 to 210 MW. The decrease from 2008 to 2009 was consistent with the reduction in reliability commitments during this period. The increase from 2009 to 2010 was attributable to the increase in supplemental commitments for system-wide reserve requirements and first contingency requirements discussed in the prior section.

D. Surplus Capacity and Real-Time Prices

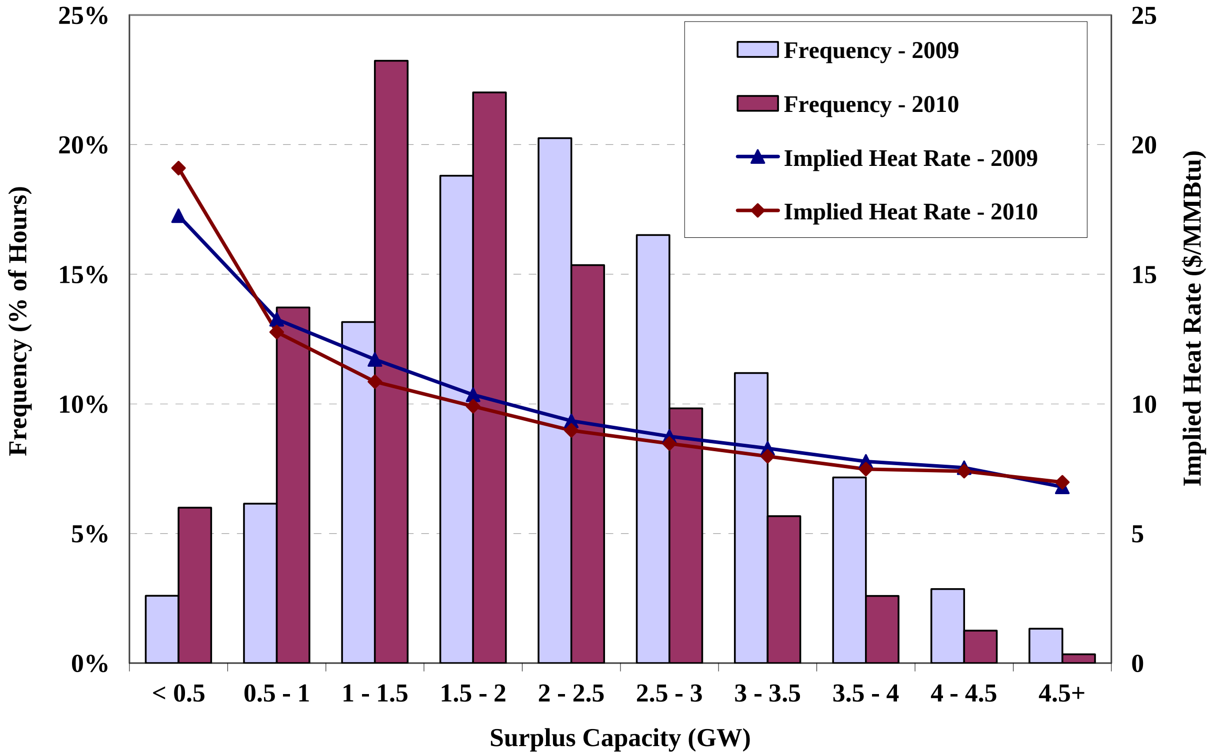
Under normal operating conditions, the available online and fast-start capacity is more than sufficient to satisfy load and reserve requirements, which suggests that some surplus capacity will exist in almost every hour. This is a normal outcome in a properly functioning market. Surplus capacity does 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 30 summarizes the relationship of surplus capacity to real-time energy prices at New England Hub in each peak hour of 2009 and 2010. Each bar shows the frequency of peak 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 13 percent of the peak hours in 2009 and 23 percent in 2010. 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 11.7 MMbtus per MWh in 2009 and 10.9 MMbtus per MWh in 2010. The implied marginal heat rate is shown in order to normalize real-time energy prices for changes in natural gas prices during 2009 and 2010.[75]

Figure 30: Surplus Capacity and Implied Marginal Heat Rates Based on Real-Time LMPs at the Hub in Peak Hours, 2009 – 2010



The figure shows a strong correlation between the quantity of surplus capacity and the implied marginal heat rate in real time. In 2010, the average implied marginal heat rate was highest (19.1 MMbtu per MWh) in hours with less than 0.5 GW of surplus capacity and lowest (7.0 MMbtu per MWh) in hours with more than 4.5 GW of surplus capacity.

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 2009 to 2010. The figure

75 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.

shows that average implied marginal heat rates were slightly lower in 2010 than in 2009 during hours with more than 0.5 GW of surplus capacity. However, when the surplus capacity was below 0.5 GW, the average implied marginal heat rates rose from 17.3 MMBtu per MWh in 2009 to 19.1 MMBtu per MWh in 2010, up more than 10 percent. This increase was primarily attributable to increased peak load conditions in 2010. The annual peak load rose 8 percent from 2009 to 2010, and the number of hours when load exceeded 20 GW increased from 250 hours in 2009 to 531 hours in 2010. As a result, high-cost generation was dispatched more frequently in 2010. Additionally, higher oil prices in 2010 contributed to higher implied marginal heat rates (based on natural gas) when oil-fired resources were needed to satisfy the system's needs. These conditions generally occurred at high load levels.

The figure shows significant reductions in the amount of surplus capacity from 2009 to 2010, 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 2.6 percent in 2009 to 6.0 percent in 2010. This is a significant development, since these hours exhibited substantially higher than average price levels with average implied marginal heat rates above 19 MMbtu per MWh. Likewise, the percentage of hours when there was less than 1 GW of surplus capacity increased from approximately 9 percent in 2009 to 20 percent in 2010. The reductions in surplus capacity primarily reflect the sizable reduction in supplemental commitments that has occurred over the past two years, as well as the increase in load and reduced imports from Hydro Quebec in 2010.

Although operating with lower surplus capacity levels increases real-time price volatility, it generally improves the efficiency and overall performance of the market because real-time prices will more completely reflect the true needs of the system and much lower costs will be incurred in uplift payments such as NCPC.

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. Both types of commitments can depress real-time prices inefficiently, while supplemental commitments by the ISO also lead to increased uplift costs.

As discussed earlier, transmission upgrades have substantially reduced the need for the ISO to commit generation to satisfy local reliability requirements. Since July 2009, the ISO's need to make supplemental commitments for local reliability has 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, the following two figures show 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 31 shows the first six months of 2010, and Figure 32 shows the last six months of 2010.

Figure 31: Daily Supplemental Commitments and Surplus Capacity January to June, 2010

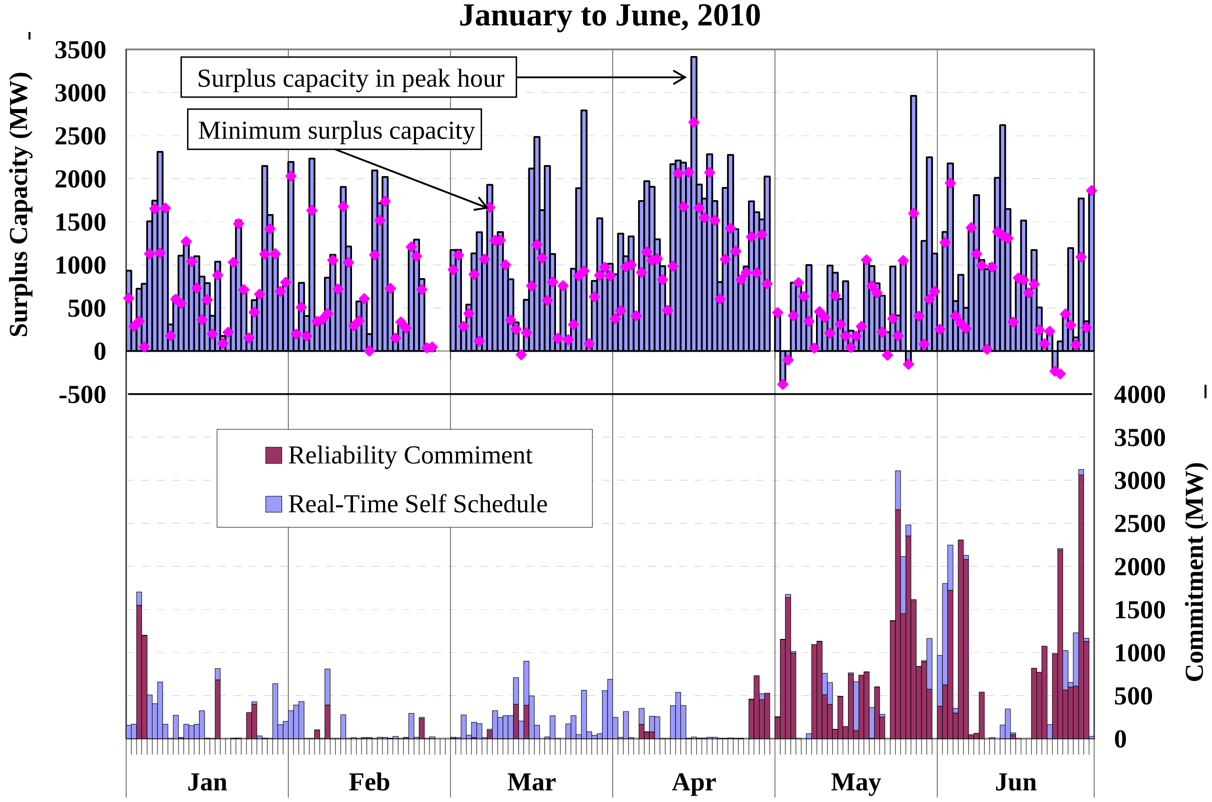
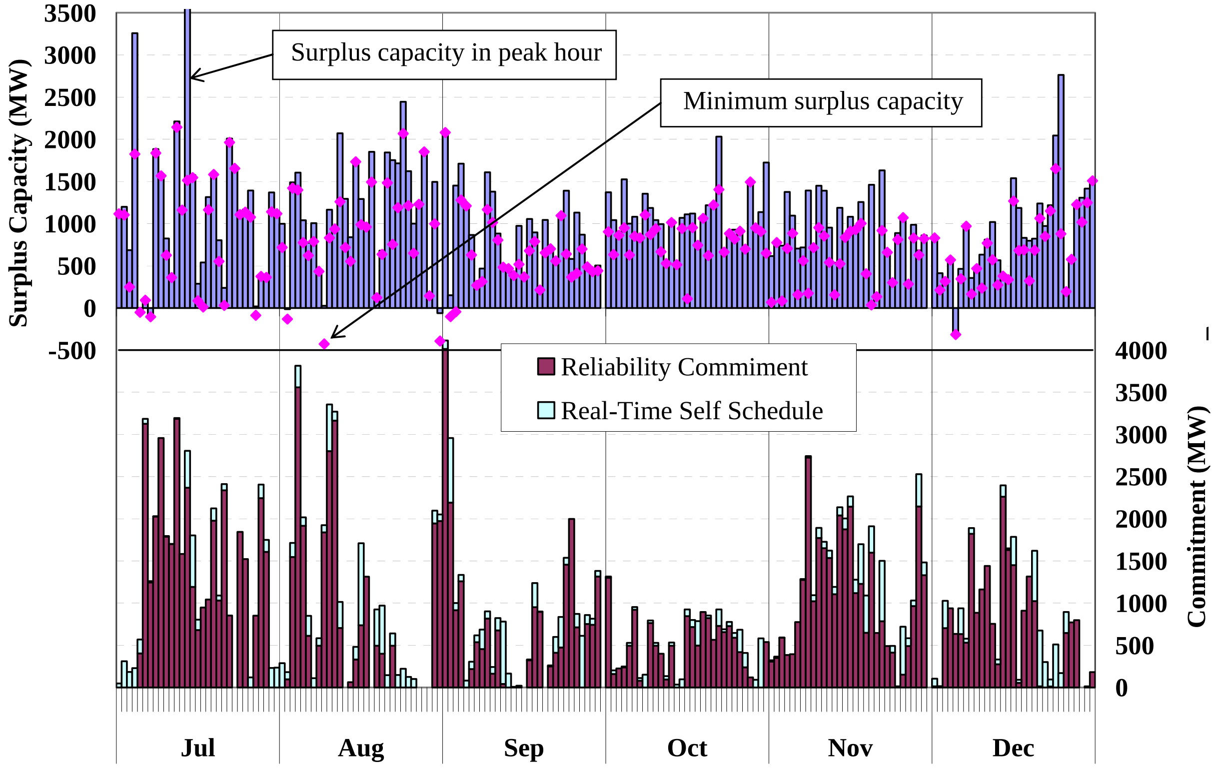


Figure 32: Daily Supplemental Commitments and Surplus Capacity
July to December, 2010



Large quantities of commitments for system-wide reliability requirements that lead to large surplus capacity levels are a concern because such commitments generally raise costs to New England's customers and distort real-time prices. Figure 31 shows that the ISO made very few supplemental commitments during the first four months of 2010, although the quantity of supplemental commitments increased substantially in May 2010 and continued into December 2010. These commitments were made primarily to maintain operating reserves at the system level and coincided with the period when a large flexible generator was unavailable to the market. The loss of this flexibility increased the need to bring additional resources online to provide dispatch flexibility to the system.

The minimum surplus capacity levels were low on most days when the ISO made supplemental commitments for system-wide capacity needs. However, there have been some days when large quantities of supplemental commitments have resulted in large quantities of surplus capacity. After reviewing the supplemental commitments and the surplus capacity levels that resulted from real-time operating conditions, we found that roughly 65 percent of the commitments in 2010 were needed to maintain system level reserves in retrospect.[76] To the extent that some of the reliability-committed capacity was not needed in retrospect, we identified several factors that have contributed to surplus capacity in these instances.

First, uncertainties tend to have a larger 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. Usually, most of the commitments of slow-starting units are made overnight, more than 12 hours before the forecasted peak.

There are two assumptions in the reliability commitment process that can make large contributions to the over-commitment on some days:

- The "desired capacity surplus" that operators have the discretion to determine to account for concerns regarding generator availability, load forecast errors, or other factors;[77] 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.

---

76 This is a simple evaluation that treats any surplus capacity (online and available offline capacity less the need to meet system load and reserve requirements) as "not needed" for the system. This simple evaluation tends to understate the necessity of supplemental commitments because: 1) the evaluation is based on hourly integrated peak rather than the higher instantaneous peak, and 2) the ISO cannot commit just a portion of a unit. For example, if the ISO needs an additional 200 MW of capacity to satisfy system reliability needs and commits the most economic unit with a capacity of 300 MW. In this evaluation, 100 MW of capacity would be deemed as "not needed".

77 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, the assumptions regarding imports and exports are based on the day-ahead scheduled transactions. By committing generation to support day-ahead exports, 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.

- The value of the day-ahead exports may not justify the costs of the supplemental commitments made to support them. This is particularly true when exports scheduled to New York when the difference in price on the New York side of the border is not significantly higher than on the New England (which represents the value of the export).

Hence, the ISO should consider whether its assumptions regarding imports and exports in its capacity evaluation process could be improved. The ISO New England is moving ahead with the New York ISO in implementing coordinated the Inter-Regional Interchange System (“IRIS”), which should rationalize the physical flow between the two markets in real-time. 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, making it unnecessary to commit generation to support day-ahead imports and exports.

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[78] and supplemental commitment to ensure reliability, particularly 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

78 Reliability agreements expired on June 1, 2010 when the first Forward Capacity Commitment Period began.

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 from 2008 to 2010. 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.[79] From January 2010 to May 2010, 31 percent of the capacity in Connecticut was under reliability agreements.
- Local Second Contingency Protection Resources – In 2010, 98 percent of the uplift from these units was allocated to Real-Time Load Obligations and Emergency Sales in the zone where the generator is located.[80] 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, export transactions, and wheel-through transactions.
- Economic and First Contingency Protection Resources – In 2010, 92 percent of this uplift was allocated to Real-Time Deviations throughout New England.[81] 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

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

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

81 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, virtual load schedules, and virtual supply schedules.

to local customers, while the uplift charges for Voltage Support Resources and other supplemental commitment are allocated to customers throughout New England.

The following table summarizes the total costs of uplift associated with reliability agreements and supplemental commitment from 2008 to 2010. The year-over-year changes in each category of uplift are shown as well.

Table 2: Allocation of Uplift for Out-of-Market Energy and Reserves Costs 2008 – 2010

	Millions of Dollars			% Change	
Category of Uplift	**2008**	**2009**	**2010**	**2008-2009**	**2009-2010**
Reliability Agreement					
Connecticut	$110	$61	$16	-45%	-74%
Other Areas	$19	$24	$6	26%	-73%
Local Second Contingencies					
Connecticut	$24	$2	$4	-91%	73%
Boston	$11	$0.3	$0.2	-97%	-50%
Southeast Massachusetts	$143	$14	$0.0	-90%	-100%
Other Areas	$4	$0.9	$0.1	-74%	-89%
Special Case Resources	$2	$0.6	$2	-61%	183%
Voltage Support	$29	$5	$5	-83%	4%
Economic*	$44	$33	$85	-27%	160%
Total	$387	$140	$117	-64%	-16%

* The category of Economic includes uplift for commitments made for system-wide reserve requirements and first contingency requirements.

Overall, uplift charges fell from $387 million in 2008 to $117 million in 2010. The largest decline occurred in uplift charges for local second contingency protection, which fell from $182 million in 2008 to only $4 million in 2010.

These sharp reductions in uplift charges were primarily due to several factors:

- Supplemental commitment for local reliability fell substantially from 2008 to 2010, due primarily to several transmission upgrades completed in Connecticut and Southeast Massachusetts during this period, which greatly reduced the amount of capacity requiring NCPC payments.

- Reliability agreement costs fell from $129 million in 2008 to $22 million in 2010, primarily because the associated reliability agreements expired. Several agreements expired in 2009 and all remaining agreements expired on June 1, 2010 when the first Forward Capacity Commitment Period began. A reliability credit of $2 million (that is not shown in the table) were paid to two units in Connecticut in 2010 under FCM because the ISO rejected their de-list requests from the first Forward Capacity Commitment Period for reliability purposes.
- Fuel prices fell significantly from 2008 to 2009, which reduces generators' costs and the NCPC payments necessary to make them whole. However, this effect was offset in 2010 as fuel prices increased.

Despite the overall reduction in uplift, NCPC payments to units committed for first contingency requirements and system-wide reserve requirements (i.e., "Economic NCPC") rose from $33 million in 2009 to $85 million in 2010. The majority of the uplift occurred after April 2010, consistent with the increase in supplemental commitment discussed earlier in this section.

G. Conclusions and Recommendations

We conclude that the ISO's operations to maintain adequate reserve levels in 2010 were reasonably accurate and consistent with the ISO's procedures. The amount of capacity committed for local reliability decreased significantly in Lower Southeast Massachusetts and Connecticut, while the amount of capacity committed for system-wide reliability increased considerably. The decline in local reliability commitment was primarily due to transmission upgrades that substantially increased the import capability into Southeast Massachusetts and Connecticut and greatly reduced the need to commit additional resources after the day-ahead market to satisfy local reliability requirements.

The increase in commitment for system reliability was due to several factors. First, local reliability commitments generally increase the total resources available for system reliability, so the reduced local reliability commitment led to more days when additional resources had to be committed in the RAA process to satisfy the system requirements. Second, the unavailability of a flexible hydro resource from May through October reduced the amount of offline reserves available to meet system reliability needs.

We regularly review patterns of supplemental commitment and the resulting out-of-merit generation because they can raise the following market issues:

- 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, and which may discourage participation in the New England market.
- Incentives for generators frequently committed for reliability to avoid market-based commitment in order 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 modifying the assumptions it makes regarding real-time imports and exports once it implements the IRIS process to improve the physical interchange with the New York ISO. 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 Section V 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.

VII. 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"), the design of 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.

The first Forward Capacity Auction ("FCA1") was held in February 2008, facilitating the procurement of installed capacity for the period from June 2010 to May 2011. Four auctions have been held to date, which have satisfied New England's planning requirements through May 2014. In June 2010, the start of the first Capacity Commitment Period allowed for the cessation of the individual reliability agreements that had been used extensively to maintain the resource requirements in Connecticut, Boston, and Western Massachusetts.

In the last year, the Commission has issued two orders on a set of market design-related matters pertaining to the Forward Capacity Market.[82] The orders' most significant determinations direct the ISO to:

- Model eight capacity zones corresponding to its eight Load Zones;
- Strengthen the supply-side market power mitigation rules;
- Extend the price floor in the auction through at least FCA6; and
- Develop buyer-side market power mitigation rules in order to address the shortcomings of the proposed Alternative Price Rule.[83]

82 See Order on Forward Capacity Market Revisions and Related Complaints, Docket ER10-787-000, *et al.* (Issued April 23, 2010). Also, see Order on Paper Hearing and Order on Rehearing, Docket ER10-787-000, *et al.* (Issued April 13, 2011).

In late May 2011, the ISO will file a proposed timeline for making these changes following an expedited stakeholder process. Hence, it is currently unclear what market rules will be in place before FCA5, which will be held in August 2011 for the period from June 2014 to May 2015.

This section of the report provides background on the FCM rules and evaluates the outcomes of the first four auctions.

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 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 ("Net CONE"), is the estimated capacity revenue that would be necessary for the investment to be profitable.[84]

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:

---

83 The Alternative Price Rule was a provision 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.

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

- *Installed Capacity Requirement* – The FCM procures the Net Installed Capacity Requirement ("NICR")[85] 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[86] installed capacity in each zone, less retirement and export bids, are compared to the zone's Local Sourcing Requirement ("LSR").[87] If the amount of capacity is greater than the LSR, the zone will not be modeled as a separate import-constrained zone in the auction. Export-constrained zones are always modeled in the auction. When the zonal requirements are modeled, the FCM produces locational prices that reflect the value of capacity in each zone.
- *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, and this review was recently strengthened to ensure that the expanded modeling of capacity zones would not provide opportunities for the exercise of market power. New capacity qualification rules and the three-year advance procurement feature allow new capacity projects to compete in the FCA.

---

85 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.

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

87 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.

B. Analysis of Forward Capacity Auction Results

Four 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"), the third in October 2009 for the commitment period of 2012/2013 ("FCA3"), and the fourth in August 2010 for the commitment period of 2013/2014("FCA4"). 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 and FCA4.

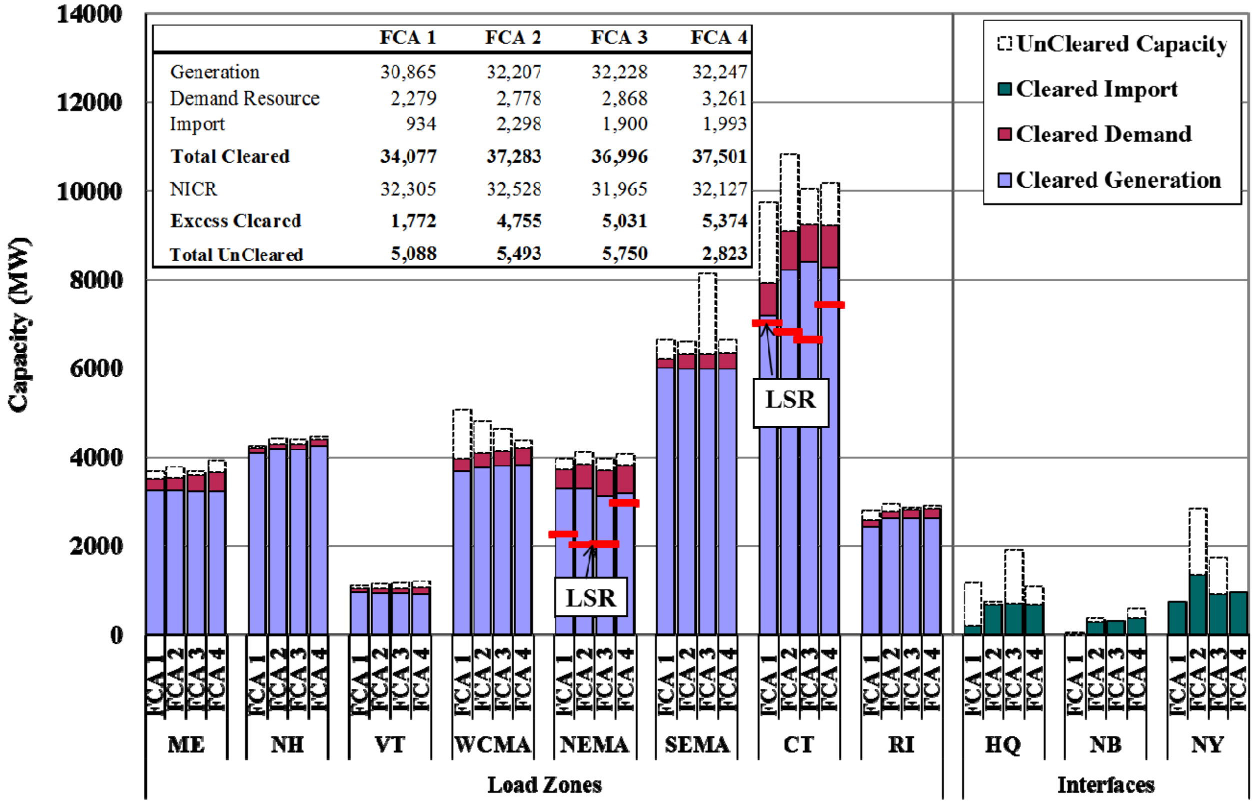
No import-constrained zones were deemed necessary because the amount of existing capacity exceeded the LSR in each area. Maine was modeled as an export-constrained zone in all four auctions, but there was no price separation between Maine and the rest of New England. This section summarizes and evaluates the overall results of the first four FCAs, the de-list bids of existing suppliers, and the procurement of new capacity.

1. Summary of Capacity Procurement

Figure 33 summarizes the procurements in the first four FCAs, showing the distribution of cleared and un-cleared capacity by location. Cleared resources are divided into generating resources, demand response resources, and imports from external areas.[88] 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.

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

Figure 33: FCM Auction Clearing Summary by Location
FCA1 – FCA4



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 1.8 GW. In FCA2, over 37 GW of resources were procured, exceeding the NICR by 4.8 GW. In FCA3, 37 GW of resources were procured, exceeding the NICR by 5.0 GW. In FCA4, over 37 GW of resources were procured, exceeding the NCIR by 5.4 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:

- 1.2 GW in FCA1, 2.3 GW in FCA2, 2.6 GW in FCA3, and 1.8 GW in FCA4 for Connecticut; and
- 1.5 GW in FCA1, 1.8 GW in FCA2, 1.7 GW in FCA3, and 0.9 GW in FCA4 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 un-cleared capacity. The un-cleared 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 slightly more than 100 percent in both FCA3 and FCA4 (i.e., cleared generating resources alone exceeded NICR in FCA3 and FCA4) . 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 7 percent of the NICR in FCA1, 9 percent in FCA2 and FCA3, and 10 percent in FCA4. Roughly 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 then decreasing modestly to 1,900 MW in FCA3 and 1,993 MW in FCA4.[89]

Substantial excess capacity cleared in the first four auctions as a result of the price floor that was originally stipulated in the Settlement Agreement. The price floor was originally supposed to be eliminated after FCA3, but it was extended until after FCA4. The price floor will be used at least through the sixth FCA. The price will likely clear at the floor in the next auction given the amount of excess supply and the vertical demand curve implicit in the FCM design.

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).

89 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.

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.

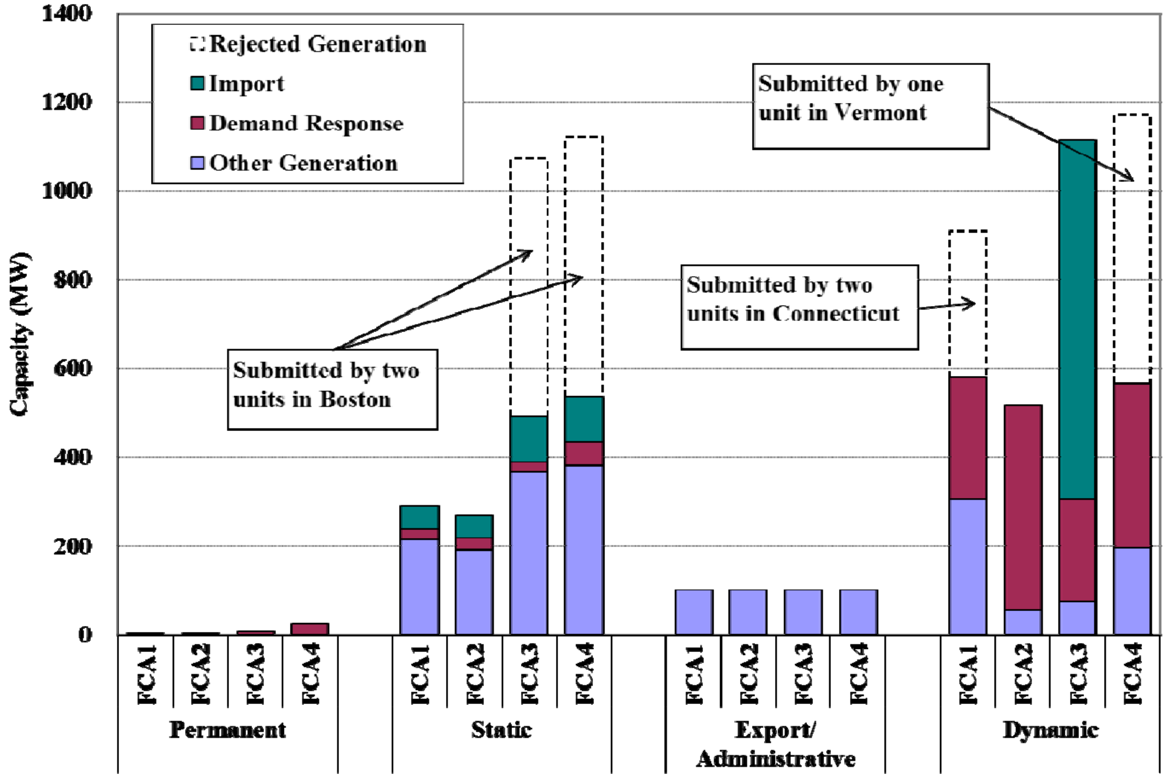
Figure 34 evaluates several categories of accepted de-list bids in the first four FCAs. The figure shows four categories of de-list bids: permanent, static, export or administrative, and dynamic.[90] Accepted de-list bids are also separated according to the type of resource: generation, demand response resources, and imports. The figure also shows the de-list bids that were rejected by the ISO-NE for reliability reasons.

The accepted de-list bids shown in Figure 34 range from 890 MW in FCA2 to 1710 MW in FCA3 in the first four auctions. Dynamic de-list bids accounted for the majority of accepted de-list bids in all four auctions. A dynamic de-list bid is 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 80 percent must be approved by the market monitor as consistent with the net going forward costs of the resource.[91] However, the threshold for evaluating de-list bids was changed to $1/kW-month by the recent Commission order.

90 Each category of de-list bid is defined in *Market Rule 1, Section 13.2.5.2*. Permanent de-list bids are submitted by resources intending to retire; static de-list bids are known in advance of the auction and can be approved by the Internal Market Monitor if they exceed 80 percent of CONE; export de-list bids are associated with resources whose capacity will be exported if not selected in New England; and dynamic de-list bids are not known in advance of the auction, but are associated with resources that may de-list at any time once prices fall below 80 percent of CONE.

91 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*.

**Figure 34: Summary of Accepted De-list Bids by Type
FCA1 – FCA4**



The total amount of accepted de-list bids from internal resources (both generation and demand) has been relatively stable in the four FCAs, ranging from 800 MW to 920 MW in the first three FCAs and rising to 1,130 MW in FCA4. With the exception of FCA2, most of these accepted de-list bids from internal resources are from generation resources (roughly two-thirds in the other three auctions). In each auction, roughly 100 MW of capacity de-listed in order to support capacity exports to New York over the Cross-Sound Cable. In FCA4, four units in historically congested areas de-listed their entire capacity. Otherwise, most of the de-listed generation was small output ranges on individual units. These levels are not surprising given the prevailing capacity surplus in New England. However, we expect that these levels will increase substantially in the future once the price floor is eliminated.

Seven de-list bids were rejected in the first four auctions. Two bids (330 MW) in Connecticut were rejected in FCA1, two bids (585 MW) in Boston were rejected in FCA3 and FCA4, and one bid associated (604 MW) in Vermont was rejected in FCA4. All seven de-list bids were rejected when the ISO determined in its Transmission Security Analysis that the units were needed for reliability. Since the rejected de-list bids were substantially smaller than the excess cleared capacity for all of New England in each of the three auctions in which bids were rejected, 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, FCA3, or FCA4. However, the rejection of the de-list bids has highlighted the following two market design concerns.

First, 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 and nearly 900 MW in FCA4. 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.

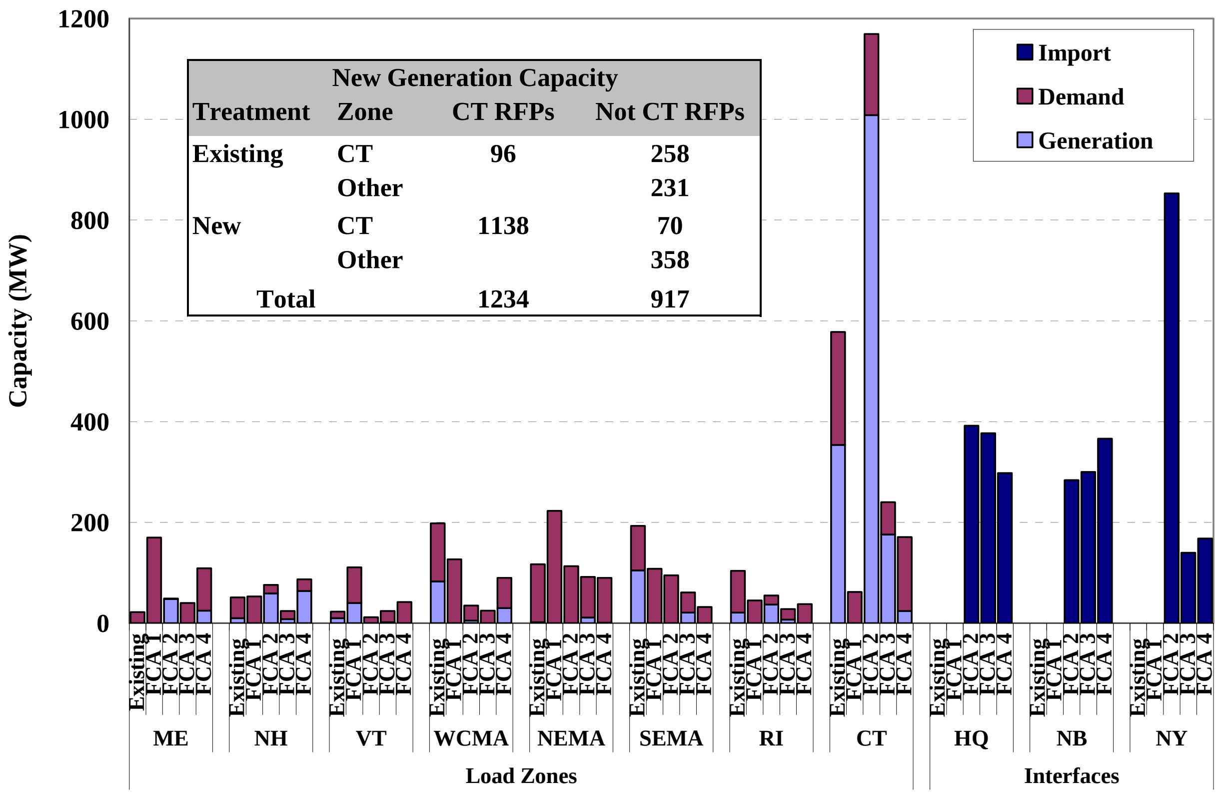
Second, the rejection of a de-list bid of a resource in Vermont shows that the need could arise for zonal price separation in areas other than Connecticut, Boston, and Maine. If all eight load zones are modeled in each FCA, then the clearing price in each zone will always be set appropriately, even if a large generator is suddenly unavailable to supply capacity.

Other than the two generators in Connecticut whose de-list bids were rejected, none of the generation that was under reliability agreements when an auction was held attempted to de-list in the first four auctions. Seventeen units (2,700 MW in total) were still under reliability agreements before the reliability agreements expired on June 1, 2010. Although the vast majority of this capacity was retained in the first four 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 its de-list bid price. It is good that out-of-market payments to retain capacity have fallen substantially starting with the first 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 35 shows the amounts of new capacity that were procured in the first four 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 through FCA4.[92]

Figure 35: New Capacity Procurement by Location in FCAs 1 through 4



New Generation Capacity			
Treatment	**Zone**	**CT RFPs**	**Not CT RFPs**
Existing	**CT**	**96**	**258**
	Other		**231**
New	**CT**	**1138**	**70**
	Other		**358**
Total		**1234**	**917**

92 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-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.[93]

In the first four FCAs, 6.9 GW of new capacity was procured from generation, demand response resources, and imports.[94] The following discussion reviews and evaluates the procurements of new capacity by resource type that are shown in Figure 35.

Import Capacity

The large quantity of new capacity sold by importers in FCA2-FCA4 indicates that the suppliers expected 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 arise if the heavier reliance on demand response were to result in much more frequent emergency load curtailments that are costly for demand response providers to satisfy. If this were to happen, it

---

93 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.

94 This excludes new resources treated as existing resources because they were already committed to enter.

would put upward pressure on capacity clearing prices or reduce the amount of capacity provided by demand response resources.

Additionally, demand response resources may not provide response comparable to supply resources during shortage or emergency conditions. The Internal Market Monitoring Unit recently conducted an evaluation of the performance of demand response resources when they were activated on June 24, 2010.[95] The evaluation concludes that there was wide variation in the performance of demand response resources with just 22 percent of resources curtailing an amount of load within 10 percent of the instructed amount, which is the performance threshold that is used for assessing uninstructed deviation penalties to generators. The results raise significant concerns about whether the demand response resources selling capacity in New England provide the same level of reliability benefits as internal generators and imports. It may be appropriate to reassess whether the performance criteria and settlements with demand response resources that do not perform as instructed should be more consistent with the criteria used for generation and imports.

The final 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 that demand response resources are not currently obligated to pay the Peak Energy Rent (PER) deduction to the ISO.[96] The fact that other types of new resources 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 (2,150 MW) has entered the market under FCM. Entry of generation resources would generally not be expected when the price clears at

95 See *2010 Annual Markets Report*, ISO New England, June 2011, Figure 3-31.

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

the price floor as it did in the first four 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 428 MW of new generating resources cleared in the FCAs that were not under the CT RFP or treated as existing resources. Most of these resources are facilities powered by renewable fuels, designed to up-rate existing resources, or made to re-power existing power plants. Such projects may have a lower Net CONE than most of the potential investments in new generation, which explains why they would clear at the floor prices in the first four 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 fourth Capacity Commitment Period exceeds the New England capacity requirement by nearly 17 percent, up from 5 percent in FCA1, 15 percent in FCA2 and 16 percent in FCA3. 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 the sixth FCA, the amount of excess capacity purchased in the auction may be greatly reduced.

It is still 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.

C. Forward Capacity – Conclusions

The Forward Capacity Market introduced by ISO New England in 2008 has operated 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 2014. 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.

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 four 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. Hence, we recommend addressing this inconsistency between the obligations of generation and demand response capacity resources.

VIII. Competitive Assessment

This section evaluates the competitive performance of the New England wholesale markets in 2010. 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. Based on the analysis presented in this section, 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.[97] We address four main areas in this section:

- 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.

Although there are some areas and conditions under which suppliers have market power, the ISO has market power mitigation measures that are employed to prevent suppliers from exercising market power. These measures have generally been effective, but it is still important to evaluate the competitive structure and conduct in the New England markets because participants with market power may still have the incentive to withhold at levels that would not warrant mitigation. A summary of our conclusions regarding the overall 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

97 See, *e.g.*, Section VIII of the "2009 Assessment of Electricity Markets in New England", Potomac Economics.

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).

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 ability and 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.[98] 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 because the supply is substantially less elastic in the higher cost ranges. 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

98 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.

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. 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 expected 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 historically 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:[99]

- All of New England;
- All of Connecticut;
- West Connecticut;
- Southwest Connecticut;
- Norwalk-Stamford, which is in Southwest Connecticut; and
- Boston.

This subsection analyzes the six 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.

[99] Lower SEMA was evaluated in prior reports, but is excluded from this report because the transmission constraints into the area was virtually eliminated since July 2009 when netowork upgrades were completed.

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. The analysis in this subsection shows several categories of supply and import capability relative to the load in each of the six regions of interest.

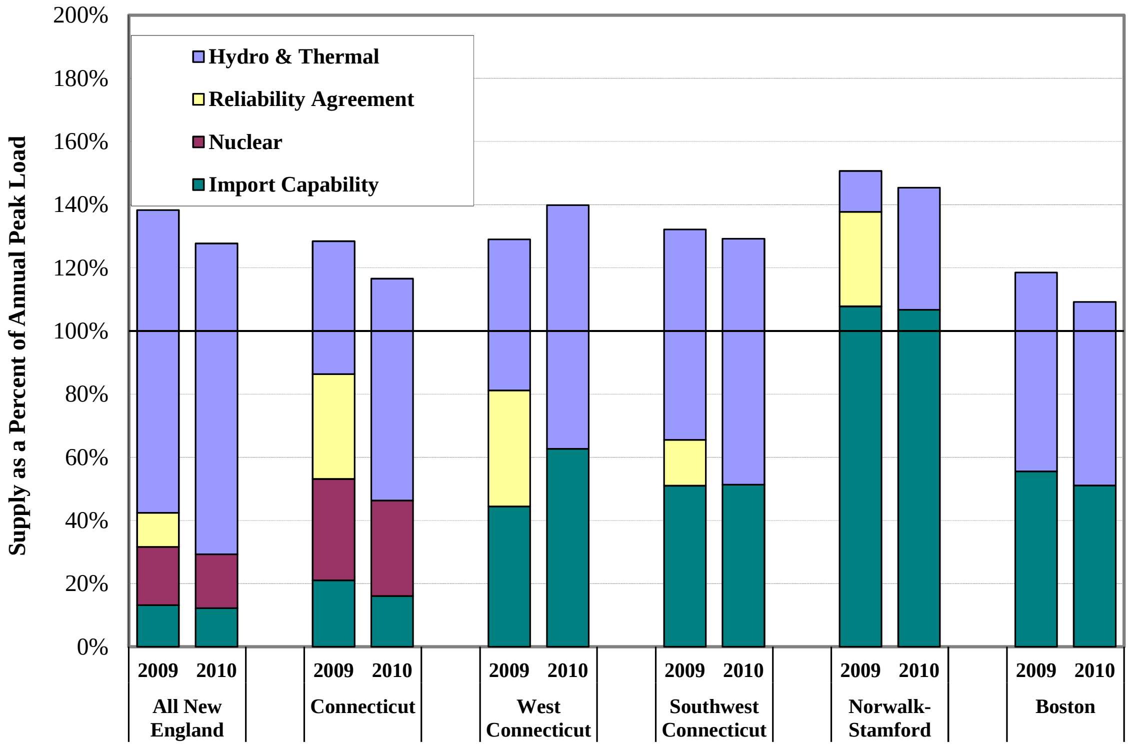
We differentiate between different types of supply because some types cannot feasibly be withheld to exercise market power. For convenience, the table below shows different categories of supply and provides comments regarding the feasibility of withholding them.

Table 3: Withholding by Type of Resource

Type of Resource	Comment
Nuclear	Nuclear resources pose fewer market power concerns than other types of resources do because they typically cannot be dispatched down substantially. This limits their owner's ability to withhold once a unit is online. They also generally have among the lowest marginal production costs making them costly to withhold.
Hydroelectric	Hydroelectric resources that can vary their output (i.e., reservoir and pump storage units) may be able to withhold. Smaller "run-of-river" hydroelectric facilities are generally more limited in their ability to change output level.
Fossil-Fired	Fossil-fired units generally have relatively wide dispatch ranges and marginal production costs that are close to the prevailing LMP. Hence, they are generally the easiest and least costly resources to withhold.
All Resources with Reliability Agreements	Supply under a reliability agreement cannot be economically withheld because the owner is obligated to offer the unit at short-run marginal cost. The supplier also has a strong disincentive to physically withhold because the fixed cost payments would be reduced if the unit failed to meet the target available hours specified in the reliability agreement. These agreements all expired in June 2010.

Figure 36 shows import capability and three categories of installed summer capability for each region: nuclear units, capacity under reliability agreements until June 2010, and all other generators.[100] These resources are shown as a percentage of 2010 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 12 percent of the load in the New England area under peak conditions, the five load pockets can serve larger shares of their peak load with imports. Norwalk-Stamford, which has the largest import capability relative to its size, was able to rely on imports to serve more than 105 percent of its load under peak conditions. This effectively eliminates it as an area of significant market power concern.

Figure 36: Supply Resources versus Summer Peak Load in Each Region 2010



100 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 36, the internal supply as a fraction of peak load decreased from 2009 to 2010. This is because the summer peak load levels rose 8 percent from 2009 to 2010, while there were very few changes to the supply of internal resources in each region. The amount of import capability into most regions also did not change significantly from 2009 to 2010 causing the import capability as a fraction of peak load to fall slightly as well.[101] The variations in import capability were primarily attributable to the differences in network topology (e.g., line outages), generation patterns, and load patterns during the peak load hours in the two years.

Therefore, supply conditions were slightly tighter overall in most areas in 2010. Figure 36 also shows the margin between peak load and the total available supply from imports and native resources. The total supply exceeded peak load in each region, ranging from 9 percent in Boston to 45 percent in Norwalk Stamford. Areas with lower margins may be more susceptible to withholding than other areas.

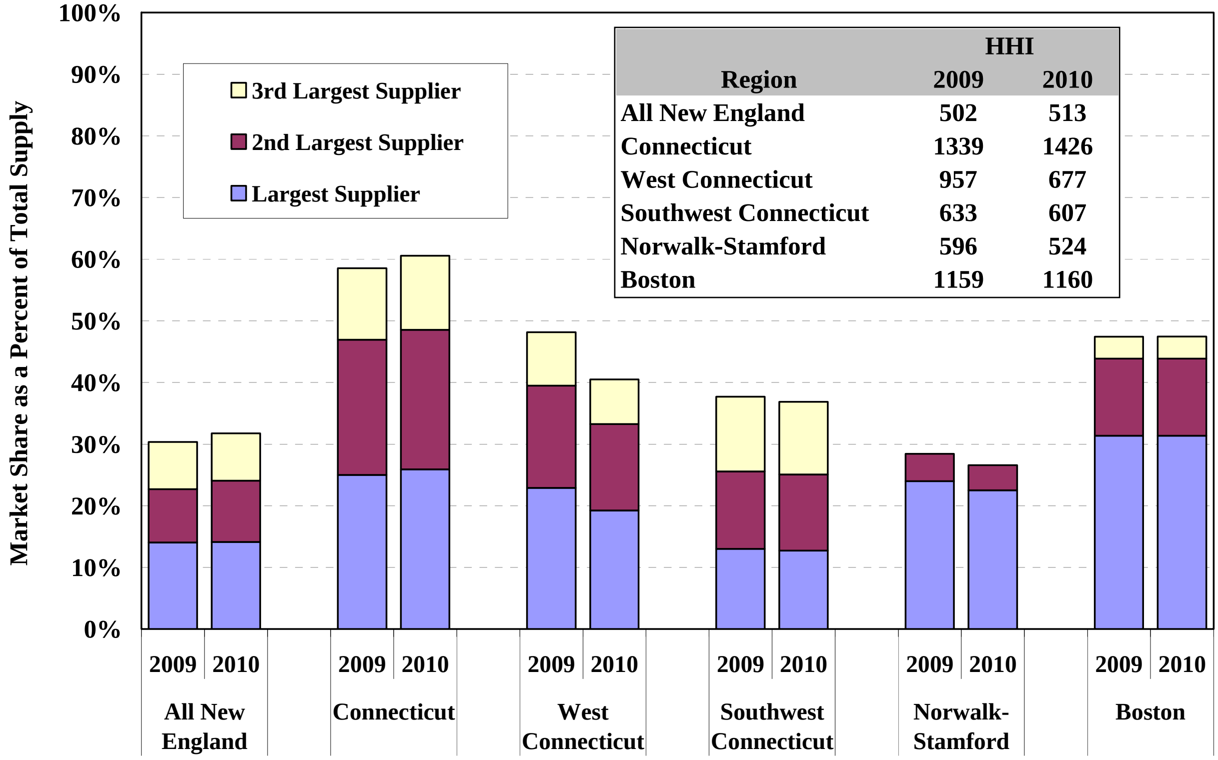
3. Market Shares and Market Concentration

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 37 shows the market shares of the largest three suppliers coinciding with the annual peak load hours in 2009 (on August 18) and 2010 (on July 6). 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

101 The transmission system in New England has evolved significantly over the past several years, particularly from 2006 to 2009 when several major transmission upgrades were completed in the historically constrained areas such as Boston, Connecticut, and Lower SEMA. These upgrades significantly improved the transmission system infrastructure and increased the transfer capability into affected regions.

area is not perfectly competitive, and because suppliers inside the area may be affiliated with resources in the market outside of the area.

Figure 37: Installed Capacity Market Shares for Three Largest Suppliers August 18, 2009, and July 6, 2010



Region	HHI 2009	HHI 2010
All New England	502	513
Connecticut	1339	1426
West Connecticut	957	677
Southwest Connecticut	633	607
Norwalk-Stamford	596	524
Boston	1159	1160

Figure 37 indicates a substantial variation in market concentration across New England. In all New England, the largest supplier had a 14 percent market share in 2010. In the load pockets, the largest suppliers had market shares ranging from 13 percent in Southwest Connecticut to 31 percent in Boston in 2010. Likewise, there is variation in the number of suppliers that have significant market shares. For instance, Norwalk-Stamford had only two native suppliers with very different market shares in 2010, while Southwest Connecticut had three native suppliers with comparable market shares.

The figure shows that these market shares did not change significantly from 2009 to 2010, because there were very few changes to the supply of internal resources in each region and the import capability into each region remained similar, absent of notable transmission upgrades. The small differences in market shares on the peak load day between 2009 and 2010 were mostly attributable to the variations in the import capability associated with differences in network topology, generation patterns, and load patterns on the two days.

The HHI figures suggest that all areas in New England were not highly concentrated in 2010.[102] The HHI for Norwalk-Stamford is 524, 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 served by imports, the need for local suppliers is very limited. Of the remaining areas, Connecticut and Boston have the highest HHI statistics in 2010, with 1426 and 1160, 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.

4. 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.[103, 104] 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

102 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.

103 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.

104 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.

relative to available supply-side resources. Since electricity cannot be stored economically in large volumes, production needs to match demand in real time. When demand rises, an increasing quantity of generation is utilized to satisfy the demand, leaving less supply that can respond by increasing output if a large supplier withholds resources. Hence, markets with higher resource margins tend to be more competitive, which is not recognized by the HHI statistics.

A more reliable means to evaluate the competitiveness of spot electricity markets and recognize the dynamic nature of market power in these markets is to identify when one or more suppliers are "pivotal". A supplier is pivotal when the output of some of its resources is needed to meet demand in the market. A pivotal supplier has the ability to unilaterally raise the spot market prices to very 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.[105] 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.

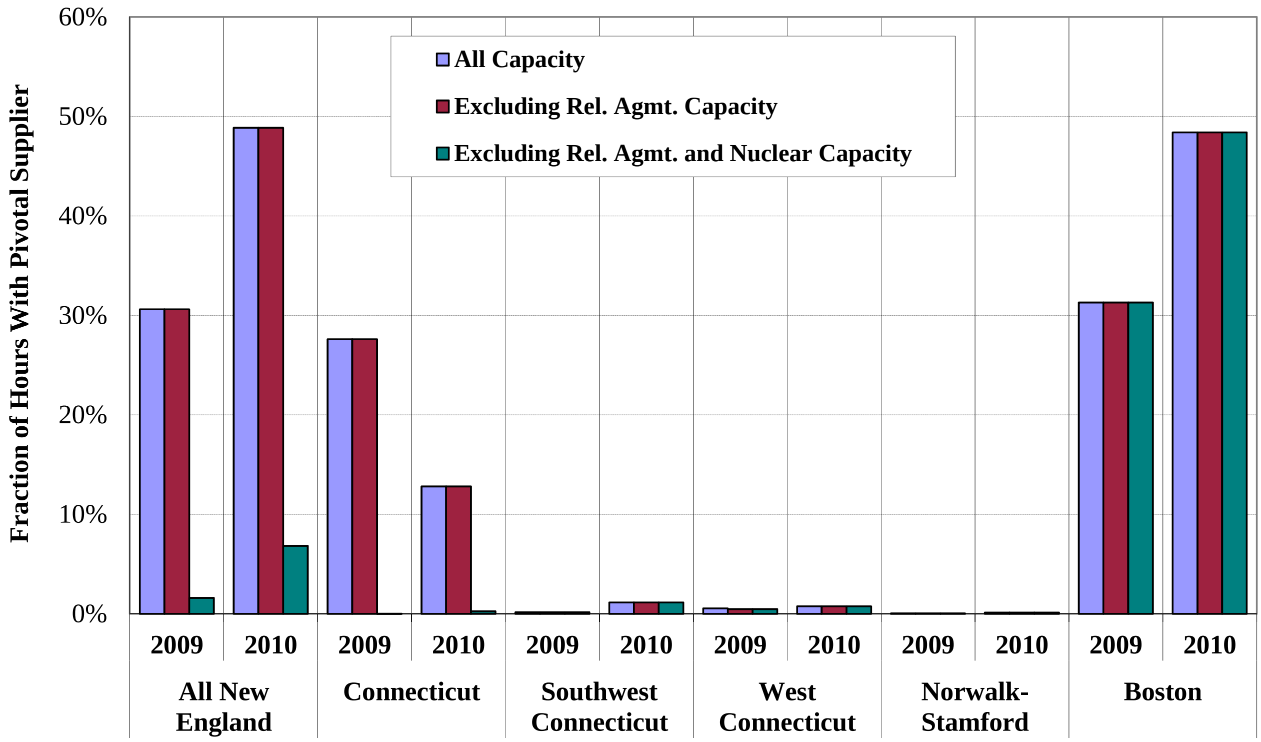
To identify the areas where market power is a potential concern most frequently, Figure 38 shows the portion of hours where at least one supplier was pivotal in each region during 2009 and 2010. The figure also shows the impact of excluding nuclear units and units under reliability

105 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.

agreements from the analysis.[106] As discussed above, owners of nuclear units are less likely to engage in economic or physical withholding of these units. Reliability agreement capacity is excluded because units under reliability agreements have a greatly diminished incentive to exercise market power as discussed above.

Including all categories of capacity, the pivotal supplier analysis raises potential concerns regarding three of the six areas shown in Figure 38. 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.

Figure 38: Frequency of One or More Pivotal Suppliers by Type of Withheld Capacity 2009 – 2010



106 Reliability agreements expired on June 1, 2010. Hence, the category of units under reliability agreements was applicable only the first five months of 2010.

The figure shows that potential local market power concerns were 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 or 2010.

Although Connecticut had a pivotal supplier in 13 percent of hours in 2010 and 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.[107] Therefore, it is appropriate to exclude the nuclear capacity from the pivotal supplier frequency for Connecticut. This leaves very few hours when a supplier was pivotal in Connecticut in the past two years.

For the entirety of New England, because none of the largest three suppliers had resources under a reliability agreement in 2009 and 2010, 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 49 percent to 7 percent of hours in 2010, and from 31 percent to 2 percent of hours in 2009). 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.

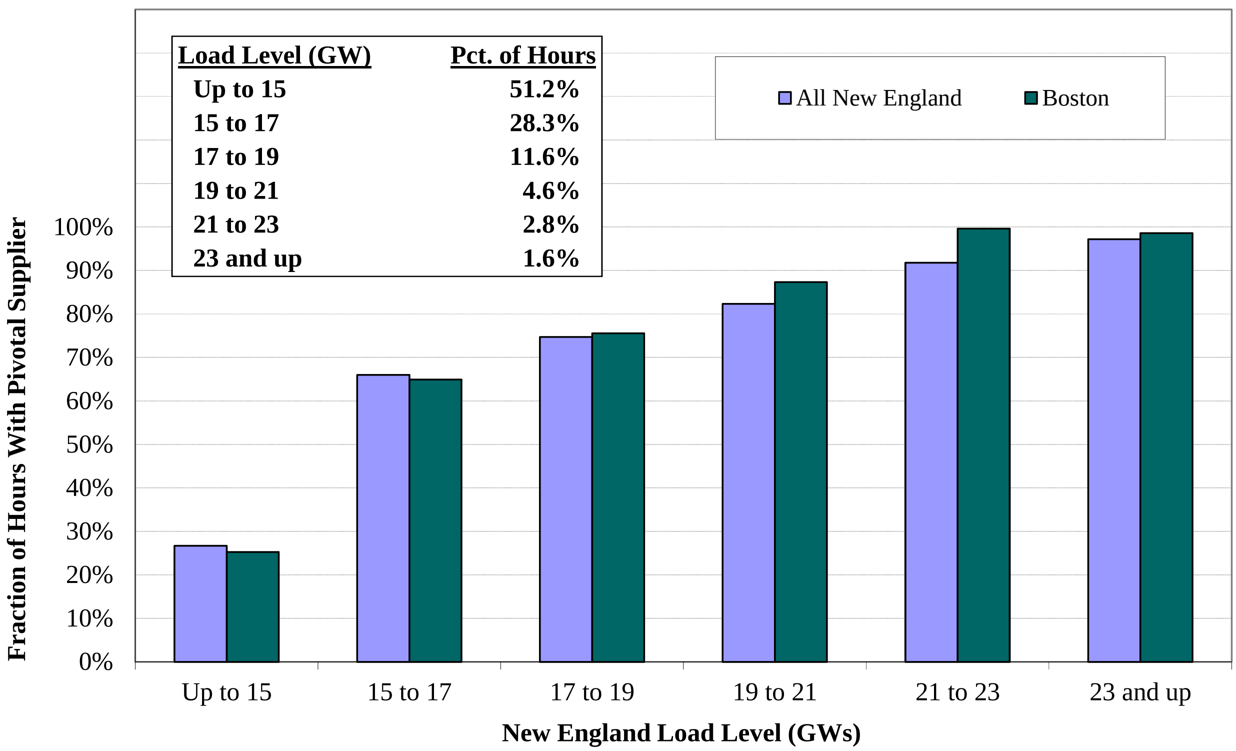
In Boston and all of New England, the pivotal frequency rose notably from the prior year, from 31 percent to 48 percent in Boston, and from 31 percent to 49 percent in all of New England. These increases were primarily driven by the substantial declines in surplus capacity due to the combined effect of higher load levels and reduced reliability commitments, which are discussed in detail in Section VI. These factors greatly reduced the supply margin in these areas in 2010.

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

The pivotal supplier summary in Figure 38 indicates the greatest potential for market power in Boston. A close examination is also warranted for all of New England, while Connecticut raises less concern. Each area had a single supplier that was most likely to have market power. Accordingly, Sections C and D closely examine the behavior of the largest single supplier in each geographic market.

As described 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 39 shows the fraction of hours when a supplier is pivotal at various load levels.

Figure 39: Frequency of One or More Pivotal Suppliers by Load Level Excluding Capacity under Reliability Agreements, 2010



The bars in each load range show the fraction of hours when a supplier was pivotal in All New England and Boston. West Connecticut, Southwest Connecticut, and Norwalk-Stamford are not shown because there were very few instances of a supplier being pivotal during 2010. Connecticut is not shown because the largest pivotal supplier had exclusively nuclear capacity, which is not expected to provide that supplier with an incentive to withhold.

A supplier in Boston was pivotal in at least 65 percent of hours when the load exceeded 15 GW in New England. In all of New England, the largest supplier was pivotal in 66 percent of the hours when load exceeded 15 GW. The pivotal frequency fell below 30 percent in Boston and all of New England during hours when load was below 15 GW in New England.

Based on the pivotal supplier analysis in this subsection, market power was most likely to be a concern in Boston and all of New England when load exceeds 15 GW during 2010. 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 we use 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. They attempt to do this by raising their offer prices so 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. 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 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;
- Connecticut; and
- All of New England.

2. Output Gap in Boston

Boston is a large net-importing region, which can cause transmission interfaces into the region to bind periodically. When this occurs, competition can be limited so it is 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 40 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 40: Average Output Gap by Load Level and Type of Supplier Boston, 2010

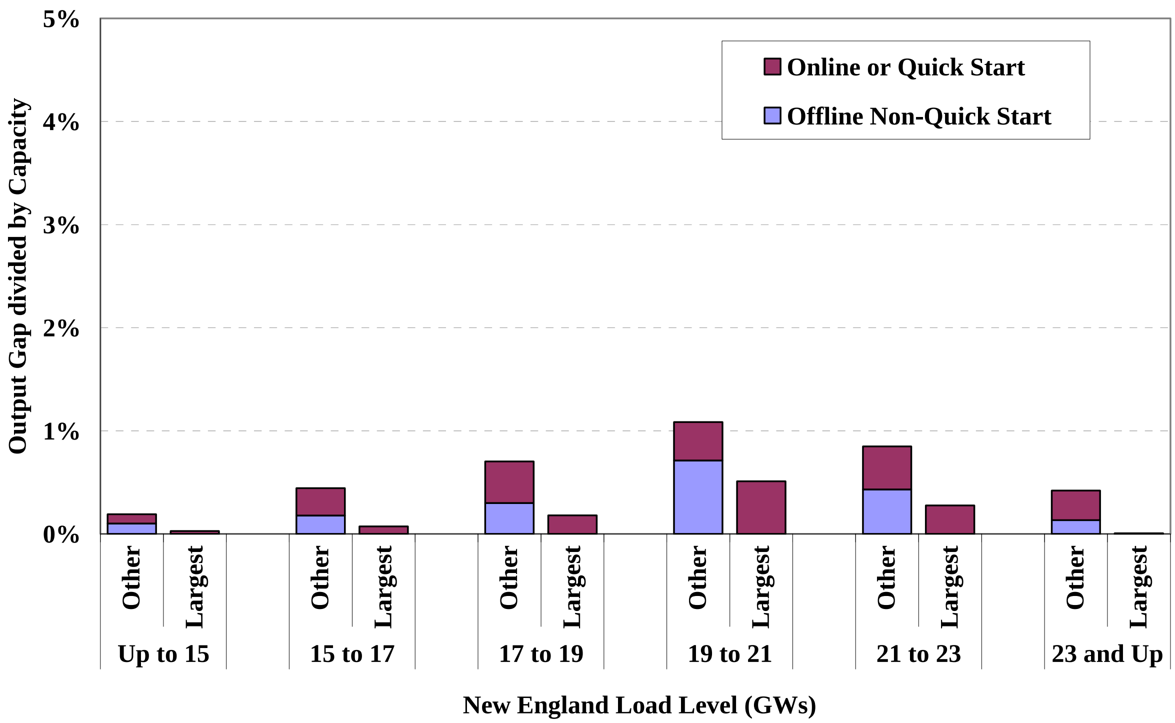
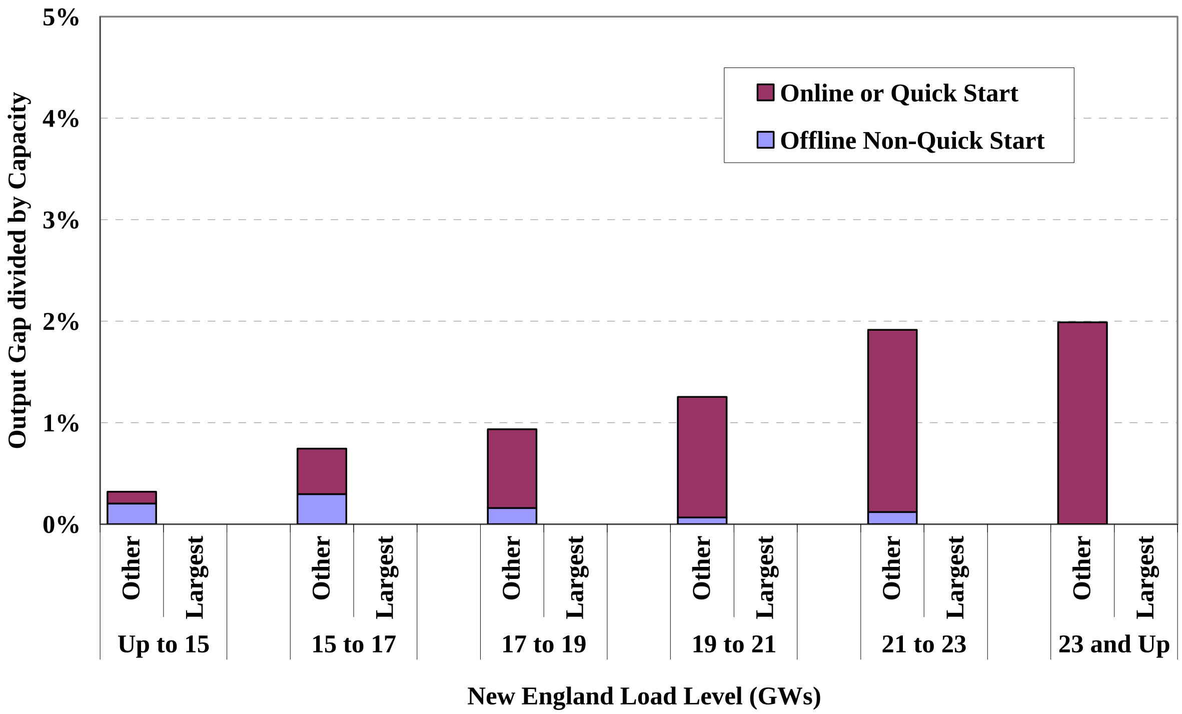


Figure 40 shows that the overall amount of output gap for the largest supplier in Boston was small as a share of its total capacity in 2010, ranging from almost zero when load was below 15 GW to 0.5 percent when load was between 19 and 21 GW. The output gap did not increase at the highest load levels (above 21 GW) and fell close to zero when load exceeded 23 GW. The output gap for the other suppliers was higher than that for the largest supplier, but the highest level was only about one percent at the load range of 19 to 21 GW. Therefore, these results do not raise significant competitive concerns.

3. 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 2010 in Connecticut. Figure 41 shows output gap results for Connecticut by load level. Output gap statistics are shown for the largest supplier compared with all other suppliers in the area.

Figure 41: Average Output Gap by Load Level and Type of Supplier Connecticut, 2010

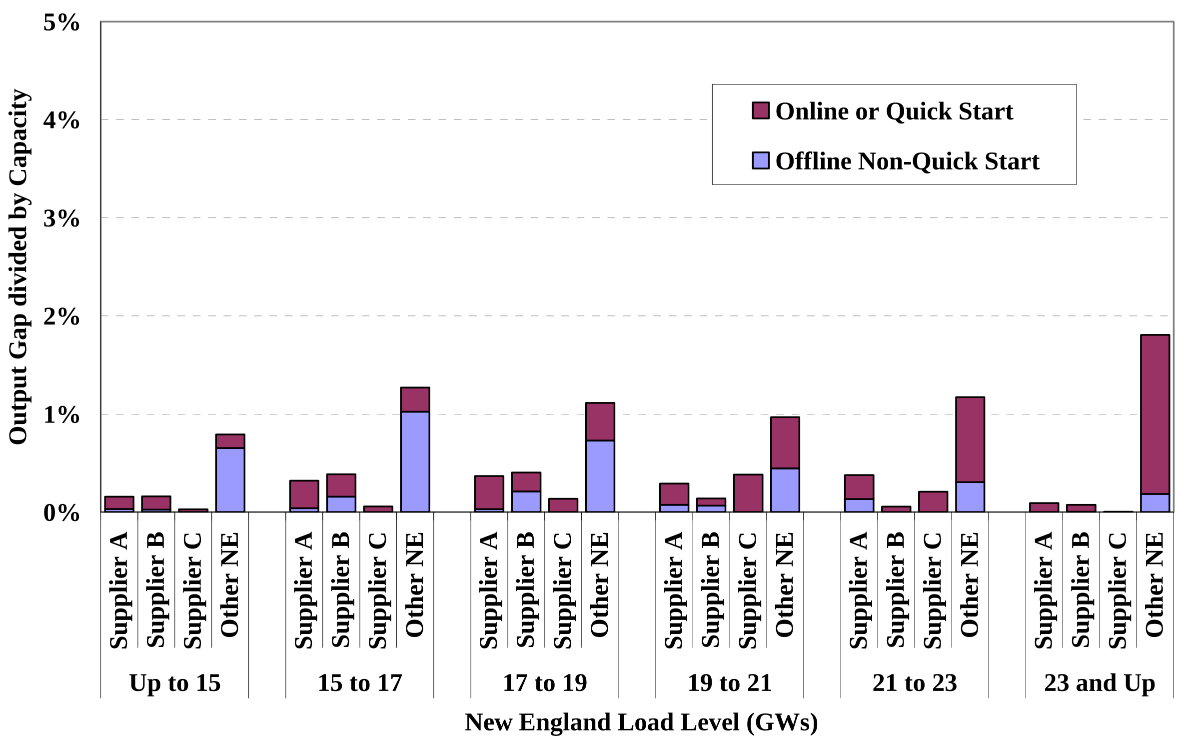


The pivotal supplier analysis indicated that the largest supplier in Connecticut was pivotal in about 13 percent of all hours when all capacity is considered, although the largest supplier owns exclusively nuclear capacity and had no output gap in 2010. Figure 41 also shows that the total output gap of all other suppliers was very low (< 2 percent) relative to the total capacity in Connecticut. Given these amounts, the results do not raise concerns regarding economic withholding in Connecticut.

4. Output Gap in All New England

Figure 42 summarizes output gap results for all of New England by load level for four categories of supply. Supplier A had the largest portfolio in New England and was pivotal in approximately 49 percent of the hours during 2010 (excluding capacity under reliability agreements). Suppliers B and C are the second and third largest suppliers in New England and were each pivotal during 19 percent and 6 percent of the hours. All other suppliers are shown as a group for reference.

Figure 42: Average Output Gap by Load Level and Type of Supplier
All New England, 2010



The figure shows that the region-wide output gap was generally low for each of the four categories of supply. Suppliers A, B, and C exhibited small output gap levels under all load conditions. It is especially notable that the output gap levels for the three largest suppliers were lower than the output gap levels of all other suppliers, which serve as a benchmark for conduct of smaller suppliers that are much less likely to have market power. Hence, the output gap levels for the large suppliers are likely to reflect only measurement error in the metric.

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 2010.

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 three geographic markets examined in the output gap analysis above: Boston, 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 43 shows declarations of forced outages and other non-planned deratings in Boston by load level. Based on the pivotal supplier analysis, the capacity of the largest supplier can be expected to be pivotal in most hours when New England load exceeds 15 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 2010, 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 43: Forced Outages and Deratings by Load Level and Supplier Boston, 2010

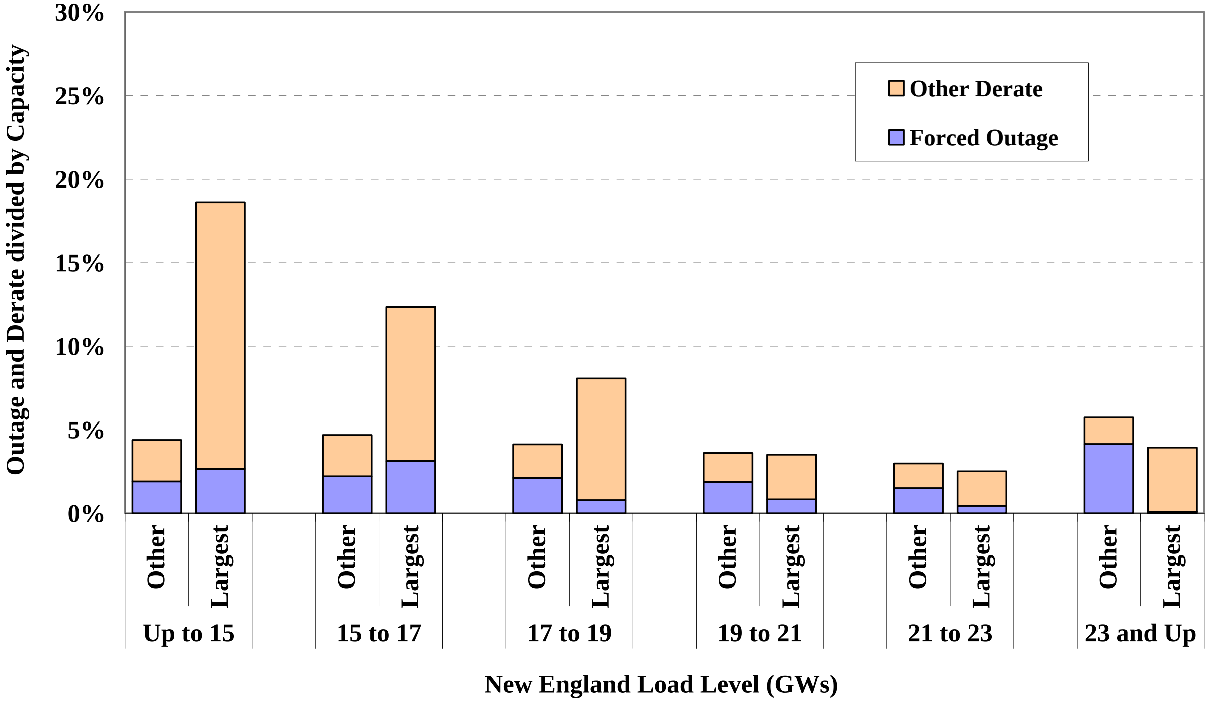


Figure 43 shows a pattern of deratings and outages consistent with expectations in a competitive market. Although levels of outages and deratings for the largest supplier were high at low load levels, they were comparable to other suppliers when load exceeded 19 GW (when withholding is most likely to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

2. Potential Physical Withholding in Connecticut

Figure 44 summarizes declarations of forced outages and other deratings in Connecticut by load level in 2010. The figure shows these statistics for the largest supplier in the area and compares them with statistics for other suppliers.

Figure 44: Forced Outages and Deratings by Load Level and Supplier Connecticut, 2010

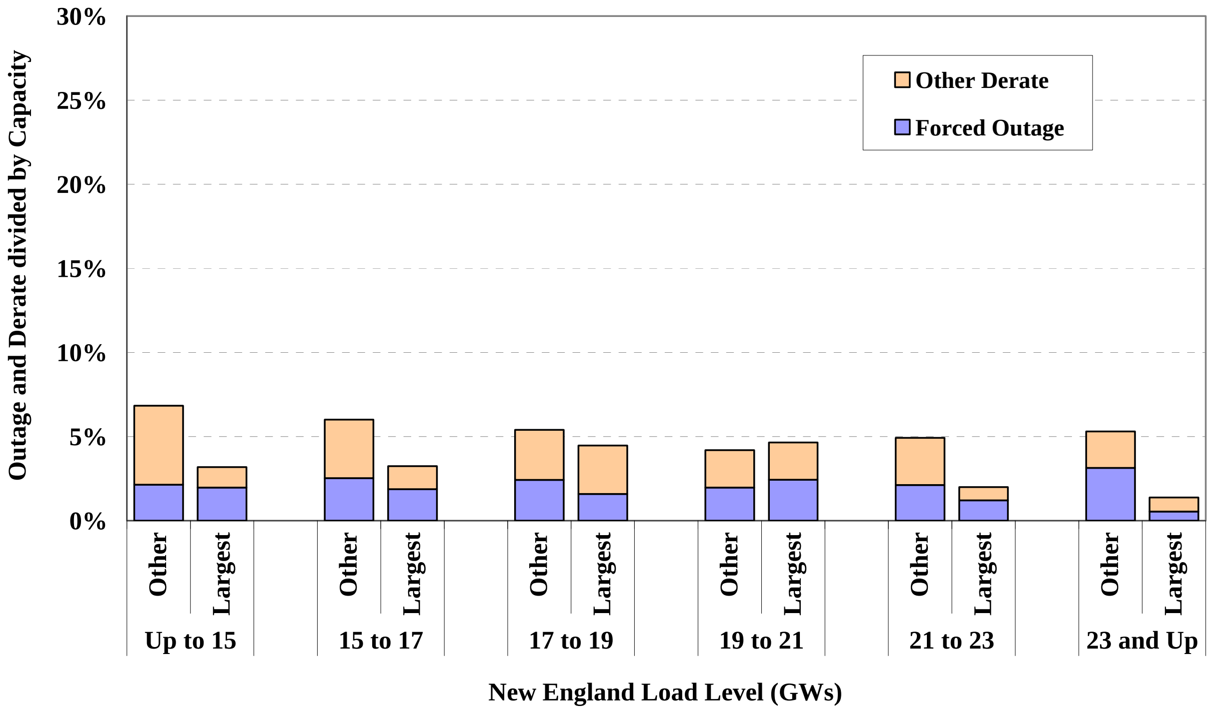
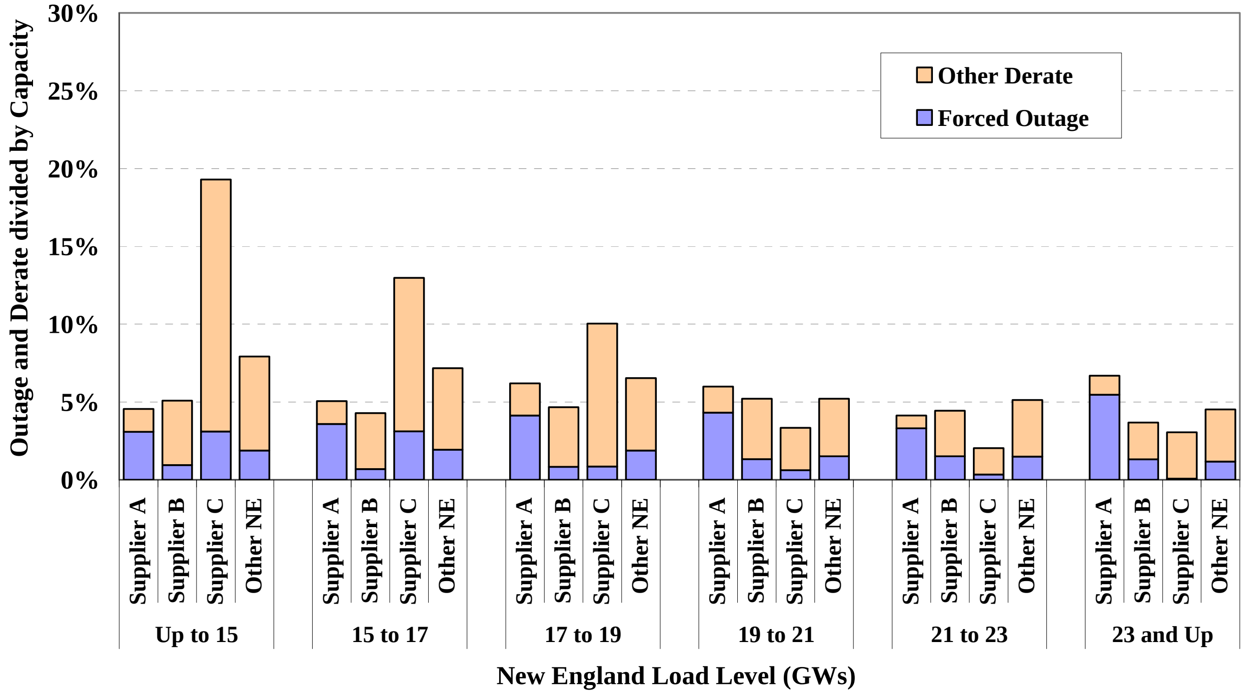


Figure 44 shows that the physical derating and forced outage quantities for the largest supplier and all other suppliers in Connecticut were moderate under all load conditions in 2010 and especially low during high load conditions. Hence, these deratings and outages do not raise concerns about physical withholding in Connecticut.

3. Potential Physical Withholding in All New England

Having analyzed the two major constrained areas in New England, Figure 45 summarizes the physical withholding analysis for all of New England by load level in 2010. 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 49 percent of the hours during 2010 (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 19 percent and 6 percent of the hours. All other suppliers are shown as a group for comparison purposes.

Figure 45: Forced Outages and Deratings by Load Level and Supplier All New England, 2010



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 19 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' 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 2010. 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. Nonetheless, ISO-NE should continue to monitor market outcomes closely for potential economic and physical withholding.