ATTACHMENT 7

2004 STATE OF THE MARKET REPORT
NEW YORK ISO

POTOMAC ECONOMICS, LTD.

Independent Market Advisor
to the New York ISO

July 2005

TABLE OF CONTENTS

Executive Summary ..... **iii**

I. Energy Market Prices and Outcomes **1**
- A. Summary of 2004 Prices and Costs 1
- B. Prices and Price Convergence 10
- C. Price Corrections 16
- D. Forced Outages in 2004 18
- E. Market Power Mitigation 19
- F. Net Revenues Analysis 22

II. Analysis of Energy Bids and Offers **26**
- A. Analysis of Supply Offers 26
- B. Analysis of Load Bidding 31

III. Market Operations **39**
- A. Transmission Congestion 39
- B. Uplift and Out-of-Merit Commitment/Dispatch 50
- C. Market Operations under Shortage Conditions 58

IV. Capacity Market **60**
- A. Background 60
- B. Capacity Market Results in 2004 61

V. External Transactions **66**
- A. Interchange between New York and Other Markets 66
- B. Scheduled Interchange by Hour of Day 71
- C. Conclusions and Recommendations 73

VI. Ancillary Services **74**
- A. Background 74
- B. Offer Patterns 76
- C. Ancillary Services Expenses 78
- D. Regulation Market 80
- E. Changes in Reserve Markets 81

VII. Demand Response Programs **83**

List of Figures

Figure 1: Energy and Natural Gas Prices 2
Figure 2: Price Duration Curve – All Hours 3
Figure 3: Price Duration Curves – Highest 5 Percent of Hours 4
Figure 4: Day-Ahead Energy Prices in 2004 5
Figure 5: Load Duration Curves 6
Figure 6: Load Duration Curves 7
Figure 7: Average All-In Price in 2002 through 2004 8
Figure 8: New York Electricity Market Expenses 9
Figure 9: Day-Ahead and Real-Time Price Convergence at Various Locations 11
Figure 10: Day-Ahead and Real-Time Prices in New York City 13
Figure 11: Average Hour-Ahead and Real-Time Energy Prices 16
Figure 12: Percentage of Real-Time Prices Corrected 17
Figure 13: Equivalent Demand Forced Outage Rates 18
Figure 14: Frequency of Real-Time Constraints and Mitigation 20
Figure 15: Frequency of Day-ahead Constraints and Mitigation 21
Figure 16: Estimated Net Revenue in the Day-Ahead Market 23
Figure 17: Comparisons of Estimated Net Revenue 25
Figure 18: Relationship of Deratings to Actual Load 28
Figure 19: Relationship of Short-Term Deratings to Actual Load 29
Figure 20: Relationship of Output Gap at Mitigation Threshold to Actual Load 30
Figure 21: Relationship of Output Gap at Low Threshold to Actual Load 31
Figure 22: Composition of Day-Ahead Load Schedules as a Proportion of Actual Load 33
Figure 23: Load Scheduled Day-Ahead versus Real-Time Load 34
Figure 24: Hourly Virtual Load and Supply New York City and Long Island 37
Figure 25: Hourly Virtual Load and Supply Outside New York City and Long Island 37
Figure 26: Monthly Congestion Expenses 41
Figure 27: Frequency of Real-Time Congestion on Major Interfaces 43
Figure 28: Value of Real-Time Congestion on Major Interfaces 44
Figure 29: Day-Ahead Congestion Costs and TCC Payments 46
Figure 30: Day-Ahead Congestion Revenue Shortfalls and Real Time Congestion 48
Figure 31: TCC Prices and Day-Ahead Congestion 49
Figure 32: Day-Ahead and Real-Time Uplift Expenses 50
Figure 33: Average Out-of-Merit Dispatch Quantities 52
Figure 34: Average Out-of-Merit Dispatch Quantities 53
Figure 35: Supplemental Resource Evaluation 54
Figure 36: SCUC Local Reliability Pass Commitment 56
Figure 37: Units Most Frequently Committed for Local Reliability and SRE 57
Figure 38: Unforced Capacity Market – Rest of State 62
Figure 39: Unforced Capacity Market – New York City 63
Figure 40: UCAP Sales – Rest of State 64
Figure 41: UCAP Sales – New York City 65
Figure 42: Real Time Prices and Interface Schedules 68
Figure 43: Day Ahead Prices and Interface Schedules 70
Figure 44: Average Net Imports from LMP Markets by Hour of Day 71
Figure 45: Average Net Imports from Canada by Hour of Day 72
Figure 46: Ancillary Services Capability and Offers 77
Figure 47: Ancillary Services Costs 78
Figure 48: Average Clearing Price and Expenses for Regulation Procurement 80

EXECUTIVE SUMMARY

Our assessment of the New York ISO wholesale electricity markets in 2004 indicates that the markets continued to perform competitively with no evidence of significant market power or manipulation by market participants. While the overall state of the market in 2004 was good, there were continuing issues relating to market rules and operations, many of which have been addressed with the implementation of the Real-Time Scheduling ("RTS") system that occurred in February, 2005. At the end of the executive summary, we provide a brief summary of the changes under RTS. In a subsequent report, we will assess the implementation of RTS and evaluate its impact on market efficiency.

In evaluating the NYISO markets in 2004, we address the following areas:

- Energy Market Prices and Outcomes;
- Market Participant Bid and Offer Patterns;
- Market Operations;
- Capacity Market;
- External Transactions Scheduling;
- Ancillary Services; and
- Demand Response Programs

The following subsections provide an overview of the findings of the Report in each of these areas.

A. Energy Market Prices and Outcomes

Summary of Prices Trends in 2004

Energy prices were generally higher in 2003 and 2004 than in the previous two years due to increased fuel prices. Changes in fuel prices are the primary driver of trends in energy prices over extended periods. Even though much of the electricity consumed in New York is generated from hydro, nuclear, and coal-fired generators, natural gas and oil units are usually the marginal generation units that set prices in the market, particularly in peak hours. Therefore, changes in the prices of these fuels will directly impact electricity prices.

The increase in fuel prices began late in 2002, and continued into 2004. Natural gas prices rose an additional 5 percent in 2004 after rising 70 percent in 2003. Distillate oil prices continued their steady rise, increasing 32 percent from 2003 to 2004 after increasing 24 percent from 2002 to 2003. Natural gas prices peaked at very high levels, exceeding $11 per MCF on average in January, when cold weather and limited inventories combined to create tight conditions in the natural gas markets. Distillate prices peaked in October, exceeding $10 per MMBTU, and remained elevated through the end of the year.

The trend toward reduced frequency and severity of electricity price spikes continued in 2004, caused primarily by mild weather conditions during the summer. The mild conditions resulted in lower peak loads on the hottest days. While there were 25 hours in the summer of 2002 when loads exceeded 30 GW, there were only three hours in the summer of 2003 and none in the summer of 2004.

While sharp increases in hourly prices were more frequent in 2002 (when there were six hours with prices over $500/MWh compared to only one hour in 2003 and none in 2004), there were more hours with moderately high prices in 2003 and 2004 due to the increase in fuel prices. In 2004 there were almost 4300 hours with prices above $50/MWh, compared to about 3500 hours in 2003 and less than 1200 hours in 2002.

Prices varied at locations throughout the state in 2004 due to transmission congestion and losses. The primary transmission constraints in New York occur at the following four locations on the system:

- The Central-East interface that separates eastern and western New York;
- The transmission paths connecting the Capital region to the Hudson Valley;
- The transmission interfaces into New York City and the load pockets within the City; and
- The interfaces into Long Island.

Congestion on the transmission paths into and within New York City resulted in average prices in the City that were $11.18 per MWh higher than in the eastern upstate region. The price

difference between eastern and western upstate regions averaged $6.06/MWh, much of which is due to the higher marginal losses in eastern New York.

Total "all-in" prices, which include the costs of energy, ancillary services, capacity, and other costs, increased slightly for all locations in 2004. Higher energy costs outweighed lower ancillary services costs in upstate New York to increase the all-in price slightly. In NYC, the increase in all-in energy prices is attributable to higher capacity costs, which exceeded the slight decline in energy and ancillary services costs.

Day-Ahead and Real-Time Price Convergence

A comparison of the average day-ahead and real-time energy prices in New York outside of New York City and Long Island generally showed a slight day-ahead price premium in 2004. The premium was largest in the Hudson Valley, which exhibited a premium of four percent. A day-ahead premium is generally consistent with expectations because most loads would pay a premium on purchases in the day-ahead market due to: a) the higher price volatility in the real-time market, and b) because Transmission Congestion Contracts ("TCCs") settle on day-ahead prices and quantities. Additionally, generators selling in the day-ahead market are exposed to some risk associated with day-ahead financial commitments. If participants are risk-averse, these factors will generate a price premium in the day-ahead market. This is consistent with historic experience from other markets.

Although the markets generally exhibit a day-ahead premium, the premium decreased in upstate locations from 2003 to 2004. This is an expected result due to more active virtual trading and reduced price volatility associated with the milder peak load conditions. In 2004, New York City and Long Island exhibited a real-time price premium, which was mitigated by the pattern of net virtual supply in upstate New York and net virtual load in New York City and Long Island. These patterns are primarily the result of modeling inconsistencies between the day-ahead and real-time markets related to the transmission limits and assumptions regarding transmission losses. These inconsistencies should be substantially reduced under the RTS. We will evaluate this after the summer of 2005.

Day-ahead to real-time price convergence varied substantially by load pocket within the City during 2004. The 345 kV system (outside most of the load pockets) generally exhibited modest premiums in the day-ahead market, while the major load pockets showed significant price premiums in real time. Price convergence has generally worsened since the implementation of load pocket modeling in 2002. Price convergence in the load pockets could be improved by introducing virtual trading within the New York City load pockets, either at the nodal level or load pocket level.[1] Limiting price-capped load bidding and virtual trading to the zonal level in New York City prevents participants from arbitraging large price differences in specific pockets.

Until 2002, there was a substantial lack of convergence between hour-ahead and real-time prices in New York. The hour-ahead prices were produced by the Balancing Market Evaluation ("BME") model, which was used to schedule external transactions and to establish the hourly dispatch levels for generators that cannot change their output every five minutes. Changes to the market rules and the BME model in 2002 dramatically improved price convergence between the hour-ahead and real-time prices. The improved convergence between hour-ahead and real-time prices improved the scheduling of non-dispatchable resources and imports and the commitment of peaking units. Implementation of the full RTS capability should improve convergence further by making scheduling and commitment decisions on a 15-minute basis.

Price Corrections

All real-time energy markets are subject to some level of price corrections to account for metering errors and other data input problems. The frequency of price corrections was relatively high in 2000, but then decreased steadily until the summer of 2002. The frequency of price corrections increased substantially in June, 2002 as a result of the introduction of changes to the modeling of New York City load pockets. In 2003, there was a spike in the frequency of real-time price corrections, resulting in slight changes to the New York City zonal prices that had been calculated with incorrect weightings. During 2004, corrections occurred at a relatively low level. These results can be attributed in part to the fact that no major enhancements were made to the market software in 2004.

[1] Virtual bids and offers can only be submitted at the zonal level in New York.

Market Power Mitigation

The conduct and impact mitigation framework ensures that mitigation will occur only when market power is exercised to increase prices. The Automated Mitigation Procedure ("AMP") software only runs when energy prices outside the City are greater than $150/MWh, when the market is most vulnerable to market power. Virtual trading, price-sensitive load bids and other factors limit potential market power in the day-ahead market outside the City. The conduct and impact tests were not satisfied in any of the relatively high-priced hours so mitigation was not imposed in 2004. The mild load conditions during 2004 limited the instances when suppliers would have been pivotal in broader areas within New York and, hence, limited the potential for market power abuses.

Inside New York City, the “ConEd” mitigation measures were replaced on May 1, 2004 with the same conduct and impact framework that is used in the State-wide AMP. However, tighter mitigation thresholds are used, reflecting the greater market power concerns in the City. The ConEd measures were triggered whenever there was congestion going into New York City. This approach resulted in mitigation in almost all hours. The conduct and impact framework applies mitigation only to the hours in which offer prices exceed the mitigation thresholds. Day-ahead mitigation has become much less frequent under the conduct and impact framework in New York City. Outside of the load pockets in the City, where the market is more competitive, mitigation occurred in 11 percent of hours while congestion occurred in 31 percent of hours, so mitigation was only invoked about 35 percent of the time when congestion was experienced. Within the load pockets, mitigation was most commonly associated with the constraint into the 138 kV system and into the Astoria West/ Queensbridge/Vernon load pocket.

Mitigation may also be applied in the real-time market for units in certain load pockets within New York City using the NYISO’s conduct and impact approach. The in-city load pocket conduct and impact thresholds are set using a formula that is based on the proportion of congested to non-congested hours experienced over the preceding twelve month period.

Economic Signals Produced in 2004

The economic signals provided by the New York markets can be measured using the net revenue metric. This metric measures the total revenue that a hypothetical new generator would have earned in the New York markets less its variable production costs. In long-run equilibrium, the market should support the entry of new generation by providing average net revenues that are sufficient to finance new entry. This may not be the case in every year since there are random factors that can cause the net revenue to be higher or lower than the equilibrium value (e.g., weather conditions, generator availability, etc.).

While revenues increased in 2003, mild summer weather conditions in 2004 reduced net revenues from the energy market while generators in upstate New York also experienced lower UCAP revenues. In the upstate locations in 2004, a combined-cycle unit would have earned approximately 70 percent of the revenue required to support the investment, while a new combustion turbine would have earned less than one-third of its required net revenue.[2] These results are consistent with expectations for two reasons. First, the mild weather of the last two summers substantially reduced the net revenue from the energy market because it reduced the peak loads and contributed to the lack of shortages in 2003 and 2004. Very high energy prices during transitory periods of shortage are an important component of the long-term economic signal that new resources are needed in a market. Second, there is a substantial surplus of generating capacity in upstate New York, resulting in relatively low UCAP prices and contributing to the lack of shortages. Therefore, the fact that net revenue has been insufficient to support the entry of new generation in upstate New York is not cause for concern.

Capacity margins in New York City have been very close to the minimum requirements, so one would expect the net revenue to be close to or exceed the entry costs of a new unit. Based on the net revenue levels in 2004, a new gas-fired combustion turbine in New York City would recover approximately 85 percent to 99 percent of the net revenue required to support such an investment depending on the location. Under normal weather conditions and, thus, higher energy net revenue over the past two years, the net revenue for a new gas turbine would exceed its entry

[2] The net revenue values for the combined-cycle unit does not consider start-up costs or minimum run times, both of which would tend to reduce a new unit's actual net revenue.

costs in the City. This is also likely true for a new combined-cycle resource, although the entry costs of such resources in the City are not known. These results are confirmed by the fact that most new construction planned in the near-term is occurring in New York City.

B. Analysis of Energy Bids and Offers

In this section of the Report, we analyze the overall patterns of conduct in the New York market, including those that could indicate attempts to exercise market power.

Potential Physical and Economic Withholding

We examined whether there was any correlation of quantities of potential withholding to load levels. The analysis is based in part on the expectation that suppliers in a competitive market should increase bid quantities during higher load periods to sell more power at the higher peak prices. Alternatively, suppliers in markets that are not workably competitive will have the greatest incentive to withhold at peak load levels when the market impact is the largest. Hence, examining how participant conduct changes under different market conditions is an effective means for evaluating the competitive performance of the market.

We first considered potential physical withholding by analyzing generator deratings. A derating occurs when a participant reduces the maximum output available from the plant. This could be for planned outages, long-term forced outages, or short-term forced outages. A derating could be partial (maximum output is reduced, but is greater than zero) or complete (maximum output is zero). We analyze only non-planned outage deratings, eliminating planned outages from our data. The remaining deratings data would then include only long-term and short-term deratings.

We focused on the hours with higher demand because, under a hypothesis of market power, we would expect to find that withholding increases as demand increases. We also limited ourselves to the locations east of the Central-East interface, as this area, which includes two-thirds of the State's load, has limited import capability and is more vulnerable to the exercise of market power. We found that no (statistically) significant relationship existed between deratings and load level in 2004, which would lead us to reject the hypothesis that market power was systematically exercised through physical withholding. Focusing only on short-term deratings,

we found the same results. Deratings are least frequent when load reaches high levels, which is consistent with workable competition.

We also examined the trend in forced outages in the New York markets to ascertain if generators are responding to economic incentives to increase reliability of their units. The Equivalent Forced Outage Rate ("EFOR") is the portion of time a unit is unavailable due to forced outages, expressed as equivalent hours of full forced outage at its maximum net dependable capability. EFOR declined substantially following the implementation of the NYISO markets. This is consistent with the incentives the deregulated markets provide to maximize availability, particularly during high load conditions. EFOR was relatively high in 2000 due to the outage of an Indian Point nuclear unit. After the Indian Point outage, the EFOR has been consistently close to 4 percent – much lower than the outage rates that prevailed prior to the implementation of the NYISO markets.

To evaluate economic withholding, we calculated the hourly "output gap". The output gap is the quantity of generation capacity that is economic at the market clearing price, but either is not running due to the owner's offer price or is setting the LBMP with an offer price substantially above competitive levels (excluding capacity scheduled to provide ancillary services). This withholding can be accomplished through high start-up cost offers, high minimum generation offers, and/or high incremental energy offers.

To determine whether an offer is above competitive levels, we use reference values based on the past offers of the participant during competitive periods. We conduct the analysis with thresholds matching the mitigation thresholds ($100/MWh or 300 percent, whichever is lower) and a lower threshold ($50/MWh or 100 percent, whichever is lower).

Like our analysis of deratings, the results would support a hypothesis of withholding if the output gap increases as load increases. We focused our analysis on Eastern New York where market power is most likely. We found that the output gap decreases to extremely low levels under the highest load conditions. This is an important result because prices are most vulnerable to market power under peak load conditions. These results indicate that economic withholding was not a significant concern in 2004.

Analysis of Load Bidding

In addition to physical and economic withholding, buyer behavior can strategically influence energy prices. Therefore, evaluating whether load bidding is consistent with workable competition is an important focus of market monitoring. Load can be purchased in one of the following four ways:

- Physical Bilateral Schedules – These allow participants to settle transmission charges (i.e., congestion and losses) with the ISO and to settle the energy portion of any underlying contract privately between the parties.

- Day-Ahead Fixed Load – This represents load scheduled in the day-ahead market for receipt at a specific bus regardless of the day-ahead price.

- Price-Capped Load Bidding – This is a price-sensitive load bid into the day-ahead market by a Load Serving Entity ("LSE"). Price-capped load bidding is only allowed at the zonal level while fixed load bidding is allowed at the bus level.

- Net Virtual Purchases – This quantity is equal to the virtual load purchases minus the virtual supply sales. Like price-capped load bidding, virtual purchases are allowed only at the zonal level.

Our analysis indicates New York City and Long Island tend to over-schedule load day-ahead. However, this pattern diminishes slightly in the highest load hours. Load scheduled day-ahead in eastern upstate New York is more variable and is usually substantially under-scheduled. This under-scheduling decreases with increases in load. In Western New York, the data reveals that day-ahead load is under-scheduled on average, and that this under scheduling becomes more acute as load rises.

These results are consistent with the differences between the day-ahead and real-time prices. Generally, real-time prices are lower than day-ahead prices in upstate New York, whereas the opposite is true in New York City and Long Island. As discussed in this report, these pricing and scheduling patterns are primarily the result of modeling inconsistencies between the day-ahead and real-time markets. The market will respond to these inconsistencies by adjusting the

purchases and sales in the day-ahead market. In this case, that arbitrage improves price convergence, but results in over-scheduling within New York City and under-scheduling outside of New York City.

C. Market Operations

Aside from operating the spot markets, a primary role of the ISO is to ensure safe and reliable grid operation. Many of the ISO's operating functions in this regard can have a substantial impact on market outcomes, especially during peak demand conditions. Reliability requires that operators carry out all of these functions, but they should be done in a way that promotes efficient market pricing and behavior. This section evaluates several operating functions and examines how they impact market outcomes.

Transmission Congestion

Congestion can arise in both the day-ahead and real-time markets when transmission capability is not sufficient to accommodate a least-cost dispatch of generation resources. When congestion arises, this will result in higher spot prices at these "constrained locations" than would occur in the absence of congestion.

The NYISO applies congestion charges to day-ahead market transactions by modeling anticipated congestion. These charges are based on the difference between day-ahead spot prices at different locations (the price at the sink less the price at the source). Congestion revenues are collected from participants, which include: a) the difference between the total payments by loads and the payments to generators and net imports (excluding losses), and b) the congestion costs collected from physical bilateral schedules. In an LMP system, this revenue will be equal to the marginal value of the transmission capacity times the amount of power flowing across the constrained interface. In the real-time market, only interface flows that are not scheduled in the day-ahead market are assessed balancing congestion charges (or credits).

This report finds that congestion charges grew from $310 million in 2001 to $629 million in 2004 due to (i) the implementation of load pocket modeling in New York City which allowed market-based congestion management, and (ii) higher fuel prices which tend to proportionately increase regional price differences associated with congestion. It is important to recognize that

these costs do not represent the net benefits of eliminating all congestion in New York, which has been estimated to be less than $100 million.

Our report includes an analysis that summarizes congestion levels on major interfaces in New York. Congestion is most frequent into the New York City load pockets, while congestion is considerably less frequent on major interfaces in upstate New York, such as the Central-East interface. Our analysis finds that the value of the major upstate transmission interfaces was approximately $70 million, while the value of the most significant downstate interfaces totaled $400 million in 2004.[3] While more analysis would be necessary to determine where transmission investment would be most profitable, this analysis suggests that the existing transmission is most valuable in New York City and Long Island.

In a well-functioning market, the level of transmission congestion should generally converge between the TCC market, day-ahead market, and real-time market. Three aspects of convergence are examined in this report. First, day-ahead congestion revenue shortfalls can occur when the revenues collected by the NYISO from congestion in the day-ahead market are less than the payments by the NYISO to the holders of TCCs. Second, balancing congestion revenue shortfalls occur when congestion revenues collected from buyers in the real-time market are not sufficient to cover congestion payments by the NYISO to sellers. Third, the prices paid for TCCs in the auction should be comparable to congestion prices in the day-ahead market that determine payment to TCC holders. The results of our analysis in these areas are as follows:

- Day-Ahead Congestion Revenue Shortfalls: The NYISO experienced substantial day-ahead congestion revenue shortfalls for several years until mid-2004. These shortfalls generally occur when the quantity of TCCs sold in the auctions exceeds the transmission flows in the day-ahead market. Ideally, the quantity of TCCs would closely match the physical capability of the transmission system. A large share of the shortfall was due to excess TCCs sold into New York City. These excess TCCs were repurchased by the NYISO in July 2004. In addition, the NYISO has implemented provisions to allocate shortfall costs to transmission operators ("TOs") with outages that cause the shortfall, and by allowing TOs to reserve transmission capacity by converting up to 5 percent of

[3] These values are not consistent with the total congestion costs reported elsewhere in the report because these values are based only on the real-time prices and the analysis did not consider transmission constraints that restrict flows within Long Island or western New York. It also does not quantify the value of flows that the NYISO schedules from western New York through PJM to New York City.

transmission capacity into six-month TCCs, which would not be available in TCC Auctions.

- Balancing Congestion Costs: the NYISO has also experienced balancing congestion revenue shortfalls for several years. If the day-ahead and real-time modeling assumptions are consistent, these costs should be close to zero. Due to modeling inconsistencies, NYISO effectively oversells transmission capability in the day-ahead market, and is therefore compelled to buy back the over-sold amount, and must uplift the associated cost to all market participants. These modeling inconsistencies may be substantially addressed through the implementation of the RTS because the RTS software platform is very similar to the day-ahead Security-Constrained Unit Commitment ("SCUC") model.

- TCC Price – Congestion Cost Convergence: In a well-functioning market, the price for the TCC should reflect a reasonable expectation of day-ahead congestion. The auction prices from the auction of 6-month TCCs during the summer capability period for 2004 resulted in a relatively accurate reflection of the value day-ahead congestion, with TCC prices slightly exceeding actual congestion. This premium is not unexpected, since a TCC is essentially a hedge against unexpected congestion. Hence, risk adverse participants would be expected to pay a premium to avoid this risk. Furthermore, actual congestion was likely lower than expected due to the mild weather conditions in 2004.

Uplift Costs

Uplift costs are incurred when market revenues are not sufficient to satisfy all market obligations while ensuring that all suppliers recover sufficient revenue to cover their full as-offered costs. The uplift costs can be divided into three areas: costs incurred in the real-time market associated with maintaining local reliability, other uplifts costs incurred in the real-time market, and uplift costs incurred in the day-ahead market. Our findings regarding these classes of uplift costs are summarized as follows.

- Real-Time Local Reliability Uplift: Uplift costs for real-time reliability fell sharply after 2002 due to the introduction of load pocket modeling in June 2002. Reduced payments for out-of-merit generation to manage congestion in the New York City load pockets are now reflected in the congestion component of the spot market price. Previously, the re-dispatch costs to manage load pocket congestion had been collected through uplift.

- Other Real-Time Uplift: Changes to the BME in 2002 to more accurately schedule units and imports for the real-time market helped reduce other uplift associated with the real-time market. However, these reductions in uplift have been partly offset by higher fuel costs.

- Day-Ahead Market Uplift: Day-ahead market uplift has tripled since 2002. This is uplift paid to units committed by SCUC, mostly in the local reliability pass of SCUC. These supplemental commitments by SCUC have a tendency to decrease day-ahead prices. As a result of lower prices, some day-ahead market uplift is paid to generators committed day-ahead. The increase in day-ahead market uplift is due to: increased quantities of generation committed in the local reliability pass of the SCUC and the increase in fuel prices.

Out-of-Merit Commitment and Dispatch

A resource is *dispatched* Out-of-Merit ("OOM") when it is dispatched by the ISO, even though its energy offer exceeds the price at its location. This can be caused by the physical parameters of the unit (e.g., minimum run-time that requires the unit to run after it has become uneconomic) or by operator action. OOM actions by NYISO operators are generally taken to ensure the reliability of the system. OOM dispatch in real-time can also be used to manage network constraints that are not included in the model. OOM actions tend to depress spot market prices, particularly during peak demand conditions when prices are most sensitive, and mask congestion. OOM dispatch quantities fell by more than two-thirds from 2002 to 2003, primarily due to the introduction of load pocket modeling in New York City and improvements in the commitment of gas turbines in the real-time market. In 2004, OOM dispatch quantities were generally very low across the state. The average quantity of OOM energy dispatched in 2004 was less than 65 MW.

Out-of-Merit *commitment* occurs when a unit is instructed to start-up, even though it has not been selected through the NYISO markets. OOM commitments include SRE actions, a process by which the ISO commits additional resources after the day-ahead market closes in order to meet reliability requirements, and local reliability commitments made by the SCUC model. Both OOM dispatch and commitment actions result in supplemental payments to owners of the OOM units that are recovered through uplift charges. The trends we identify in the OOM commitment include:

- Improvements in day-ahead modeling and commitment have reduced the quantity of SRE actions outside of New York City since 2001.
- The average quantity of capacity committed through SRE in New York City has increased three-fold since 2002. This increase is partly due to nitrous oxides (NOx)

emission limits that require certain baseload units to operate in order for gas turbines to operate.

- Capacity committed in the day-ahead market for local reliability tends to reduce prices from levels that would result from a purely economic dispatch; and can increase uplift incurred to make guarantee payments. The average capacity committed for local reliability was approximately 440 MW in 2004, which is a 50 percent increase from 2003.

Virtually all of the local reliability commitments made by SCUC involved three units in New York City. These units are generally scheduled at their minimum generation level. It would be more efficient for these units to be committed within the economic pass of SCUC because it may cause SCUC to not commit units in other locations, which would reduce uplift and improve energy prices. The means to do this is discussed in this report.

Market Operations in Shortage Conditions

Two market reforms were implemented prior to the summer of 2003 to improve the efficiency of the energy pricing during shortage conditions. First, Reserve Shortage Pricing ("scarcity pricing") became effective in June 2003. When the system is in shortage (that is, when available capacity is not sufficient to meet both energy and reserve requirements), the ISO meets the system's energy demands by foregoing a portion of its required reserves. Because the ISO will pay a supplier up to the offer cap of $1000 for energy in order to hold 10-minute reserves, the scarcity pricing provisions set the LBMP at $1000/MWh in New York City when a 10-minute reserve shortage occurs.

Second, the pricing provisions were modified to allow demand response resources to set energy prices. The NYISO can call on demand-side resources -- Special Case Resources ("SCRs") and Emergency Demand-Response Program ("EDRP") resources – to reduce their consumption and be paid up to $500 for these load reductions. When these reductions are needed to avoid a shortage, they will set the energy price. Due to the relatively mild weather in the summer of 2004 and increased imports from New England, there were no shortages in 2004. Hence, these pricing provisions were not triggered.

The scarcity pricing provisions have been replaced by reserve demand curves as part of RTS, which more fully and efficiently reflect the allocation of resources between production of energy

and reserves that occur under shortage conditions. The reserve demand curves have been designed to reflect the current operating requirements and reflect the implicit value of the operating reserves based largely on the $1000 bid cap. The reserve demand curves are included in both the day-ahead and real-time market models, ensuring that the day-ahead commitment, hour-ahead scheduling, and real-time dispatch are all consistent.

D. Capacity Market

The capacity market is intended to provide efficient economic signals for capital investment and retirement decisions for generating capacity. To improve the performance of the capacity market, a demand curve was implemented in May 2003. The capacity demand curve stabilized the capacity prices and substantially improved the consistency of prices in the strip, monthly, and deficiency auctions. The capacity demand curve also caused a larger share of the capacity to be sold in the deficiency auction, when previously the small volumes purchased had contributed to erratic prices in this auction. The increase in spot procurements corresponds to a reduction in self-schedules. This trend toward increased purchases in the spot market reversed in 2004 when the capacity purchased in this market fell to roughly 20 percent of all capacity purchased.

Overall, the capacity prices in the "rest-of-state" area were not substantially higher following the implementation of the demand curve. A summary of the prices in the rest-of-state capacity market after the introduction of the demand curves include:

- Compared to the prior year, capacity prices in the strip auction decreased slightly in the summer 2003 and increased slightly in the winter 2003-2004;
- Capacity prices in the summer 2004 strip auction remained stable, though shorter term prices declined; and
- Prices in the winter of 2004-2005 declined significantly to pre-demand curve levels.

In New York State, the capacity demand curve contributed to higher purchases in the rest-of-state. The capacity demand curve resulted in additional purchases in the summer 2003 of 2200 to 2500 MW. A few hundred MW of additional capacity were purchased in the summer of 2004, due in part to the start-up of the Athens plant in May 2004. In the winter, the demand curve resulted in slightly higher purchases ranging from 2500 to 3300 MW. The additional purchases in the winter are due to the higher unit ratings during the winter months that increase available

UCAP supplies. A substantial share of the additional UCAP in all seasons came from sources external to the NYISO after the implementation of the capacity demand curve.

Capacity purchased in New York City increased significantly in 2003 and 2004. The increased UCAP purchases are primarily due to increased requirements in the City rather than the demand curve. Virtually all of the capacity in the City was sold, i.e., much less went unsold than in the rest-of-state. Finally, it is important to note that revenues from the capacity market play a critical role in the conclusion that the economic signals in New York City would support new investment. This is an important result because New York City capacity levels are close to the minimum required to maintain reliability.

E. External Transactions

The performance of the wholesale electricity markets depends not only on the efficient utilization of the internal resources, but also the efficient utilization of the transmission interfaces between New York and other areas. Absent transmission constraints, trading should occur between neighboring markets to cause prices to converge.

Based on our analysis in this report, the real-time markets continue not to be efficiently arbitraged by participants. The dispersion in prices during unconstrained hours is considerable. In a significant number of hours for each interface, power is scheduled from the high-priced market to the lower-priced market. These results are similar to results presented in prior years. Several factors prevent real-time prices from being fully arbitraged between New York and adjacent regions.

- Market participants do not operate with perfect foresight of future market conditions at the time that transaction bids must be submitted. Without explicit coordination between the markets by the ISOs, complete arbitrage will not be possible.

- Second, differences in scheduling procedures and timing in the markets serve as barriers to full arbitrage.

- Third, there are substantial transmission fees and other transaction costs associated with scheduling imports and exports that diminish the returns from arbitrage. Risks associated with curtailment and congestion will reduce participants' incentives to engage in external transactions at small price differences.

- A significant portion of imports and exports reflect long-term bilateral agreements (rather than arbitrage of hourly prices) which tend to be insensitive to real-time prices and contribute to the price divergence.

The introduction of the LMP markets in New England in March 2003 was expected to improve arbitrage between New York and New England, but transactions costs and other factors continue to hinder efficient arbitrage. In 2005, export fees between New York and New England were eliminated, which will help improve the arbitrage of the adjacent markets. However, exports from New England scheduled after the day-ahead market continue to be allocated uplift charges for certain types of supplemental commitment. These charges can be significant.

We continue to encourage New York and New England to develop and implement new scheduling procedures, such as "intrahour transaction scheduling". Intrahour transaction scheduling is a process that would allow the physical interchange to be adjusted within an hour when prices diverge at the interface between the two markets. These adjustments would ensure that the interchange levels are efficient, eliminating the price distortions and other inefficiencies caused by poor market arbitrage. This will lead to less volatility and more predictability in the New York to New England prices. Likewise, we recommend that the NYISO work with PJM to eliminate export fees and improve the scheduling procedures.

F. Ancillary Services

Ancillary Services Costs

Ancillary services costs declined slightly as a percentage of total market expenses from close to 2.5 percent in 2002 to roughly 1.5 percent in 2004. Over this time-frame, total ancillary services expenses decreased by $15 million to a total of approximately $94 million in 2004. Ancillary services costs peaked in 2003 at a total of almost $130 million. Decreased expenditures for ancillary services were primarily due to reductions in the cost of 10-minute total reserves, 30-minute reserves, and regulation. Prices of 10-minute non-spinning reserves declined even after the removal of the $2.52 bid cap.

The decline in reserve costs since 2001 can be attributed to three market design changes. The reserve-sharing agreement with ISO-NE permitted a reduction in the ten-minute reserve requirement for the East (from 1200 MW to 1000 MW), locational ancillary services prices

limited the impact of reserve shortages in constrained areas on state-wide reserve prices, and changes in the BME model to recognize latent 30-minute reserves prevents the BME model from setting irrationally high prices for reserves when plenty of 30-minute capability is available.

Efficient pricing of reserves with the implementation of RTS will likely increase total reserve costs, despite cost reductions due to other RTS improvements. This is an important feature of the RTS operating reserves markets because it provides the necessary economic signals to attract and retain resources that are primarily needed to meet the NYISO's reserve requirements, such as gas turbines.

Ancillary Services Offer Patterns

Our findings in previous analyses in New York have indicated that a substantial portion of the capability of certain services is not offered in the day-ahead ancillary services markets, particularly for 30-minute reserves and regulation. With the exception of the 10-minute non-synchronous resources, a substantial portion of the capability of all other ancillary services was not offered in the day-ahead markets. The average quantity of regulation offered to the market is approximately one-half of the total capability, while less than one-third of the 30-minute operating reserves capability was offered on average. Generally, this is not a significant concern given the excess reserve and regulation capability that is available. However, these offer patterns can result in very high prices under peak load conditions since these markets are jointly optimized and the same resources are offered in multiple markets. Prior recommendations to increase the portion of the capability offered have now been implemented as part of the RTS system, which we will evaluate following the summer 2005.

Regulation Market

The regulation market is the only market-based ancillary service that is not a type of operating reserve. Regulation prices have increased from 2002 levels, primarily due to modeling changes in SCUC and BME in May 2002 to recognize that units' minimum generation level may limit the range in which a unit can regulate down. This reduced the supply available on some units, particularly during off-peak periods. This constraint on assigning regulation no longer exists with the implementation of RTS.

The second factor that contributed to the rise in regulation prices is higher fuel prices that increase opportunity costs to provide regulation and raise regulation prices, though the impact is much smaller than the effect of higher fuel costs on the energy market. Nonetheless, regulation costs still remain a relatively low portion of the total electricity market expenses for the NYISO (slightly more than 1 percent). Since regulation capacity far exceeds the demand and is controlled by a diverse set of suppliers, market power is not a significant concern.

Changes in Reserve Markets

The implementation of RTS in 2005 will lead to major changes in the markets for reserves and regulation. Under the multi-settlement system, real time ancillary services schedules will be settled against the day-ahead schedules. Since suppliers are liable for the real-time cost of reserves that they schedule day ahead, they will have an incentive to be available in real time and to perform when called. Reserve market clearing prices will be set on a locational basis using the shadow prices of the reserve constraints in both the day-ahead and real-time markets. Both day-ahead and real-time clearing prices of ancillary services will cover the lost opportunity cost of the marginal supplier. This is intended to give efficient incentives to the lowest-cost reserve providers to provide reserves rather than energy, and eliminate the need for separate lost opportunity cost payments previously recovered through uplift charges.

In RTS, the prior reserve shortage scarcity pricing provisions were be superseded by the reserve demand curve. These demand curves establish an economic value for reserves that will be reflected in energy prices at times when the energy market must bid scarce resources away from the reserve markets. Because reserves should generally be substituted to maintain the highest quality reserve, the total value of a specific reserve type will incorporate the demand curve values of lower quality reserves. The demand curve values have been set at levels that are consistent with the actions normally taken by the NYISO operators in reserve shortage conditions.

G. Demand Response

The New York ISO has some of the most effective demand response programs in the country. There are currently three demand response programs in New York State:

- Special Case Resources ("SCR") – These are loads that must curtail within two hours. SCR participants are paid the higher of a strike price that they bid (up to $500/MWh) or the real-time clearing price.

- Emergency Demand Response Program (EDRP) – Loads that curtail on two hours notice on a voluntary basis. EDRP resources are paid the higher of $500/MWh or the real-time clearing price.

- Day-Ahead Demand Response Program ("DADRP") – This program schedules physical demand reductions for the following day, allowing resources with curtailable load to offer into the day ahead market as any supply resource.

The EDRP and SCR programs are among the most effective of their kind in achieving actual load reductions during peak conditions. The total registered quantity of more than 1700 MW is much larger than comparable programs in other ISOs. In 2004, the quantity of SCR/ICAP subscribers that sold capacity was 175 MW in NYC; 98 MW in Long Island; and 707 MW in upstate New York. The state total has increased 30 percent from 2003. This success is the result of the pricing incentives that induce a high-level of participation and contribute to efficient pricing in time of shortage.

During times when EDRP and SCR are the marginal sources of supply in the market that allow the system to satisfy its reserve requirements, the LBMP typically will be set at $500/MWh. This price in a range that is consistent with the marginal value of reserves to the system. Hence, these payments and the associated pricing provisions contribute to efficient pricing during shortage (or near-shortage) conditions. However, EDRP and SCRs were not utilized in 2004 due to mild load conditions and good resource availability.

The DADRP has not resulted in substantial quantities of real-time demand reductions. There were 2818 hours with day-ahead demand response bids. The average quantity bid was approximately 2 MW per hour, and the average quantity scheduled was less than half a megawatt. There were 222 hours when day-ahead demand response bids amounted to 10 MW or more, with a high of 17 MW, and these bids were accepted in 132 hours. The hours with these large bids primarily occurred around holidays such as New Year’s Day, Thanksgiving, and Christmas week. The low participation may be due to the alternatives available for demand to bid in the markets (virtual trading and price-capped load bidding).

H. Real-Time Scheduling

The NYISO began implementation of the RTS market enhancements on February 1, 2005. The RTS system uses a common computing platform, algorithms, and network models for both the real-time commitment and real-time dispatch functions, and is comprised of three major components:

- Real-Time Commitment ("RTC") model replaces the BME software. RTC co-optimizes energy, reserves and regulation, and commits resources as necessary to meet the demands of the next hour. RTC runs and posts results every 15 minutes, instead of every hour, and makes commitment decisions that are optimized over a 2 ½ hour period. RTC issues binding commitments to 10-minute and 30-minute gas turbines. It also determines transaction schedules and the dispatch level for off-dispatch resources for the upcoming hour at the top of each hour. However, it has the capability to schedule external transactions and off-dispatch generation on a 15-minute basis.

- Real-Time Dispatch ("RTD") replaces the Security Constrained Dispatch ("SCD") software to dispatch the system. RTD issues a 5-minute basepoint, co-optimizing energy, reserves, and regulation for the forecasted conditions up to 60 minutes ahead. RTD recognizes the transaction schedules, self-committed unit schedules, and units committed by RTC in making dispatch decisions.

- Real-Time Dispatch - Corrective Action Mode ("RTD-CAM") is a tool that the NYISO system operators can run on-demand to address abnormal or unexpected system conditions. RTD-CAM can produce a new set of basepoints, and/or commit 10-minute resources on demand. It also includes innovative algorithms that support reserve pick-ups, and the return to normal system operations.

The RTS was a vehicle for introducing a number of market enhancements, including:

- Generators may submit hourly start-up costs, specified as a discrete dollar cost or as a function of elapsed time since the most recent shutdown.

- Generators may submit three-part bids in real-time for purposes of intraday commitments (start-up costs, minimum generation costs, and incremental energy costs);

- A second settlement in real-time was introduced for the reserves and regulation markets, providing better incentives for these services to be provided by the lowest cost suppliers in real time, regardless of the day-ahead schedules;

- Reserve and regulation market-clearing prices set based upon the marginal system cost of providing the service, including marginal lost opportunity costs;

- Demand curves for reserves and regulation services are included in the Day-Ahead and Real-Time Markets to provide consistent prices (and scheduling results) in the event that the desired reserves are unavailable or more costly to schedule than they are worth.

The last element of the RTS system is particularly important because it ensures efficient energy and operating reserve prices during periods of shortage. When operating reserves are sacrificed to meet energy demands, the value of the foregone reserves will be reflected in the energy and reserve prices. This replaces the prior scarcity pricing provisions that had been utilized prior to the introduction of RTS.

Because the RTS system was implemented in 2005, it is not analyzed or evaluated in this report. However, we will be reviewing the early performance of the RTS system in a report to be performed after the summer of 2005.

I. ENERGY MARKET PRICES AND OUTCOMES

In anticipation of significant changes with the implementation of the new Real-Time Market software and systems in 2005, modifications to the markets in 2004 were limited to improvements that could not be reasonably postponed. While the NYISO continued to make minor refinements to its market rules and procedures over the last year, there were no changes to the basic structure of the multi-settlement energy markets that are the central feature of the New York electricity markets.

The multi-settlement system consists of a financially-binding day-ahead market and a real-time market. Through these markets, the NYISO commits generating resources, dispatches generation, procures ancillary services, schedules external transactions, and sets market-clearing prices based on supply offers and demand bids. The day-ahead and real-time markets were augmented with the hour-ahead BME scheduling process in 2004. The main functions of the BME model was to commit 30-minute gas turbines, establish dispatch levels for units that only receive hourly dispatch signals (i.e., off-dispatch units), and schedule external transactions. The BME schedules are important because the NYISO guarantees that any offer taken through the BME will recover its as-bid costs and because the BME schedules can substantially affect the market outcomes in the real-time market. This section of the report provides an overview of the market results in 2004 and evaluates the performance of these markets. This evaluation includes an assessment of the long-term economic signals provided by the New York markets that govern new investment and retirement decisions.

A. Summary of 2004 Prices and Costs

We begin in this sub-section by summarizing the 2004 energy price trends, load levels, overall market expenses, and trends in individual components of the market expenses.

1. Energy Prices

Energy prices were generally higher in 2003 and 2004 than in previous years due to increased fuel prices. Changes in fuel prices are the primary driver of trends in electricity prices over extended periods. Even though much of the electricity used by New York consumers is generated from hydro, nuclear, and coal-fired generators, natural gas and oil units are usually the

marginal generation units which set prices in the market, especially during peak hours. Therefore, changes in the prices of these fuels will directly impact market prices.

The following analysis of monthly variations demonstrates that much of the changes in average electricity prices during 2003 and 2004 can be explained by corresponding changes in natural gas prices. Figure 1 shows average natural gas prices and electricity prices on a monthly basis during 2003 and 2004.

Figure 1: Energy and Natural Gas Prices 2003 – 2004

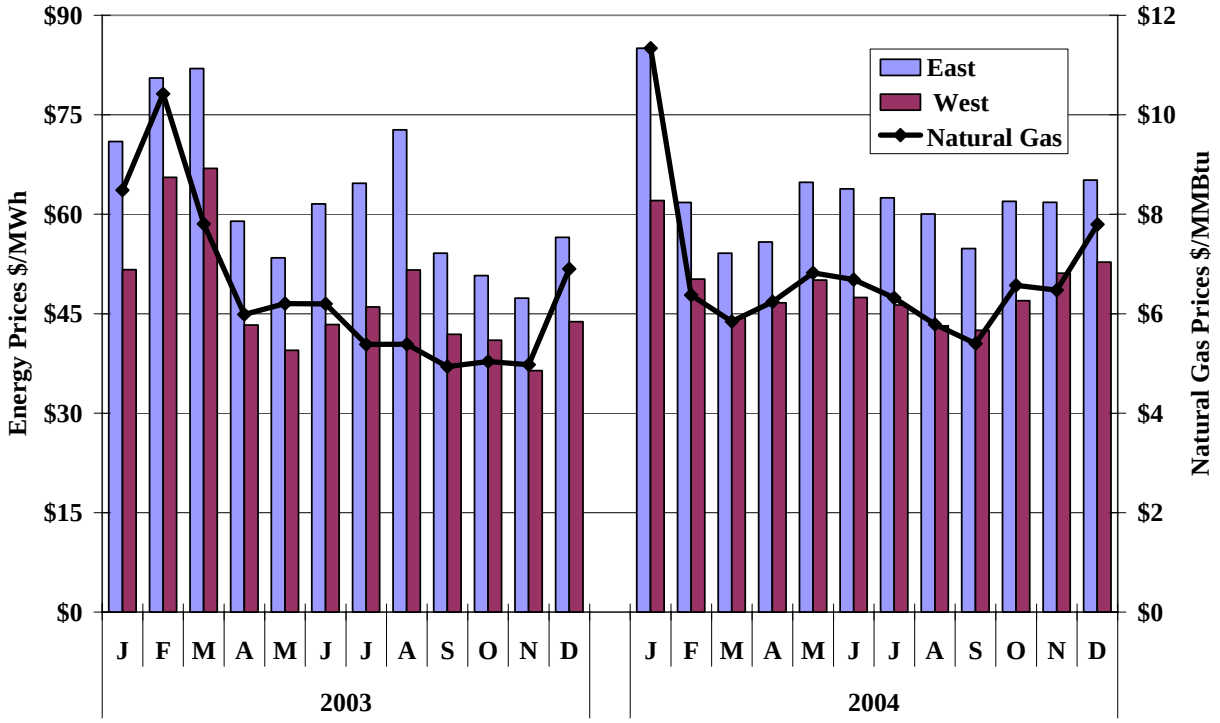


Figure 1 demonstrates that monthly average electricity prices are closely correlated with natural gas prices. Natural gas prices were very high during the first three months of 2003 due to extreme cold weather and low storage levels, but dropped to $5 to $6/MMbtu during the spring, summer, and fall. During the winter months at the end of 2003 and beginning of 2004, natural gas prices spiked again due to extreme cold temperatures and low inventories, but then settled down to $6/MMbtu and remained there until the winter. The average prices in the East and West of New York are shown generally moving in proportion with natural gas prices, although electricity prices rose less than natural gas prices during the months with extreme natural gas price spikes. This suggests that during periods of extreme natural gas price spikes, there is some

switching to other fuels, particularly oil, that moderate the impact of natural gas price fluctuations on electricity prices.

The highest electricity demand levels occur during the hot summer months and typically result in relatively high electricity prices, particularly during periods of supply shortages. During periods of shortage, prices can rise to more than 10 times the average price levels. Hence, a small number of price spikes can have a very significant impact on overall price levels. The figure above shows a modest rise in electricity prices during the summer of 2003 with no corresponding rise in natural gas prices. The weather during the summer of 2003 was cooler than in the two prior years when peak demand levels had led to more frequent price spikes and higher average prices during the summer. During the summer of 2004, the figure shows no significant rise in prices because the weather was very mild and there were no shortages.

In the NYISO markets, like other markets in the Northeast, electricity prices are primarily driven by natural gas prices and weather conditions during the summer. The following two figures show price duration curves that characterize the impact of these factors on electricity prices from 2002 to 2004.

Figure 2: Price Duration Curve – All Hours
Average Real-Time Price, 2002 – 2004

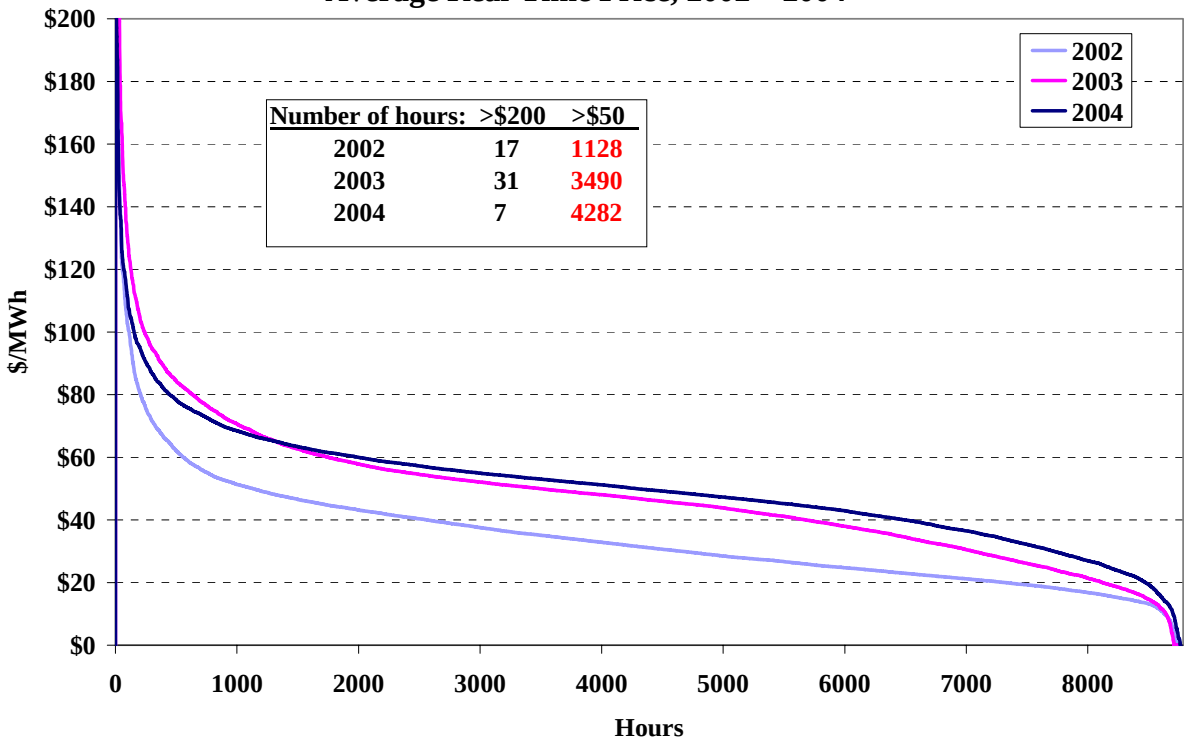


Figure 2 presents a price duration curve, which shows the number of hours on the x-axis in which the market settled at or above a given price level which is shown on the y-axis. Figure 2 shows that in most hours in 2003 and 2004, prices were significantly higher than in 2002. For example, there were almost 4300 hours with prices above $50 in 2004, compared to about 3500 hours in 2003 and less than 1200 hours in 2002. This primarily reflects the higher natural gas prices, which increased significantly after 2002, but were similar in 2003 and 2004.

Figure 3 is also a price duration curve that focuses on the highest priced 5 percent of hours, which account for a disproportionate share of the economic signals in any electricity market. Extreme price spikes generally occur during periods of peak demand, so the number and magnitude of relatively high prices depends on the severity of summer weather and the availability of generation. During 2004, there were only 7 hours when prices exceeded $200 per MWh, compared to 31 hours in 2003 and 17 hours in 2002. High prices were most frequent in 2002, when there were 6 hours with prices over $500 compared to only one hour in 2003 and none in 2004.

Figure 3: Price Duration Curves – Highest 5 Percent of Hours
New York State Average Real-Time Price, 2002 – 2004

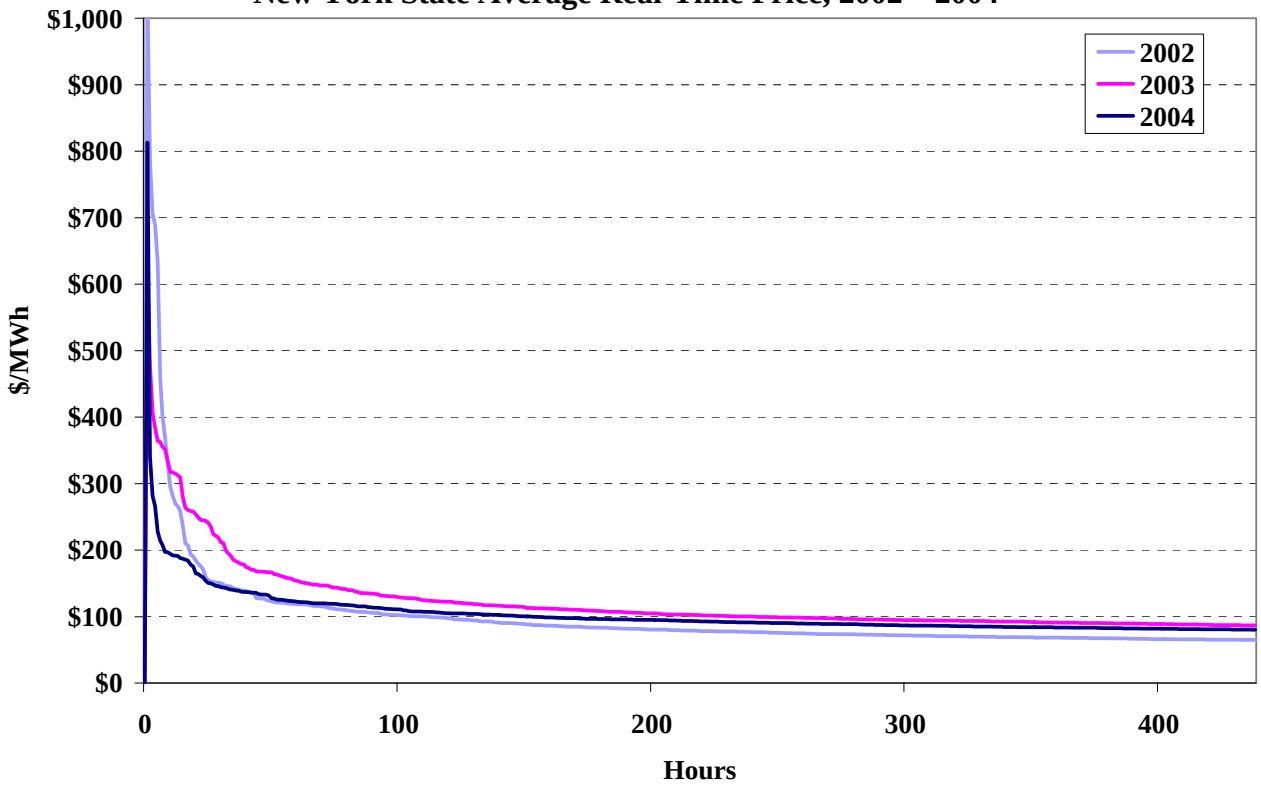
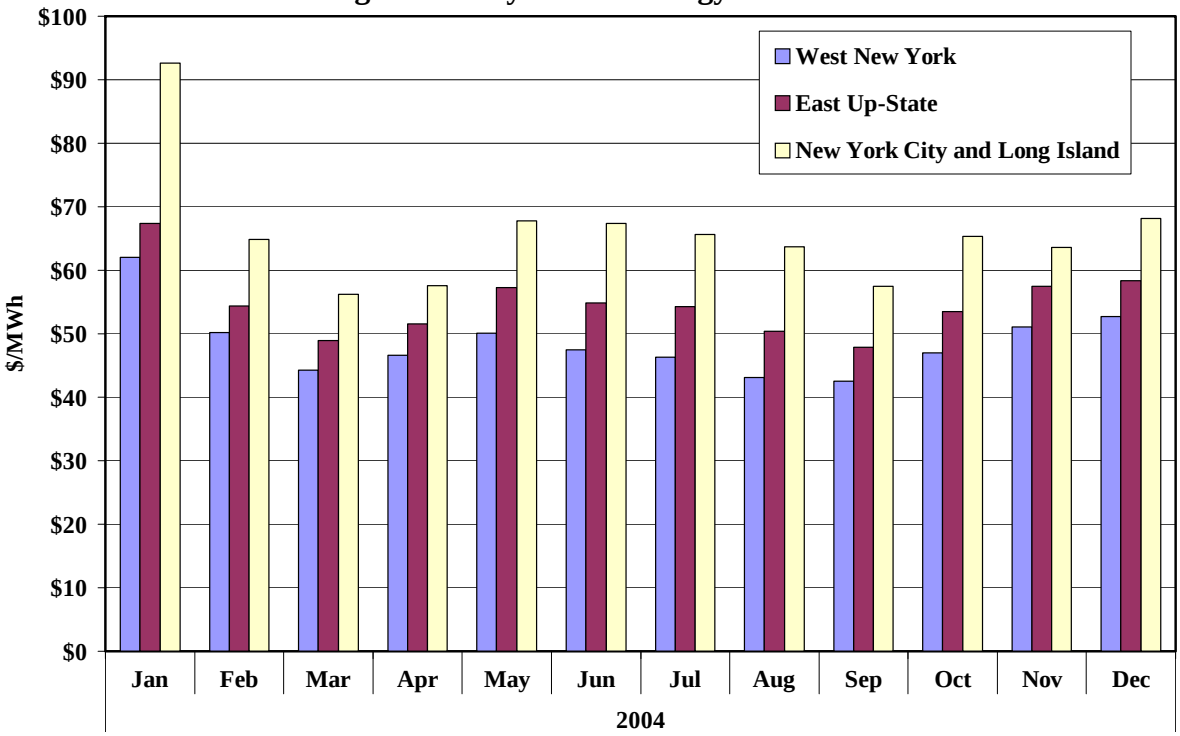


Figure 4 presents the monthly day-ahead energy prices in three regions in the State for 2004. Prices are lowest in Western New York, which exports significant amounts of power to Eastern New York. The prices are highest in New York City and Long Island which import a large portion of their power. Most of the power that flows from Western New York to New York City and Long Island passes through the Eastern upstate portion of the New York system. These west to east flows result in transmission losses and congestion that cause the pricing patterns shown in the figure.

Figure 4: Day-Ahead Energy Prices in 2004



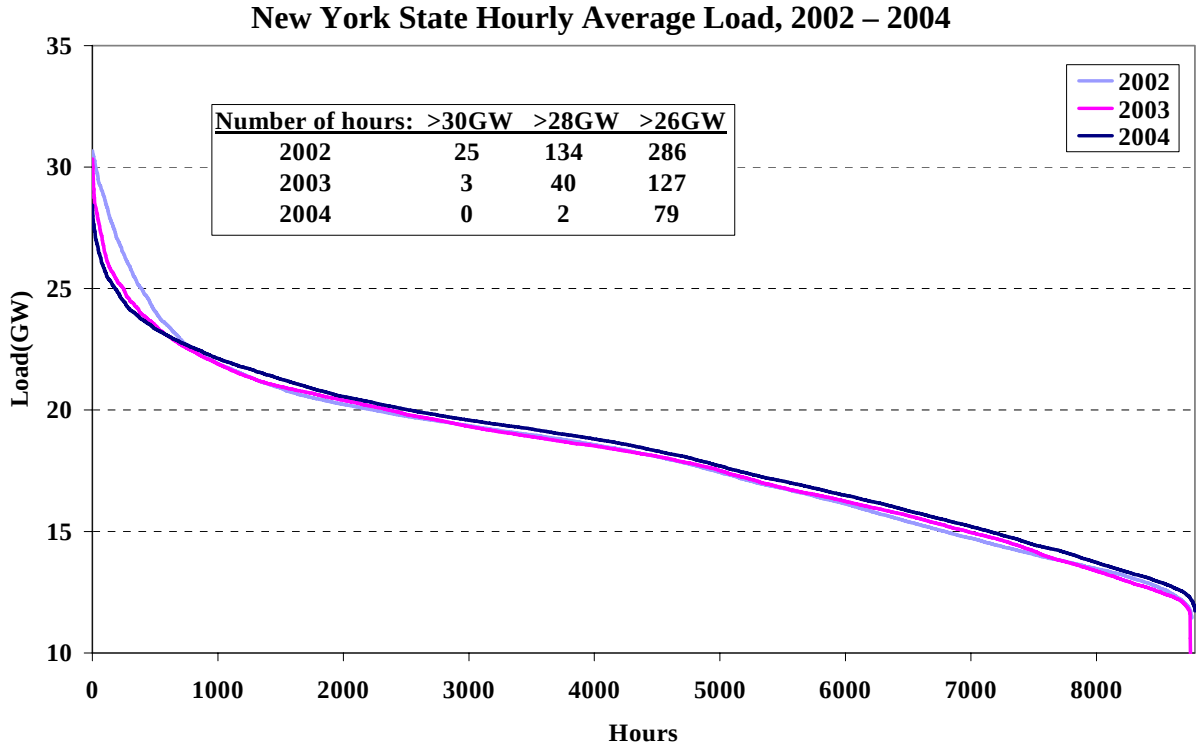
Transmission losses make up a significant share of the $6.06/MWh average difference between East and West prices, while transmission constraints, particularly on the Central-East Interface and in the Hudson Valley, make up the rest of the difference. There are transmission constraints into New York City and Long Island from upstate New York, as well as local load pockets within the City which resulted in price differences between the City and the eastern upstate region averaging $11.18/MWh.

The northeast experienced unusually cold weather and high fuel prices during January. The largest regional price difference occurred during January when the difference between Western New York and New York City and Long Island was more than $30/MWh. The greater

dependence on inefficient gas turbine units as the marginal source of energy in New York City and Long Island caused prices to be disproportionately impacted there by the January spike in natural gas prices. Other than January, June through August exhibited the largest differences between the West and downstate, at approximately $20/MWh. Load is highest during the summer months, which tends to increase the transmission flows between the West and New York City. This leads to more frequent congestion and a higher proportion of transmission losses.

One factor that has a major impact on changes in energy prices in the New York markets is the duration and timing of electricity demand. High prices resulting from a few days of extreme load conditions can raise the average price significantly for the entire year. During peak demand conditions, the relatively high prices are assessed to a larger volume of electricity purchases. Therefore, it is the hours when load peaks that result in both high prices and a substantial portion of all revenues received by energy suppliers, and thus a significant portion of costs experienced by customers. The summer of 2004 was notable because of the lack of days with extreme load levels. The following two figures compares load for the years 2002 - 2004.

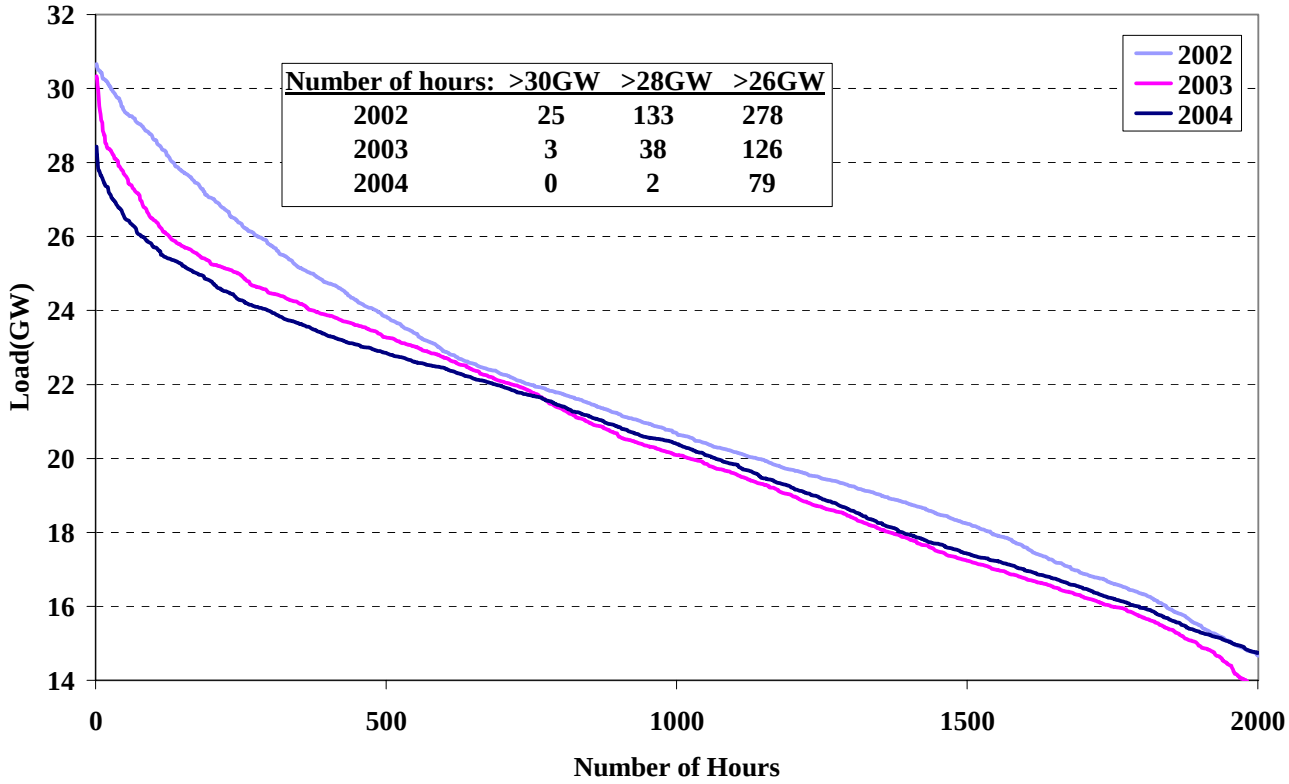
Figure 5: Load Duration Curves
New York State Hourly Average Load, 2002 – 2004



* Hours during the August 2003 blackout period are excluded.

The load duration curves in Figure 5 show that for the large majority of hours, load grew steadily from 2002 to 2004. However, mild conditions in the summer of 2003 and particularly in 2004 significantly reduced the frequency of extreme peak demand conditions. This is presented more clearly in Figure 6, which focuses on the summer months. At the highest load levels, the load duration curve for summer 2002 lies above the load duration curves for summer 2003 and summer 2004. While there were twenty-five hours in the summer of 2002 when actual loads exceeded 30 GW, there were only three hours in the summer of 2003 and none in the summer of 2004. Likewise, there were 133 hours when loads exceeded 28 GW in the summer of 2002 versus 38 hours in the summer 2003 and only two hours in the summer of 2004.

Figure 6: Load Duration Curves
Summer Hours, 2002 – 2004

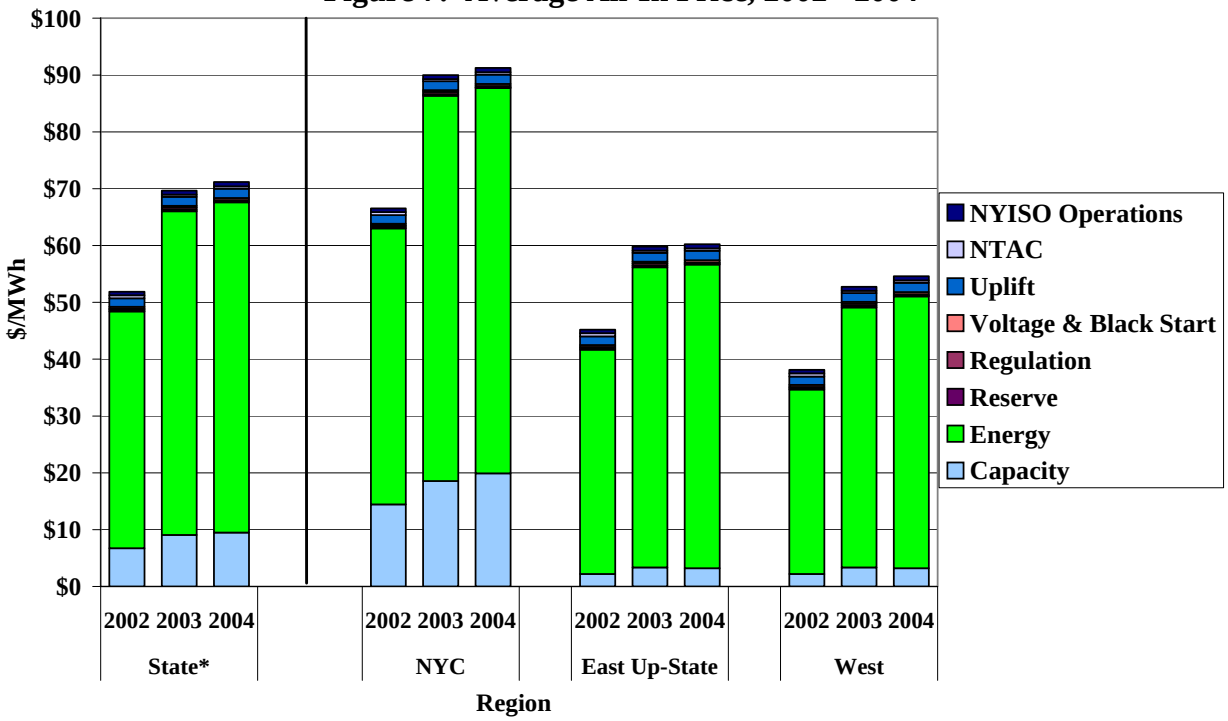


* Hours during the August 2003 blackout period are excluded.

Next we examine the all-in price for electricity which includes the costs of energy, uplift, capacity, ancillary services, congestion, losses, operating expenses, transmission, and other components of wholesale energy costs. The all-in price is calculated for various locations within New York because both capacity and energy prices vary substantially by location. The energy prices used for this metric are real-time energy prices. The capacity prices are a weighted

average of the capacity sold in each UCAP auction (6-month strip, monthly, and spot auctions). For the purposes of this metric, costs other than energy and capacity are distributed evenly for all locations. Figure 7 presents the average annual all-in price of electricity for the past three years.

Figure 7: Average All-In Price, 2002 - 2004



*Capacity portion of state price excludes Long Island.

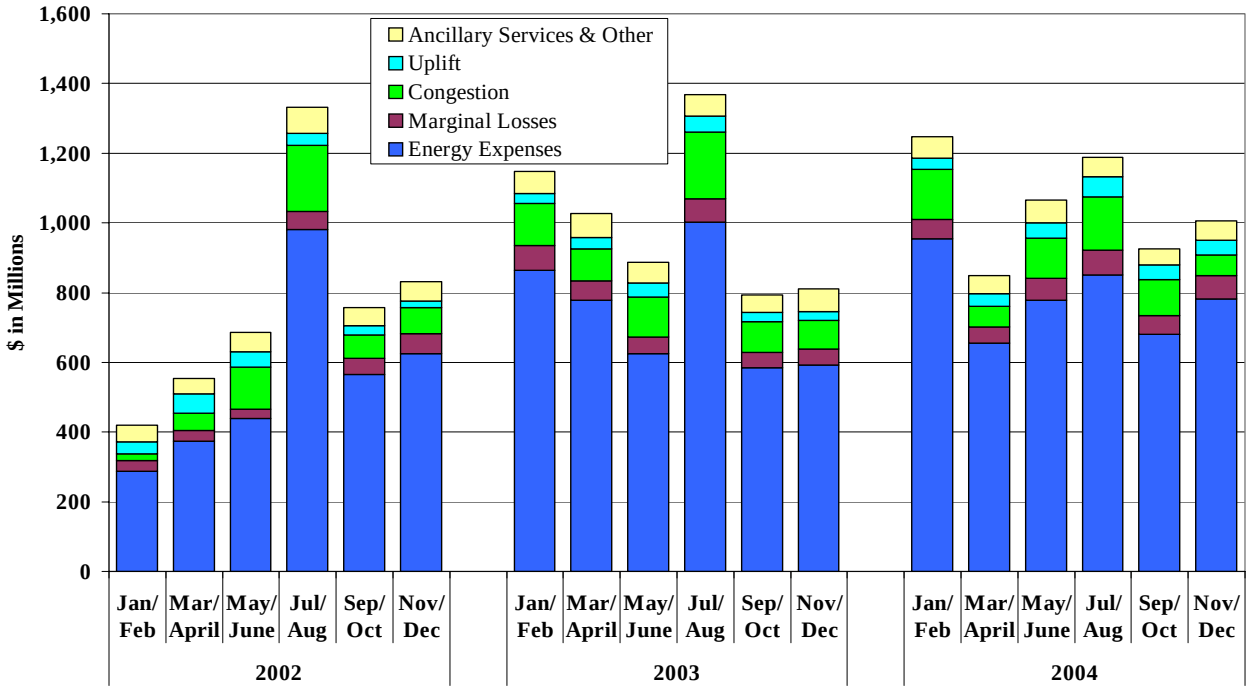
Figure 7 shows that the all-in price rose considerably in 2003 for all locations and increased slightly more in 2004. The increase is primarily caused by higher energy prices in 2003, which rose 36 percent in 2003 due to higher fuel prices. Fuel prices increased an additional 5 percent in 2004, but the impact on electricity prices was mitigated by mild summer weather. The capacity component also rose in 2003 and 2004 due primarily to: a) rising forecasted peak load resulting in a higher obligations, and b) additional purchases under the demand curve. Upstate, higher energy costs outweighed lower ancillary services and capacity costs, while in New York City the increase in capacity costs exceeded the slight decline in energy and ancillary services costs.

2. Total Market Expenses

Total market expenses include energy, ancillary services, congestion, losses, and uplift. The market expenses reflect settlements by the participants through the ISO markets and will, therefore, not include all costs of serving load. For example, physical bilateral schedules do not

settle the energy component of the schedule through the NYISO, only the congestion and losses costs. Between 40 and 50 percent of the load is scheduled in this manner. Figure 8 shows the total expenses for market participants of the NYISO for 2002 to 2004.

Figure 8: New York Electricity Market Expenses 2002 - 2004



Total electricity costs for 2004 were approximately $6.2 billion – two percent higher than in 2003 and a substantial increase from the $4.6 billion in total costs in 2001 and 2002. Changes in market expenses from 2002 to 2003 and 2004 were caused primarily by higher average energy prices due to higher fuel prices. Lower peak loads due to mild weather reduced total energy costs in 2003 and 2004 below the level that might have otherwise prevailed. A secondary contributor to higher costs has been a small reduction in scheduling of physical bilateral transactions, which increases the share of energy in New York settled through the NYISO markets. This does not mean that loads are more exposed to the NYISO market prices since they can execute forward financial contracts as hedges against price exposure that are not reflected in the NYISO settlements.

B. Prices and Price Convergence

In this section, we evaluate the convergence of prices between day-ahead and real-time markets, and the convergence of hour-ahead prices and real-time prices. Price convergence is an important measure of market performance because it indicates whether the market is efficiently arbitraging intertemporal prices, something that is desirable because it promotes the efficient commitment of generating resources and scheduling of external transactions.

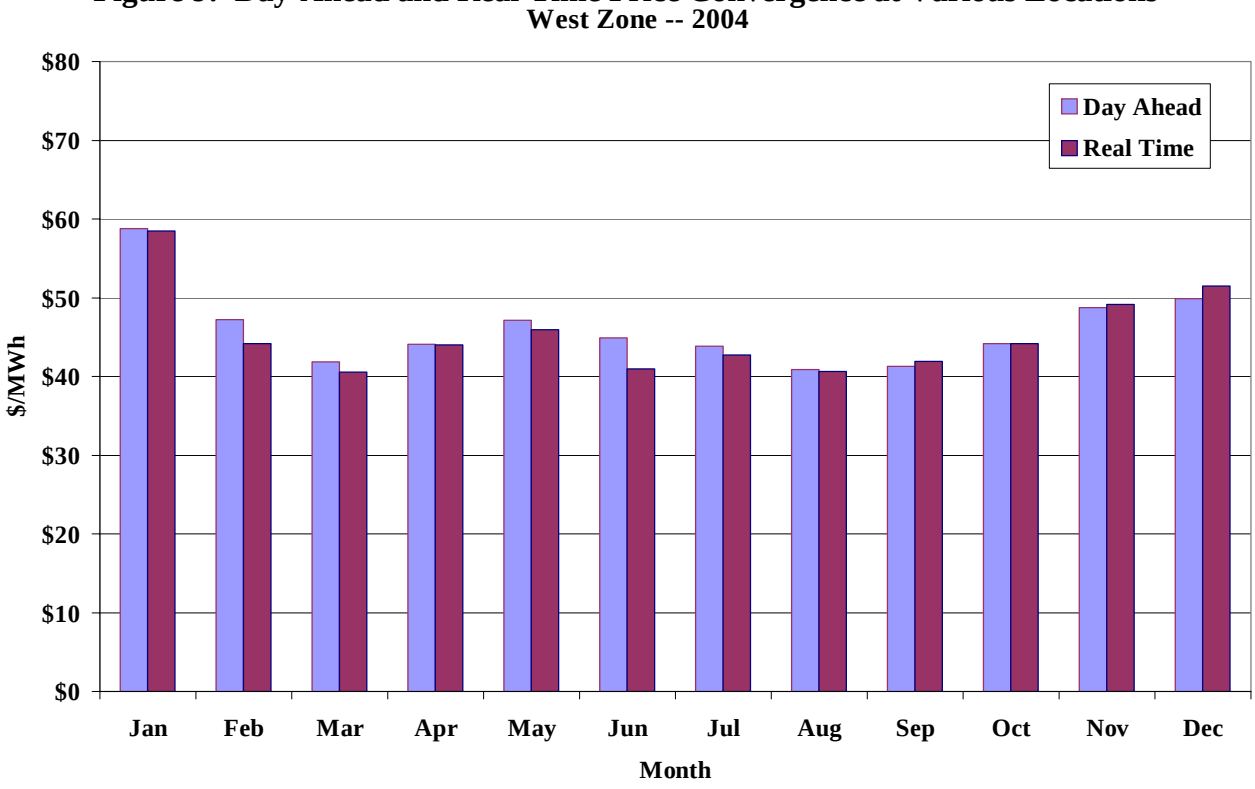
1. Day-Ahead and Real-Time Price Convergence

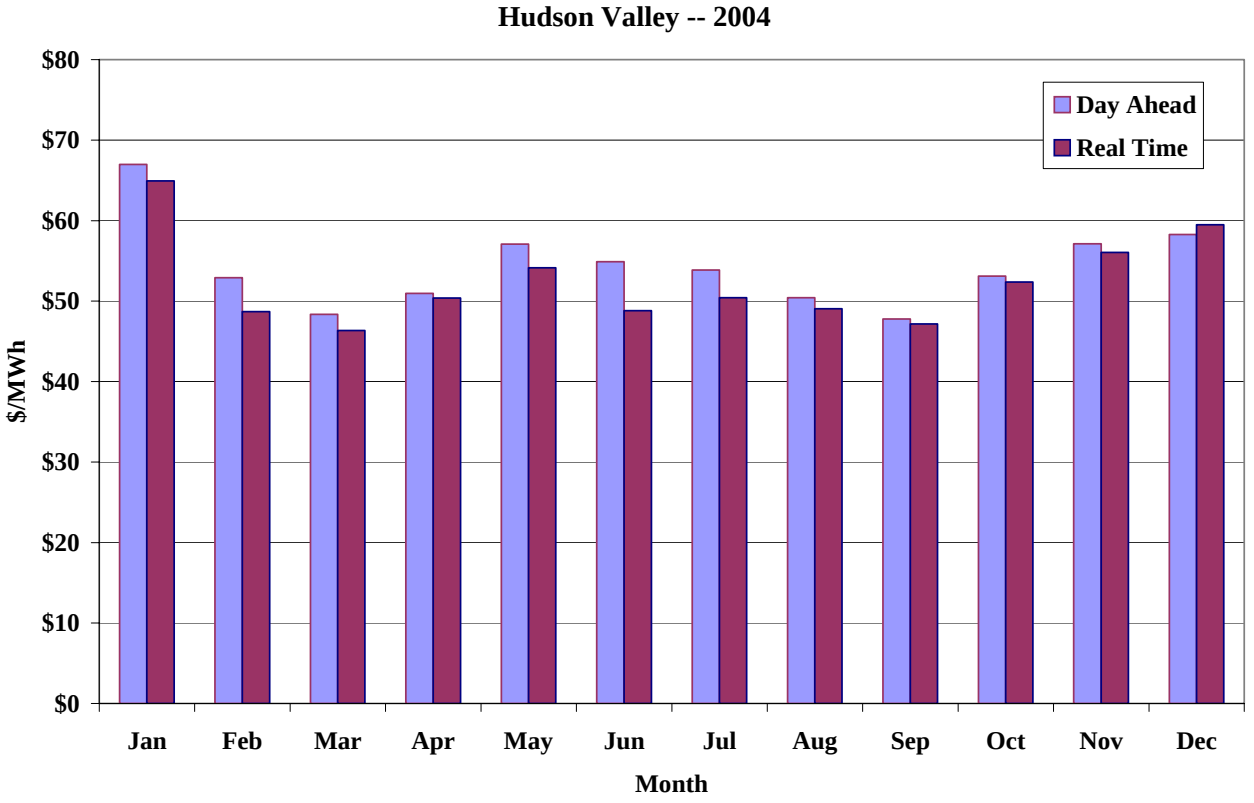
The day-ahead market allows participants to make forward purchases and sales of power for delivery in the real-time. This is a valuable financial mechanism that allows participants to hedge their portfolios and manage risk. Loads can insure against volatility in the real-time market by purchasing in the day-ahead market and by using TCCs in the day-ahead market to hedge against congestion. Generators selling in the day-ahead market are exposed to some risk associated with committing financially day-ahead. This is because they are committed to deliver physical quantities in the real-time market and an outage could force them to purchase replacement energy from the spot market during a price spike.

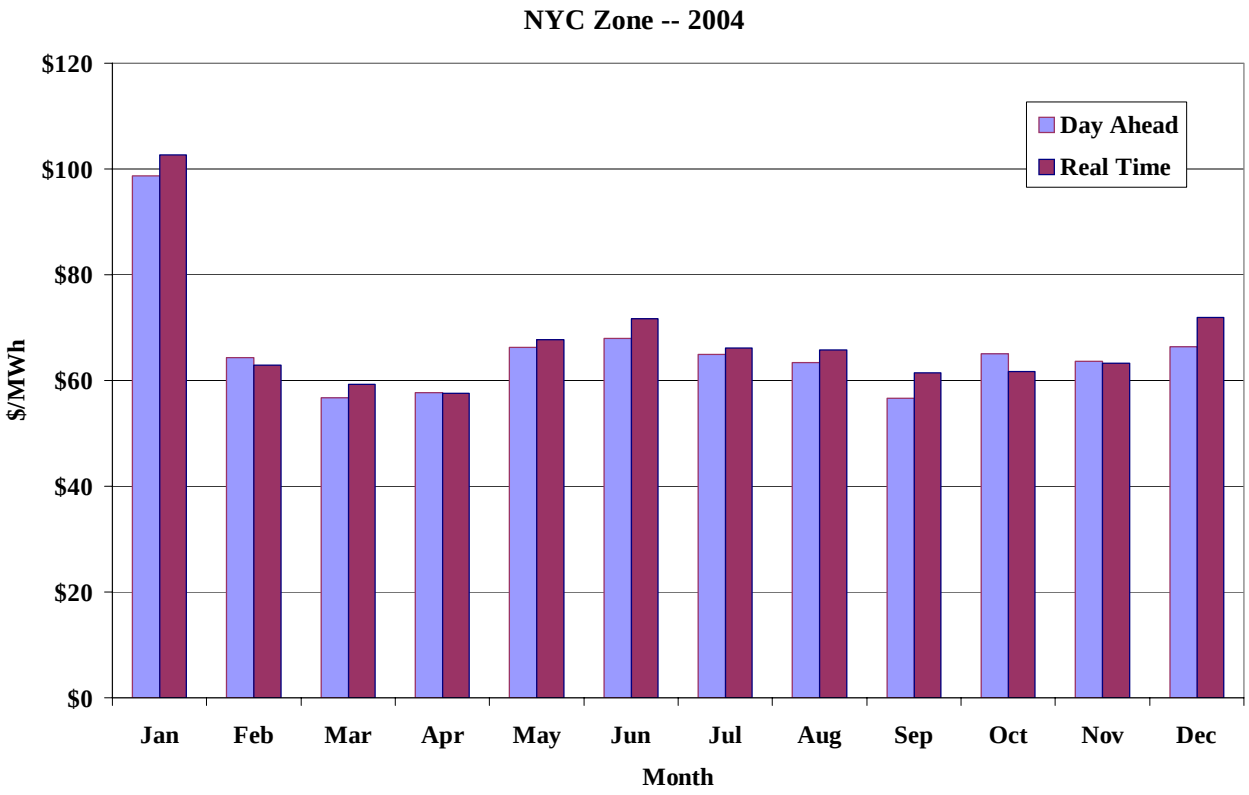
If participants are risk-averse, these factors will induce a premium in the day-ahead prices, which is generally consistent with the experience from other markets. However, day-ahead and real-time prices should not systematically diverge to a significant degree. Figure 9 shows a comparison of the average day-ahead and real-time energy prices in the West zone, Hudson Valley, and New York City for 2004.

The results generally show a premium associated day-ahead prices in the West zone (approximately 2 percent) and in the Hudson Valley (approximately 4 percent). While a day-ahead premium is consistent with risk averse behavior by generators and load serving entities, virtual trading activity tends to diminish the size of the premium. Generally, virtual traders can profitably arbitrage the price differences by selling virtually in areas with a day-ahead premium. Indeed, Section II of this report shows that in the areas with a day-ahead premium, there is far more scheduling of virtual supply than virtual load. This is improves convergence since virtual supply tends to reduce a day-ahead price premium.

Figure 9: Day-Ahead and Real-Time Price Convergence at Various Locations







The comparison of day-ahead and real-time prices yields different results for New York City, which shows a small real-time price premium (approximately 3 percent). When day-ahead prices are systematically lower than real-time prices, virtual traders have strong incentives to take advantage by scheduling virtual load. Hence, they purchase virtually at the lower day-ahead price and sell back at the higher real-time price. Section II of this report indicates that virtual traders predominantly schedule virtual load in New York City, which puts upward pressure on the day-ahead price and improves convergence. As discussed later in the report, the virtual load and higher real-time prices in New York City can be attributed, in part, to modeling inconsistencies between the day-ahead and real-time markets.

2. Price Convergence in Load Pockets

Prior to June 2002, the NYISO market software did not recognize the load pockets that existed inside New York City. This often resulted in inefficient commitment and dispatch decisions. Modeling of the load pockets within New York City, which was implemented in June 2002, has resulted in more accurate locational energy prices because prices now reflect the transmission constraints within New York City. This change has increased congestion expenses incurred in

the energy market while decreasing the amount of uplift paid to generators that are re-dispatched to resolve the load pocket constraints. The congestion within New York City leads to a wide variation in prices throughout the City. Therefore, convergence needs to be examined at various locations specific within the City.

Day-ahead and real-time prices differed by 3 percent on average for the New York City zone. However, the New York City zone price is a load-weighted average price based on the locational prices in each of the load pockets in the City. Therefore some locations may experience significant divergence in day-ahead and real-time prices that are off-set by divergences in the opposite direction at other locations. Hence, we conducted a further analysis of day-ahead and real-time prices at different locations throughout New York City. These results are shown in Figure 10 for 2002, 2003, and 2004.

Figure 10: Day-Ahead and Real-Time Prices in New York City 2002 – 2004

Figure 10 shows the ratio of the average day-ahead to real-time price. Thus, a ratio significantly higher than 100 percent indicates a day-ahead premium, while a ratio significantly below 100

percent denotes a real-time premium. For instance, the West zone and Hudson Valley zone are shown with small day-ahead premiums in 2004, while New York City zone and Long Island zone are shown with small real-time premiums. Figure 10 shows that day-ahead to real-time price convergence varied substantially by load pocket within the City during 2004. All four load pockets shown exhibited significant price premiums in real time, particularly Astoria East where the real-time premium was 13 percent. The 345 kV system (within the City, but outside the load pockets) exhibited a large premium of 6 percent in the day-ahead market.

From 2002 to 2004, Astoria East has consistently had lower day-ahead prices and the 345 kV system has consistently had higher day-ahead prices. However, the general pattern of divergence has changed for other load pockets in New York City over the three years. In 2002, Astoria West, Vernon/Greenwood, and Greenwood/Staten Island showed day-ahead premiums of 6 to 10 percent. In 2003 and 2004, the day-ahead prices have dropped relative to real-time prices so that in 2004 they all showed significant real-time premiums. Overall, price convergence has been far better at the zonal level than at the load pocket level.

Price convergence in the load pockets could be improved by introducing virtual trading within the New York City load pockets. Currently, virtual trading only occurs at the zonal level in New York. Limiting price-capped load bidding and virtual trading to the zonal level in New York City limits the ability of participants to arbitrage large price differences in specific load pockets. Allowing virtual trading at the load pocket or nodal level would likely improve convergence inside New York City load pockets.

Inconsistencies between the day-ahead and real-time market models may contribute to the lack of consistency between day-ahead and real-time prices at the load pocket level. Due to limitations of the model that was used to clear the real-time market in 2004, the Security-Constrained Dispatch ("SCD") model, a simplified representation of the intra-New York City constraints was used in real time market. NYISO uses a more detailed representation of the New York City system in the day-ahead market. In addition, the modeling of transmission losses was different in the two markets. These differences can contribute to divergence between the day-ahead and real-time prices within New York City. New real-time scheduling ("RTS") software, implemented in 2005, utilizes the same platform as the day-ahead market software, and should

make it easier to address these inconsistencies. We will be evaluating the performance of the post-RTS market after the summer 2005. If price convergence issues persist in New York City after the implementation of RTS, we recommend allowing virtual trading at a more disaggregated level.

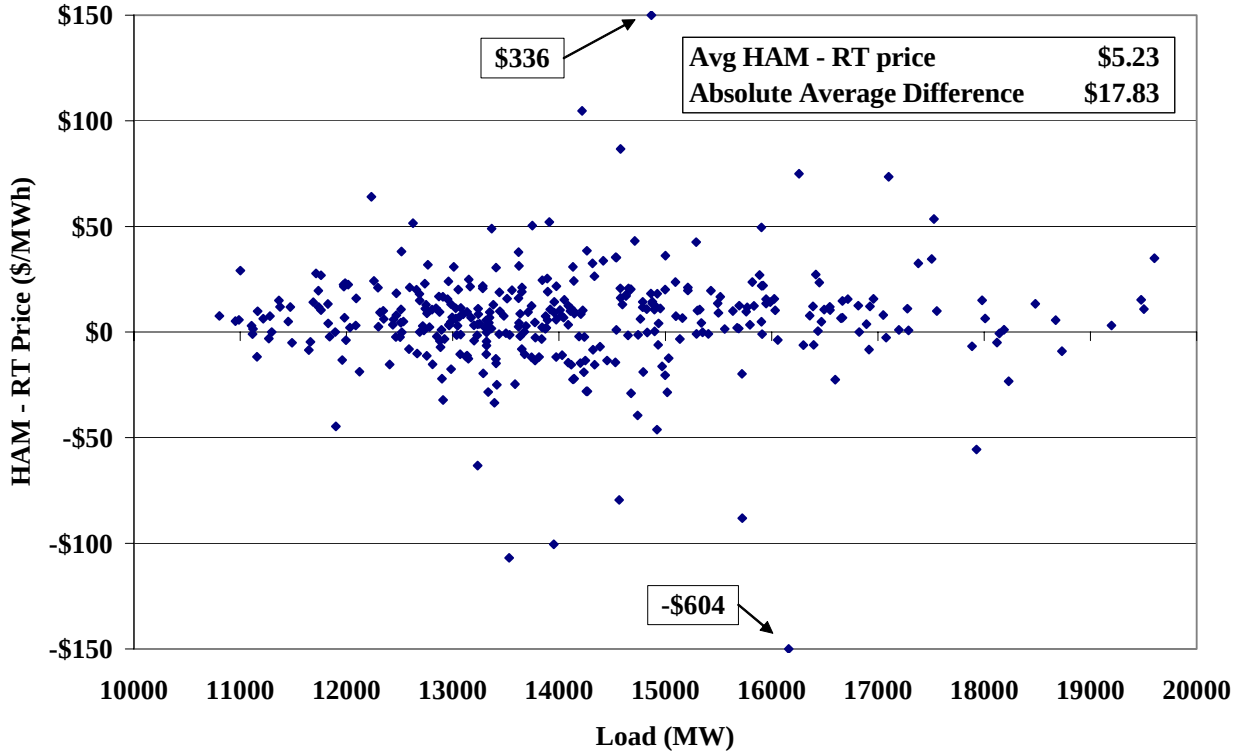
3. Hour-Ahead and Real-Time Price Convergence

Advisory hour-ahead prices are produced by the Balancing Market Evaluation ("BME") which commits generation and schedules external transactions and non-dispatchable units. Lack of convergence between hour-ahead and real-time prices can be a substantial concern because large price differences can result in external transactions and off-dispatch generation being scheduled inefficiently; resulting in increased uplift costs and inefficient real-time prices. Convergence tends to be the worst in the highest demand hours when prices are more volatile.

Changes to the market software and rules in 2002 dramatically improved price convergence between the hour-ahead and real-time models. These changes implemented in 2002 had the effect of treating 30 minute reserves more consistently between the hour-ahead and real-time models. This has led to more efficient commitment and scheduling by the BME. To measure the consistency between the hour-ahead and real-time outcomes, Figure 11 shows a comparison of hour-ahead and real-time prices in Eastern New York in the peak hour of the day during 2004.

Figure 11 shows the dispersion of price differences on the y-axis compared with the actual load level on the x-axis. Points above $0/MWh correspond to hours when the BME price was higher while points below $0/MWh denote hours when the SCD price was higher. Particularly during high demand periods, the commitment and scheduling decisions of the BME are very important for both preventing unnecessary real-time shortages and not over-committing uneconomic generation that can lead to substantial uplift costs. Thus, it is favorable that BME prices are slightly higher than the SCD prices on average and that there is no strong relationship between price convergence and load level.

Figure 11: Average Hour-Ahead and Real-Time Energy Prices
Eastern New York – Daily Peak Load Hours



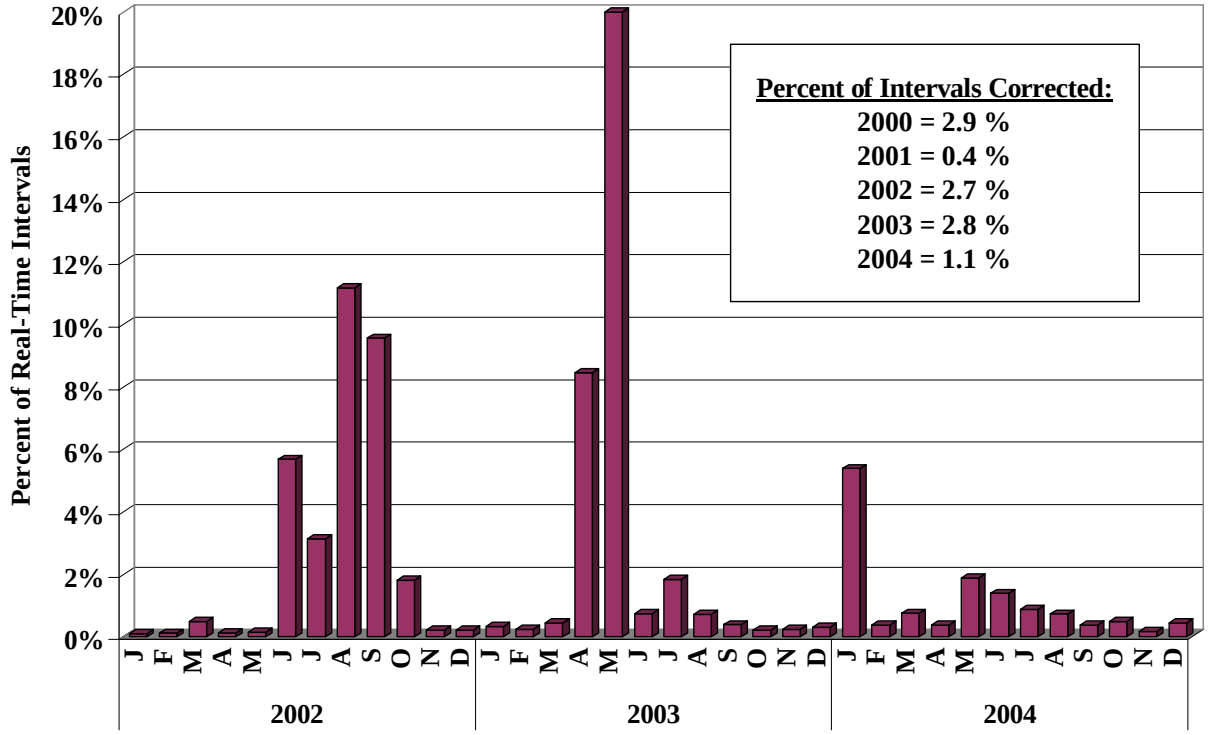
While significant improvements have already been made that drastically reduced these price differences, some differences remain. RTS, the new scheduling software that will replace the BME and SCD, will likely lead to further improvements in convergence between hour-ahead and real-time prices. The commitment software, RTC, is capable of scheduling externals and off-dispatch generation and committing resources every 15 minutes rather than hourly. The 5-minute dispatch model, RTD, co-optimizes reserves in a manner similar to the RTC. The similarities of the RTD and RTC models should further improve consistency between their outcomes.

C. Price Corrections

All real-time energy markets are subject to some level of price corrections to account for metering errors and other data input problems. Accurate prices are critical not only for the obvious need to settle market transactions fairly, but also for sending reliable real-time price signals to participants that have to make continual buy and sell decisions. Therefore, the incidence of these problems should be minimized. Price corrections are required when flaws in the market software or flaws in operating procedures cause prices to be posted erroneously. It is

important to resolve these errors as quickly as possible to maximize price certainty. Figure 12 summarizes the frequency of price corrections in the real-time energy market from 2002 to 2004.

Figure 12: Percentage of Real-Time Prices Corrected



The frequency of price corrections was relatively high in 2000 after the market opened, but then decreased steadily until the summer of 2002. The frequency of price corrections increased substantially in June, 2002 as a result of changes to the modeling of New York City load pockets. Once the modeling issues related to load pockets were addressed, the level of corrections returned to the low frequency that was experienced prior to the summer of 2002.

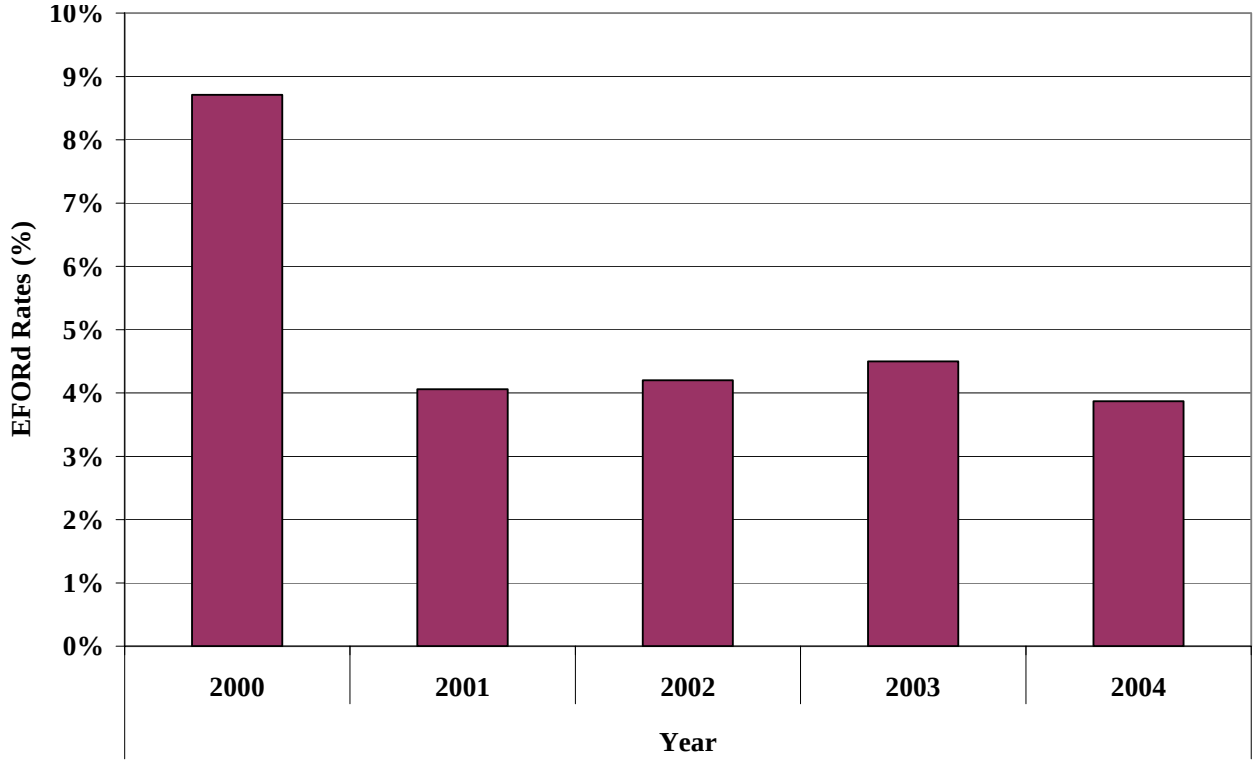
In 2003, there was a spike in the frequency of real-time price corrections in April and May before the frequency of price corrections returned to the historical norm. These price corrections occurred for most hours on fourteen days in April and seven days in May, resulting in slight changes to the New York City zonal prices that had been calculated with incorrect weightings.

During 2004, corrections occurred at a relatively low level. These results can be attributed in part to the fact that no major enhancements were made to the market software in 2004.

D. Forced Outages in 2004

We examined the trend in forced outages in the New York markets to ascertain if generators are responding to economic incentives to increase availability of their units. Figure 13 shows the Equivalent Forced Outage Rate ("EFOR"), which is used as a measure of forced outages. The EFOR is the portion of time a unit is unavailable due to forced outages, expressed as equivalent hours of full forced outage at its maximum net dependable capability.

Figure 13: Equivalent Demand Forced Outage Rates 2000 – 2004



EFOR declined substantially following the implementation of the NYISO markets. This is consistent with the incentives the deregulated markets provide to maximize availability, particularly during high load conditions. EFOR was relatively high in 2000 due to the outage of an Indian Point nuclear unit. After the Indian Point outage, the EFOR has been consistently close to 4 percent – much lower than the outage rates that prevailed prior to the implementation of the NYISO markets.

E. Market Power Mitigation

1. Background

The NYISO applies a conduct-impact test that can result in mitigation of participant bid parameters (i.e., energy offers, start-up and no-load offers, and physical parameters). The conduct test first determines whether bid parameters exceed pre-defined conduct thresholds. If at least one of the participant's bid parameters exceeds a conduct threshold, the bid parameter may be mitigated if the conduct results in sufficient impact on the energy price. While the NYISO tariff allows for mitigation to be invoked manually according to pre-defined criteria, this rarely occurs. Instead, the day-ahead and real-time market software are automated to perform most mitigation according to pre-defined conduct and impact thresholds.

Mitigation is applied in the real-time market for units in certain load pockets within New York City using the NYISO's conduct and impact approach. The in-city load pocket conduct and impact thresholds are set using a formula that is based on the number of congested hours experienced over the preceding twelve-month period.[4] An in-city bid will be mitigated if it exceeds the reference level by this threshold. This approach permits the in-city conduct thresholds to increase as the frequency of congestion decreases, whether due to additional generation or increases in transmission capability.

Prior to May 1, 2004, the day-ahead software used the conduct and impact test framework only for determining whether to mitigate outside New York City. Inside New York City, the day-ahead software would mitigate all units to their reference level (based on variable production expenses) whenever it detected at least a small amount of congestion between Indian Point and New York City. These mitigation procedures were referred to as the Consolidated Edison or "Con Ed" mitigation procedures as they were developed by Con Ed when it divested its generation. Under the Con Ed procedures, mitigation occurred nearly every day.

The Con Ed procedures were replaced on May 1, 2004 by the conduct and impact mitigation framework which was already being applied to the rest of the state. This framework significantly reduced the frequency of mitigation by making it more focused on potential market power in the

[4] $$\text{Threshold} = \frac{2\% * \text{Avg. Price} * 8760}{\text{Constrained Hours}}$$

NYC load pockets. This prevents mitigation from occurring when it is not necessary to address market power and allows high prices to occur during legitimate periods of shortage.

2. Mitigation in 2004

Figure 14 summarizes the frequency of constraints into the load pockets and the actual frequency of real-time mitigation. When the constraints shown are binding, resources with offers exceeding their reference levels by more than the load pocket's conduct threshold may warrant mitigation. The total height of each column in the figure shows the percentage of the hours in which the constraints into a load pocket are binding. Of those intervals, the lower portion of the columns shows the portion of the intervals in which one or more units in the given load pockets were mitigated.

Figure 14: Frequency of Real-Time Constraints and Mitigation
New York City Load Pockets, 2004

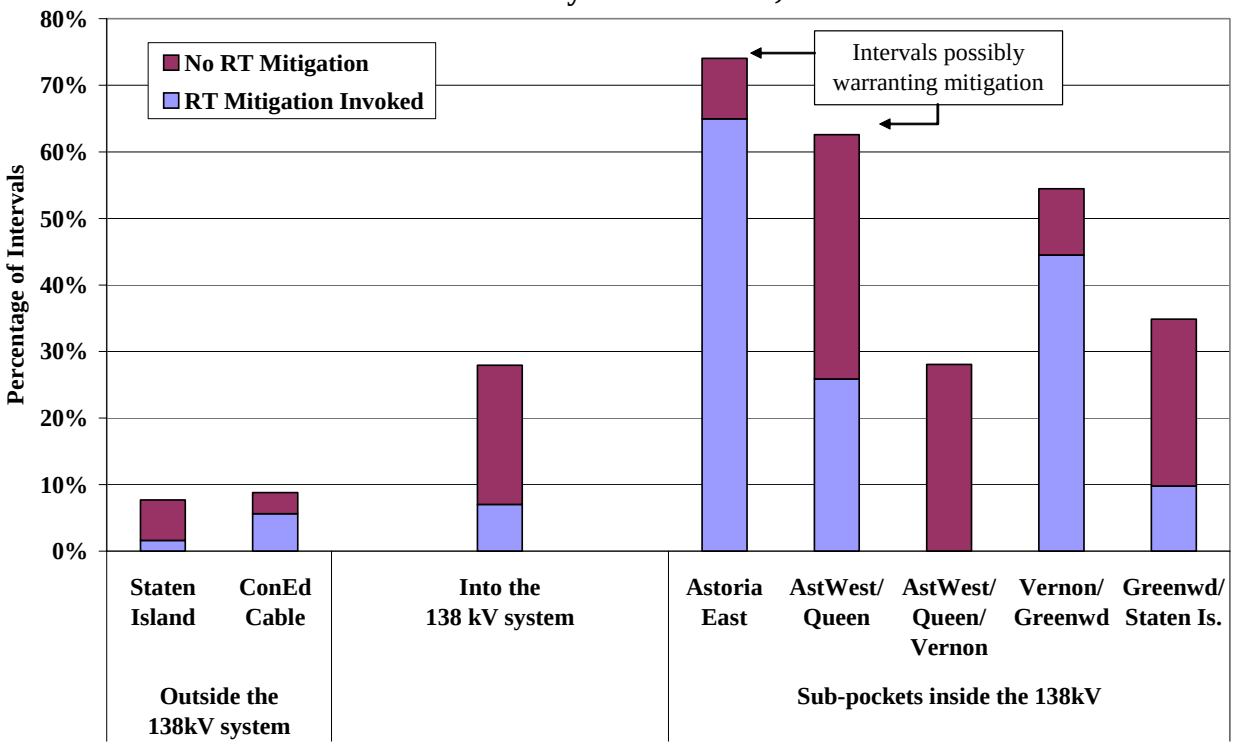


Figure 14 shows that mitigation was generally more frequent in the most congested load pockets where market power is of greater concern. Mitigation was much less frequent in the more competitive areas, occurring less than 10 percent of the time in the areas outside the 138 kV

system. The low frequency of real-time congestion is partly due to the fact that day-ahead mitigated offers are carried into the real-time up to the day-ahead schedule of the unit.

As noted above, prior to the conduct and impact framework being implemented in the day-ahead market for New York City load pockets, day-ahead mitigation occurred in nearly every hour of every day under the ConEd measures. Figure 15 shows the day-ahead mitigation results in 2004, following the implementation of the conduct and impact framework in New York City.

Figure 15: Frequency of Day-ahead Constraints and Mitigation
New York City Load Pockets, June to December 2004

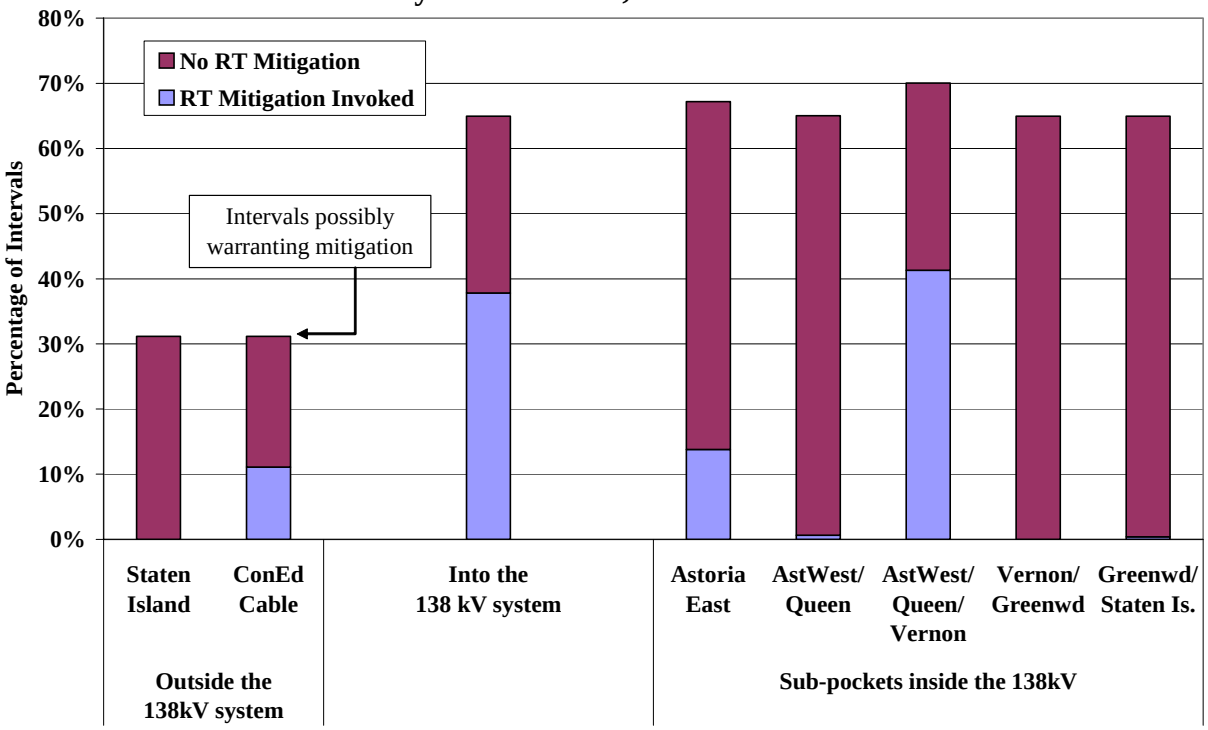


Figure 15 shows that day-ahead mitigation has become much less frequent under the conduct and impact framework in New York City. Outside of the load pockets in the City, mitigation occurred in 11 percent of hours while congestion was noted in 31 percent of hours, so mitigation was only invoked about 35 percent of the time that congestion was experienced. Within the load pockets, mitigation was most commonly associated with the constraint into the 138 kV system and into the Astoria West/ Queensbridge/Vernon load pocket. These results are consistent with our expectations under the conduct and impact framework.

F. Net Revenues Analysis

Revenues from the energy, ancillary services, and capacity markets provide the key signals for investment in new generation and retirement of existing generation. The decision to build or retire a generation unit will depend on the expected net revenues that unit will receive in the market from sales of energy, ancillary services, and capacity. Net revenue is defined as the total revenue that a generator would earn in the New York markets less its variable production costs.

If there is not sufficient net revenue in the short-run from these markets to justify entry of a new generator, then one or more of the following conditions may be present: (i) new capacity is not needed because there is sufficient generation already available; (ii) load conditions, due to mild weather and/or a reduction in demand, and thus energy prices, are below long-run expected values; and/or (iii) market rules are causing revenues to be reduced inefficiently. Likewise, the opposite would be true if prices provide excessive revenues in the short-run. If a revenue shortfall persists for an extended period, without an excess of capacity, this is a strong signal that markets need modifications.

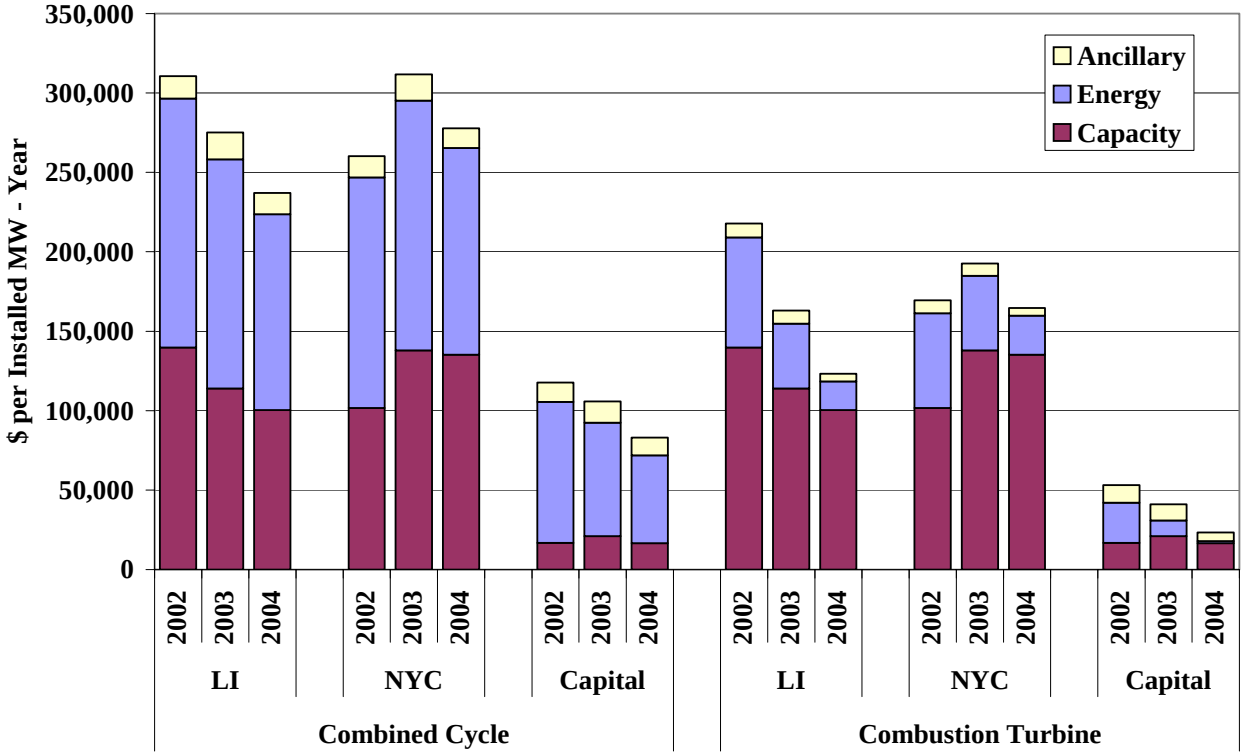
In this section we analyze the net revenues that would have been received in 2004 by various types of generators at three different locations, New York City, Long Island and the Capital zone. We calculated the net revenue the markets would have provided to two different types of units at these locations for the last three years. The two types of units are:

- Gas combined-cycle: heat rate assumed of 7000 BTU/kWh and
- New gas turbine: heat rate assumed of 10,500 BTU/kWh.

It is important to note that combined cycle generators have significant start-up costs and start-up times and minimum run time requirements that exceed one hour. Because this analysis calculates net revenue equal to the LMP minus the variable production costs on an hourly basis, ignoring these commitment considerations, the net revenue values shown below for the combined cycle units will tend to be overstated. In addition, gas turbines frequently purchase natural gas in the intraday market, which generally trades at a slight premium to the day-ahead price used in this analysis. Therefore, the net revenue for the gas turbine shown below is also likely to be slightly higher than an actual new unit would realize. Nonetheless, the assumptions for this analysis have been standardized by FERC and the market monitors in the various

markets to provide a comparable basis for comparison of the net revenue values from the different markets.

Figure 16: Estimated Net Revenue in the Day-Ahead Market 2002 - 2004



As Figure 16 indicates, a new gas fired combustion turbine (with a heat rate of 10,500 BTU/kWh) would earn revenue in New York City in the range of $150,000 to $175,000 per MW-year. This would recover approximately 85 percent to 99 percent of the net revenue required to support such an investment.[5] The results for the combined-cycle unit are less clear. While a new combined-cycle plant would earn from $250,000 to a little over $300,000 per MW-year, there is no publicly available data on the costs of investing in a new combined-cycle plant inside New York City.[6]

5 Levitan & Associates, Inc., *Independent Study to Establish Parameters of the ICAP Demand Curves for the New York Independent System Operator (2004).*

6 Energy Information Administration Annual Energy Outlook 2005 includes an estimate that does not take into account the extra costs of building in a densely populated area such as New York City.

Net revenues increased from 2002 to 2003 in New York City due to a moderate rise in UCAP prices. However, net revenues decreased in 2004 associated with reductions in the net revenue from the energy market due to the mild summer weather. Likewise, net revenues in the Capital zone decreased over the three-year period due primarily to relatively low peak-load conditions during the summer that have prevented any instances of shortages.

It is apparent that entry is not likely to be economic outside of New York City based on the net revenue results for the past three years. There are two principal reasons why this result is consistent with our expectations. First, the mild weather of the last two summers substantially reduced the net revenue from the energy market because it reduced the peak loads and contributed to the lack of shortages in 2003 and 2004. Very high energy prices during transitory periods of shortage are an important component of the long-term economic signal that new resources are needed in a market. Second, there is a substantial surplus of generating capacity in upstate New York, resulting in relatively low capacity prices and contributing to the lack of shortages. Therefore, the fact that net revenue has been insufficient to support the entry of new generation in upstate New York is not cause for concern.

In New York City, however, the net revenue results are less clear. Capacity margins in New York City have been very close to the minimum requirements, so one would expect the net revenue to be close to or exceed the entry costs of a new unit. This may well be the case for a new combined cycle resource, although the entry costs of such resources in the City are not known. In addition, under normal weather conditions and, thus, higher energy net revenue over the past two years, the net revenue for a new gas turbine would exceed its entry costs in the City. These results are confirmed by the fact that most new construction planned in the near-term is occurring in New York City.

We also compared the net revenue in the Capital zone of New York with net revenue in other centralized wholesale markets. Figure 17 compares estimates of net revenue for each of the auction-based wholesale electricity markets in the U.S.: (a) the ERCOT North Zone, (b) NP15 in the California ISO, (c) the Capital zone in New York, (d) ISO New England Hub, and (e) the average for PJM. The figure includes estimates of net revenue from (a) energy, (b) reserves and regulation, and (c) capacity.

Figure 17: Comparisons of Estimated Net Revenue
Five U.S. Markets, 2002 – 2004

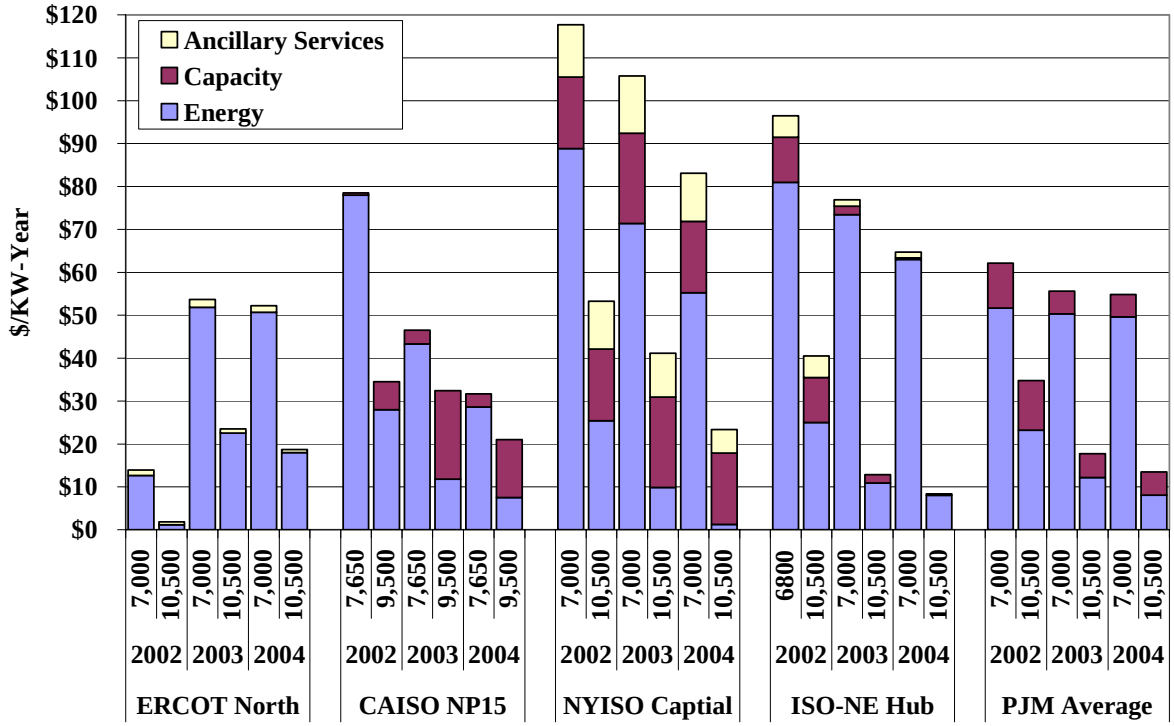


Figure 17 shows that net revenues fell moderately across all markets from 2003 to 2004 as most areas experienced a very mild summer in 2004. Net revenues decreased slightly or remained flat from 2002 to 2003 for every market except ERCOT. The Capital zone of New York exhibits estimated net revenues from energy that are comparable to New England. In previous years, ERCOT and PJM had substantially lower energy net revenues than upstate New York. However, upstate New York's energy net revenue has decreased, causing its energy net revenue to be comparable to ERCOT's and PJM's in 2004.

Overall, New York has the highest estimated net revenue due to its capacity and ancillary services markets, which send more accurate and efficient economic signals than the other markets in this analysis. New York is the only ISO or RTO with markets for locational capacity, regulation, 10-minute spinning reserves, 10-minute total reserves, and 30-minute reserves.

II. ANALYSIS OF ENERGY BIDS AND OFFERS

In this section, we examine bidding patterns to evaluate whether market participant conduct is consistent with efficient and effective competition. On the supply side, the analysis seeks to identify potential attempts to withhold generating resources as part of a strategy to increase prices. On the demand side, we evaluate load-bidding behavior to determine whether load bidding has been conducted in a manner consistent with competitive expectations. We also analyze virtual trading in this section.

A. Analysis of Supply Offers

Wholesale electricity production is attributable primarily to base-load and intermediate-load generating resources. Relatively high-cost resources are used to meet peak loads and comprise a very small portion of the total supply. The marginal cost of base-load and intermediate-load resources do not vary substantially relative to the marginal cost of resources used at peak times. This causes the market supply curve to be relatively flat at low and moderate output levels and steeply sloped at high output levels. Therefore, as demand increases from low load levels, (as an almost vertical demand curve shifts along the supply curve) prices remain relatively stable until demand approaches peak levels, where prices can increase quickly as the more costly units are required to meet load. The shape of the market supply curve has critical implications for evaluating market power.

Suppliers exert market power in electricity markets by withholding resources and increasing the market clearing price. This can be accomplished through physical withholding or economic withholding. Physically withholding occurs when a resource is derated or not offered into the market when it is economic to do so. Economic withholding occurs when a supplier raises the offer price of a resource to reduce its output below competitive levels or to otherwise raise the market price. Demand must be high enough that withholding a resource has the potential to significantly impact market price. When the market clears along the flat portion of the supply curve, prices will be relatively insensitive to withholding.

An analysis of withholding must distinguish between strategic withholding aimed at exercising market power and competitive conduct that could appear to be strategic withholding.

Measurement errors and other factors can erroneously identify competitive conduct as market power. For example, a forced outage of a generating unit may be either legitimate or a strategic attempt to raise prices by physically withholding the unit.

To distinguish between strategic and competitive conduct, we evaluate potential withholding in light of the market conditions and participant characteristics that would tend to create the ability and incentive to exercise market power. Under competitive conditions, suppliers maximize profits by increasing their offer quantities during the highest load periods to sell more power at the higher peak prices. Alternatively, a supplier that possesses market power will find withholding to be profitable during periods when the market supply curve becomes steep (i.e., at high-demand periods). Therefore, examining the relationship between the measures of potential withholding and demand levels will allow us to test whether the conduct in the market is consistent with workable competition.

1. Deratings and Physical Withholding

We first consider potential physical withholding by analyzing generator deratings. A derating occurs when a participant reduces the maximum output available from the plant. This could be for planned outages, long-term forced outages, or short-term forced outages. A derating could be partial (maximum output is reduced, but is greater than zero) or complete (maximum output is zero). We analyze only the summer months to effectively eliminate planned outages from our data. By eliminating planned outages, we implicitly assume that planned outages are legitimate and are not aimed at exercising market power.[7] The remaining deratings data would then include only long-term and short-term deratings. We first analyze both long-term and short-term deratings together. In our second analysis, we focus on short-term deratings because short-term deratings are more likely to reflect attempts to physically withhold since it is more costly to withhold via long-term deratings or outages.

[7] Planned outages are usually scheduled far in advance, and are almost always scheduled for a period during the year when demand is historically at low levels, in New York, that would be the spring and autumn months. Since weather forecasters are currently incapable of predicting unusual weather events, like record setting heat waves in May, the fact that a planned outage results in higher prices in those circumstances is not evidence of the exercise of market power. Thus, only outages which occur during periods when the supplier can anticipate a benefit from withholding are relevant to the market power analysis.

We focused on the afternoon hours which have higher demand because, under a hypothesis of market power, we would expect to find that withholding increases as demand increases. We also limited ourselves to the locations east of the Central-East interface, as this area, which includes two-thirds of the State's load, has limited import capability, and is more vulnerable to the exercise of market power.

Figure 18 and Figure 19 show our analysis for all deratings and for short-term deratings, respectively.

Figure 18: Relationship of Deratings to Actual Load
Day-Ahead Market - East New York
Weekdays, Noon to 6 PM, Summer 2004

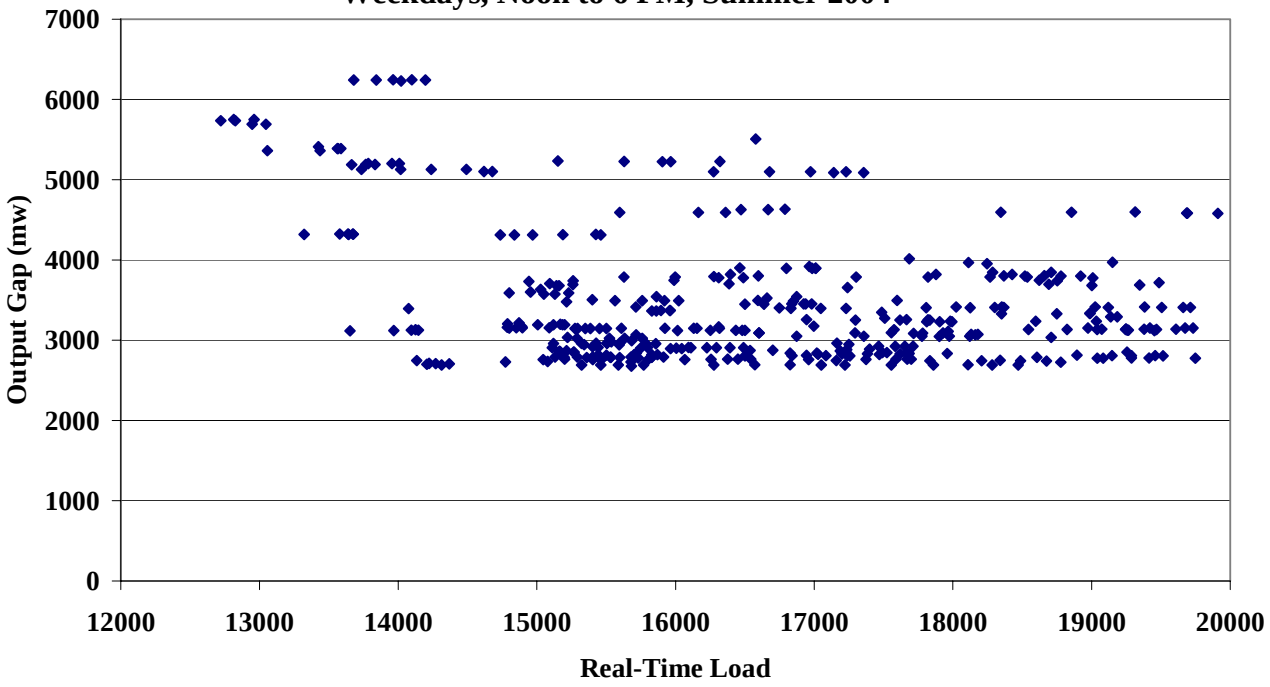


Figure 19: Relationship of Short-Term Deratings to Actual Load
Day-Ahead Market – East New York
Weekdays, Noon to 6 PM, Summer 2004

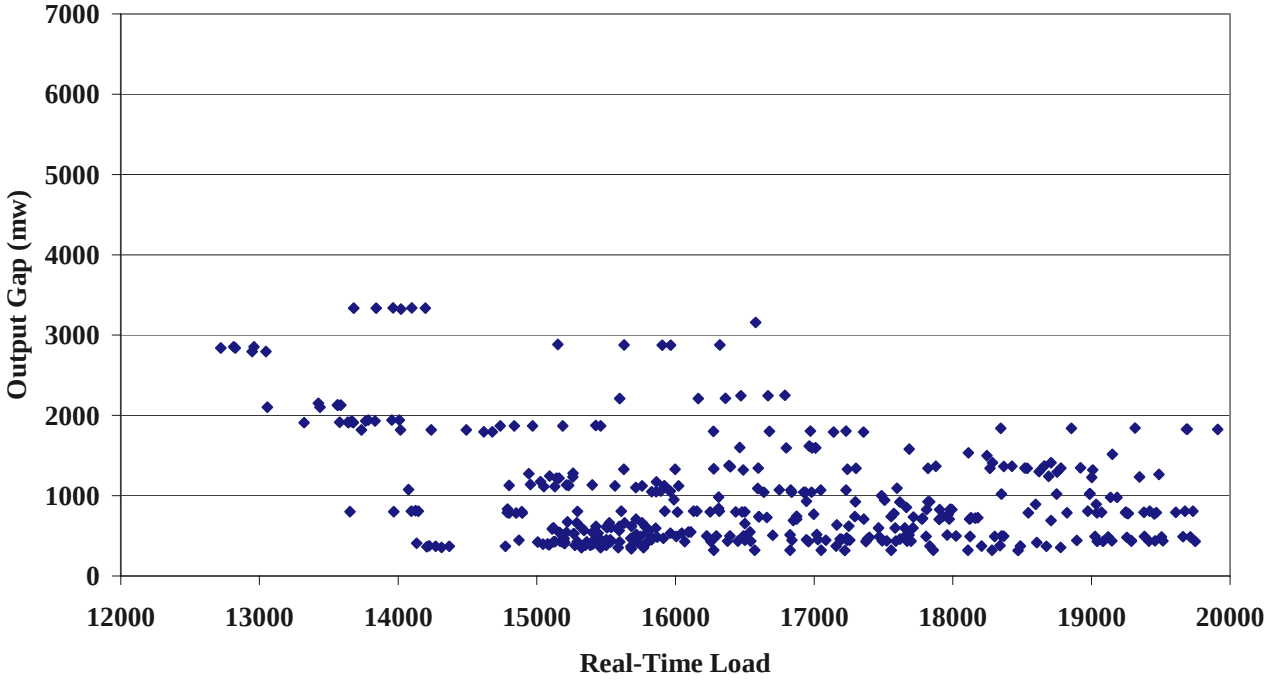


Figure 18 indicates that the total quantity of deratings is generally smaller when demand reaches very high levels. Focusing on short-term deratings in Figure 19, we also found less deratings during high demand periods. This indicates good competitive performance since the incentive to physically withhold resources would generally increase under high demand conditions for participants with market power. Furthermore, while forced outages are generally random, the total forced outages would be expected to rise slightly under peak demand conditions when the ISO must call on units that operate infrequently. Therefore, we find that the overall pattern of outages and deratings was consistent with workable competition during the summer of 2004.

2. Output Gap and Economic Withholding

To evaluate economic withholding, we calculated the hourly "output gap". The output gap is the quantity of generation capacity that is economic at the market clearing price, but is not running due to the owner's offer price or is setting the LBMP with an offer price substantially above competitive levels (excluding capacity scheduled to provide ancillary services). This withholding can be accomplished through high start-up cost offers, high minimum generation offers, and/or high incremental energy offers.

To determine whether an offer is above competitive levels, we use reference values based on the past offers of the participant during competitive periods. A supplier will normally offer at levels near marginal cost, because during periods when market power is unlikely to be exercised, excessive offers will cause the unit not to be dispatched and cost the owner lost profits. We allow considerable tolerance in our threshold. An offer parameter is indicated as above competitive levels if it exceeds the reference values by a given threshold. We conduct the analysis with thresholds matching the conduct threshold used by the state-wide automated mitigation procedure ($100/MWh or 300 percent, whichever is lower) and a lower threshold ($50/MWh or 100 percent, whichever is lower).

Like our analysis of deratings, we examine the relationship of the output gap to the market demand level. We focus our analysis on Eastern New York where market power is most likely. Figure 20 shows the output gap using the state-wide mitigation thresholds of $100/MWh or 300 percent. To assess whether there have been significant attempts to withhold by offering just below the state-wide mitigation threshold, Figure 21 shows the output gap results using a lower threshold of $50/MWh or 100 percent.

Figure 20: Relationship of Output Gap at Mitigation Threshold to Actual Load
Real Time Market – East New York
Weekdays, Noon to 6 PM

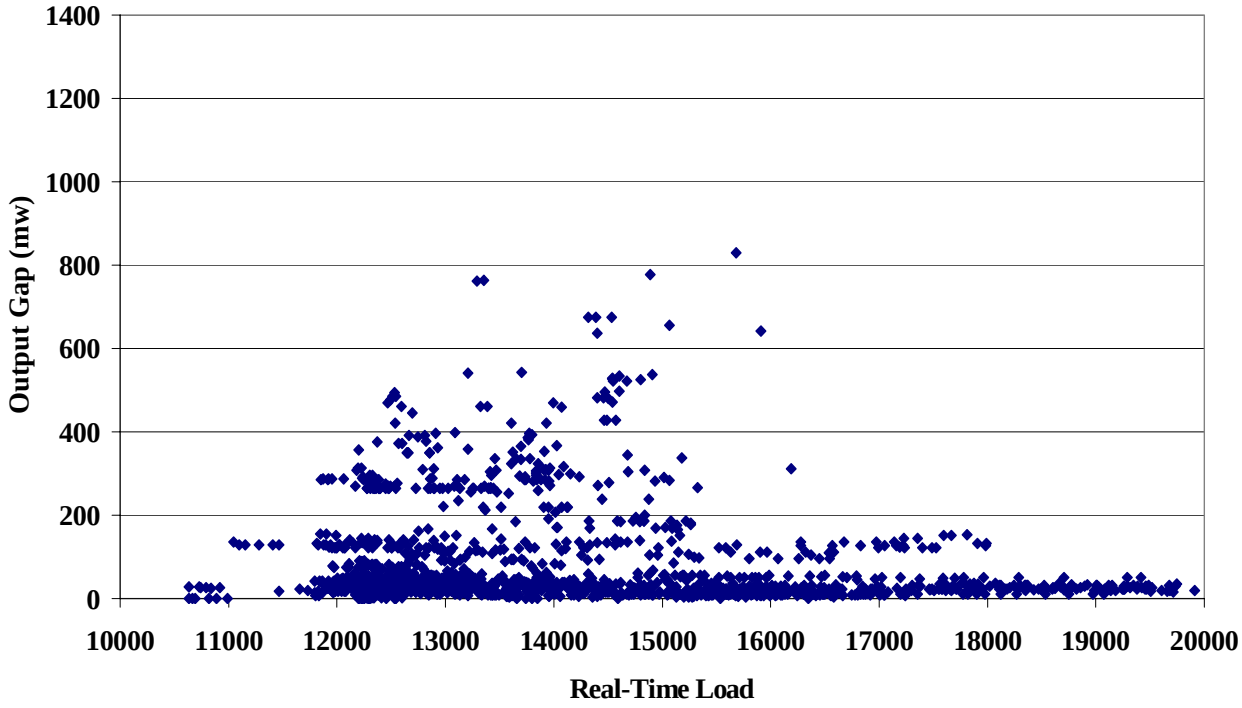
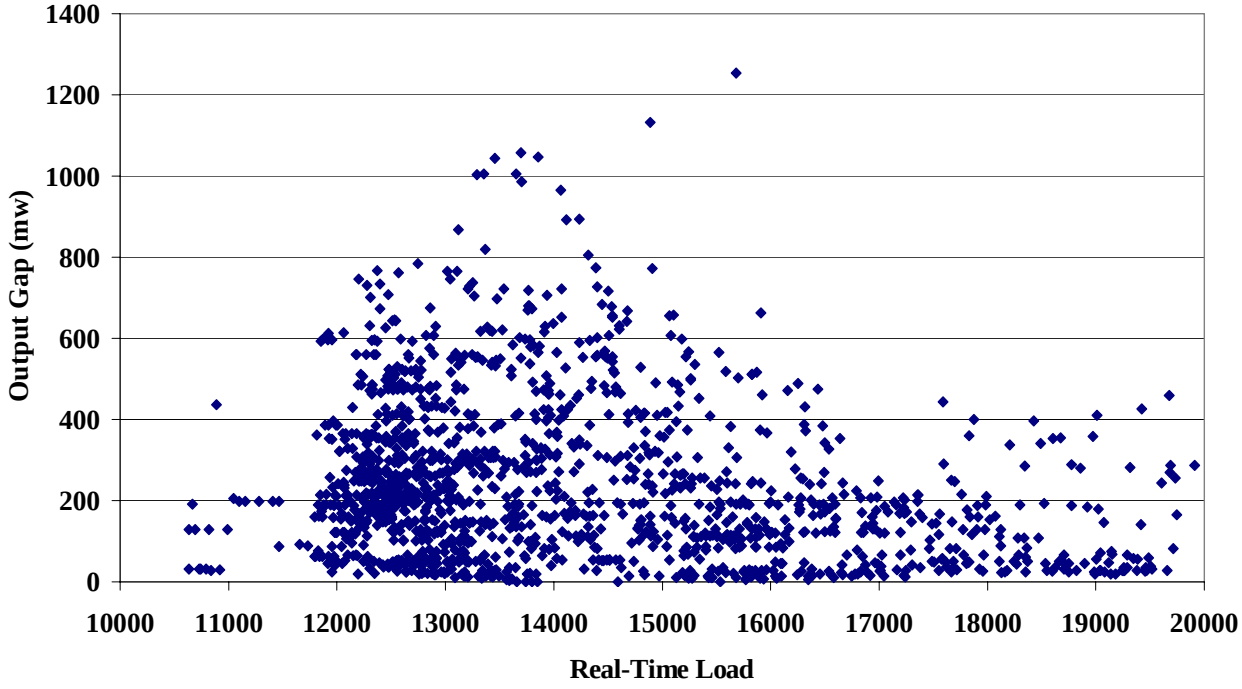


Figure 21: Relationship of Output Gap at Low Threshold to Actual Load
Real Time Market – East New York
Weekdays, Noon to 6 PM



These figures both show that output gap decreases to extremely low levels under the highest load conditions. Figure 20 shows that the output gap measured at the high threshold was less than 100 MW during all hours when Eastern New York load exceeded 18 GW, while Figure 21 indicates that the output gap measured with the lower threshold was less than 500 MW in the same high load hours. Even if all of the output gap measured at the lower threshold was actual withholding, it would still amount to a small portion of the Eastern New York load. This is an important result because prices are most vulnerable to the exercise of market power under peak load conditions. These results indicate that economic withholding was not a significant concern in 2004.

B. Analysis of Load Bidding

In addition to physical and economic withholding, buyer behavior can strategically influence energy prices. Therefore, evaluating whether load bidding is consistent with workable competition is an important focus of market monitoring. Load can be purchased in one of the following four ways:

Physical Bilateral Contracts. These are schedules that the NYISO provides to participants that allow them to settle transmission charges (i.e., congestion and losses) with the ISO and to settle on the commodity sale privately with their counterparties. It does not represent the entirety of the bilateral contracting in New York, however, because participants have the option of constructing identical arrangements by other means that would settle through the NYISO. In particular, participants may sign a "contract-for-differences" ("CFD") with a counterparty to make a bilateral purchase. Financial bilateral contracts such as CFDs are settled privately and generally would show up as day-ahead fixed load.

When the CFD is combined with a TCC, the participant can create a fully-hedged forward energy purchase. Therefore, the trends in the quantity of physical bilateral contracts scheduled with the NYISO do not indicate the full extent of forward contracting.

Day-Ahead Fixed Load. This represents load scheduled in the day-ahead market for receipt at a specific bus regardless of the day-ahead price. It is the equivalent of a load bid with an infinite bid price, which is difficult to rationalize from an economic perspective.

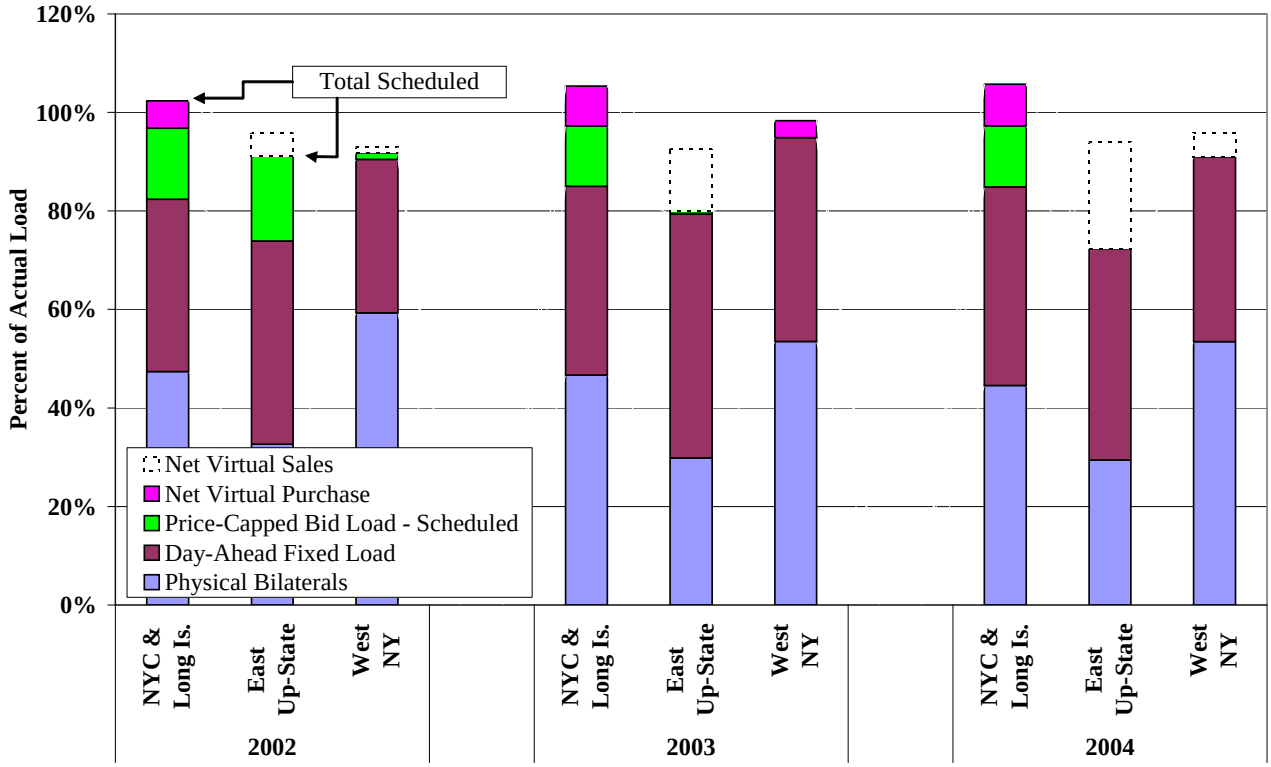
Price-Capped Load Bidding. This represents load bid into the day-ahead market with a bid price indicating the maximum amount the Load-Serving Entity ("LSE") is willing to pay. For example, an LSE may make a price-capped bid for 500 MW at $60 per MWh. If the day-ahead market at its location clears above $60, the energy would not be purchased in the day-ahead market. If the load is actually realized in real-time, it would be served with energy purchased in the real-time market. This is a more rational form of load-bidding than the non-price sensitive fixed load schedules. However, price-capped load bidding is only allowed at the zonal level while fixed load bidding is allowed at the bus level.

Net Virtual Purchases. This quantity is equal to the virtual load purchases minus the virtual supply sales. Virtual trading was introduced in the NYISO markets in November 2001.

Figure 22 shows the load that was scheduled in each of these categories during 2002 through 2004 at various locations in New York. The share of the actual load supplied through physical bilaterals has decreased slightly over the past three years, averaging just below 50 percent of the actual load for New York State. This does not mean that over 50 percent of the load is exposed

to prices in the NYISO energy markets. Physical bilateral scheduling does not include all bilaterals. Participants with financial bilaterals, such as contracts for differences ("CFDs"), settle with the NYISO at the energy price in the NYISO market and settle separately with their counterparties to achieve a total settlement consistent with their bilateral contract. Hence, these participants will appear to be settling at the NYISO prices, but incur only the bilateral contract price in reality.

Figure 22: Composition of Day-Ahead Load Schedules as a Proportion of Actual Load 2002 - 2004



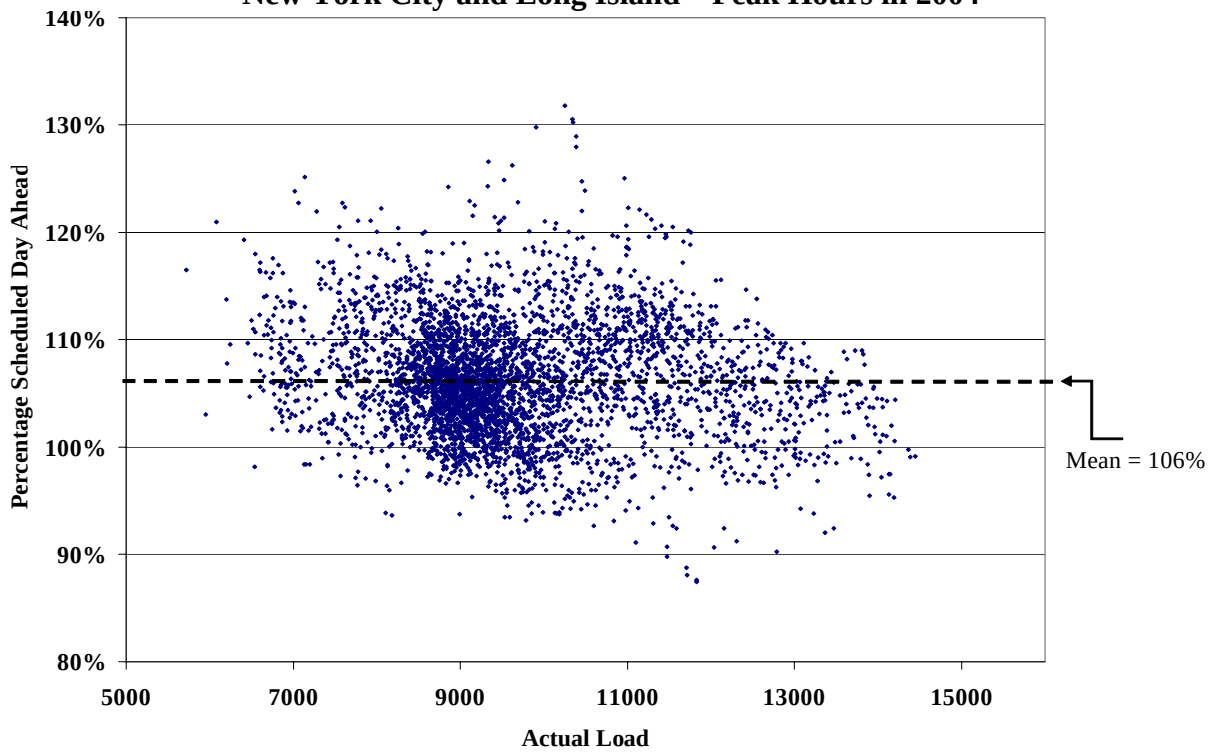
The figure also shows that load was over-scheduled in New York City and Long Island, whereas load was under-scheduled in the upstate regions. The ratio of day-ahead scheduling to actual load decreased in East upstate New York from 91 percent in 2002 to just 72 percent in 2004. Figure 22 shows that this decrease was primarily driven by a rise in net virtual sales. Net virtual sales are consistent with the incentives presented by the day-ahead price premium in that region. Thus, the lack of scheduling convergence in East upstate New York is caused by virtual trading activity that has improved price convergence. Convergence between day-ahead scheduled load and actual load worsened slightly from 2002 to 2004.in New York City and Long Island. The higher day-ahead purchases in New York City are consistent with the incentives facing virtual

traders and load-serving entities by the persistent real-time price premium in that area. As discussed in this report, these pricing and scheduling patterns are primarily the result of modeling inconsistencies between the day-ahead and real-time markets.

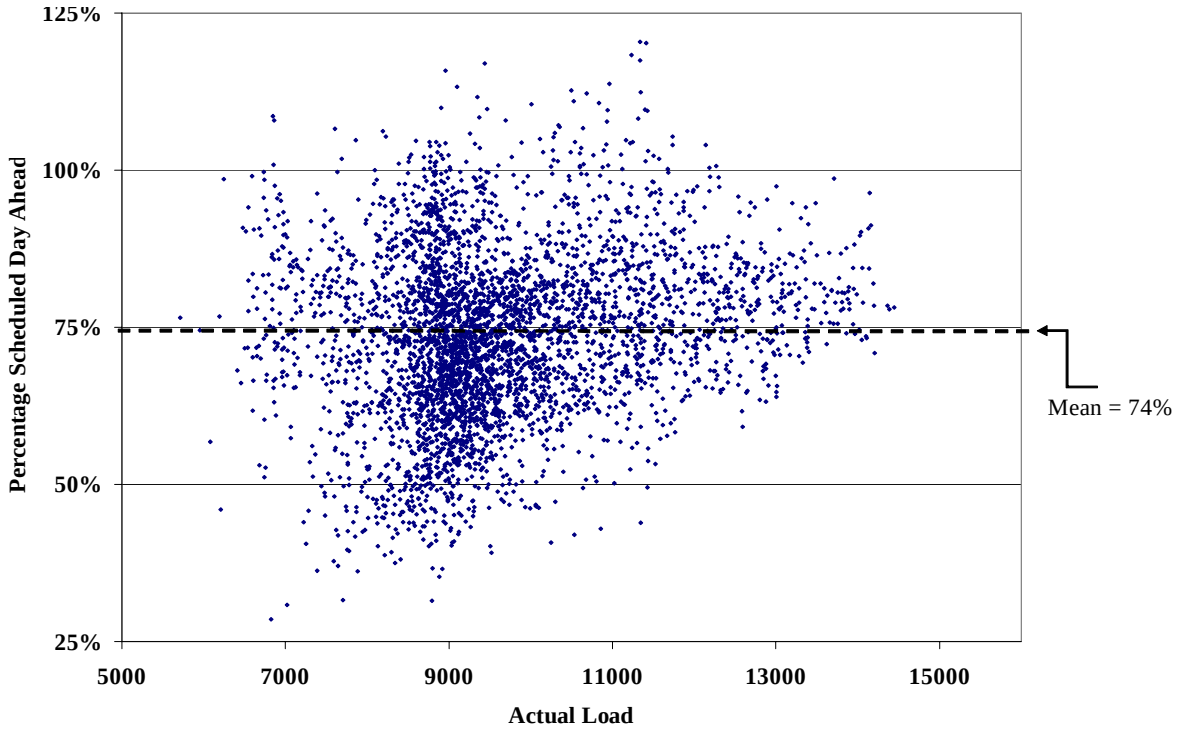
In order to further evaluate the pattern of load bidding, we calculated day-ahead hourly load schedules (including virtual load bids) as a percentage of real-time load for all peak hours during 2004. This analysis is shown in Figure 23, which includes scatter-plot diagrams for New York City, eastern New York outside of New York City, and western New York.

Consistent with the previous figure, this analysis indicates New York City and Long Island tend to over-schedule load day-ahead. However, this pattern diminishes slightly in the highest load hours. Load scheduled day-ahead in eastern upstate New York is more variable and is usually substantially under-scheduled. This under-scheduling decreases with increases in load. In Western New York, the data reveals that day-ahead load is under-scheduled on average and that this under scheduling becomes more acute as load rises.

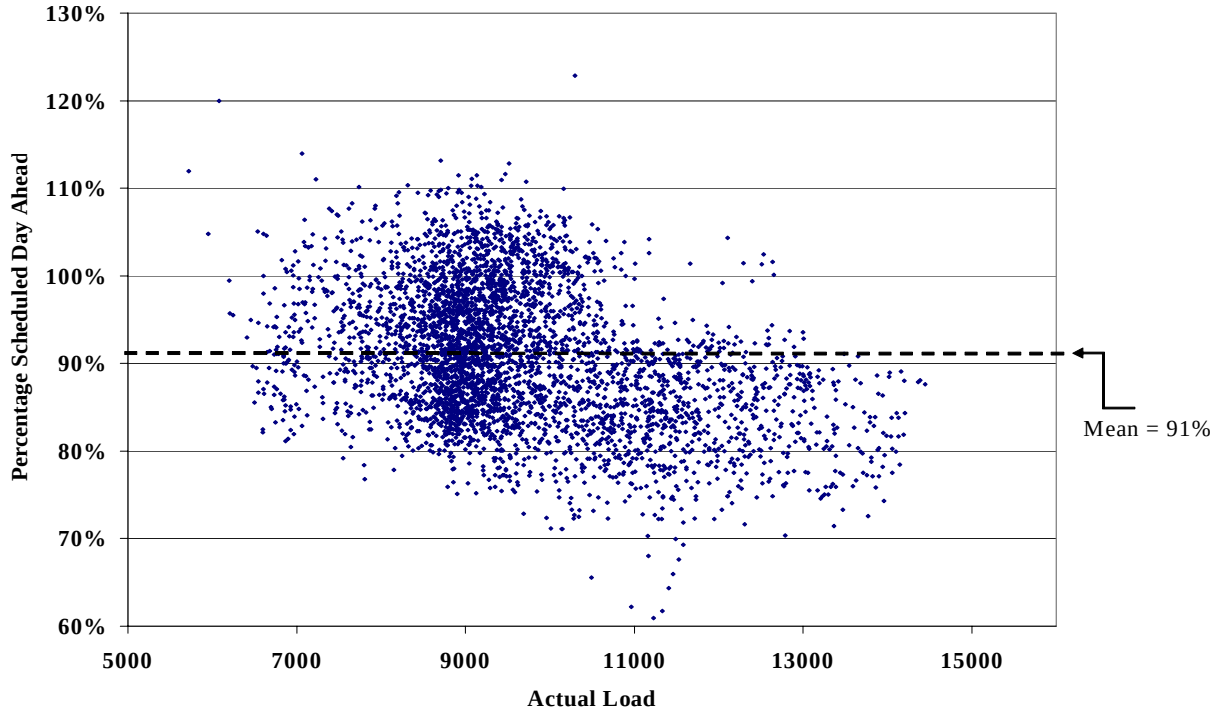
Figure 23: Load Scheduled Day-Ahead versus Real-Time Load
New York City and Long Island – Peak Hours in 2004



Eastern Upstate New York – Peak Hours in 2004



Western New York – Peak Hours in 2004



The figures in this sub-section clearly show that load was over-scheduled in New York City and Long Island, and under-scheduled in upstate New York. This pattern is consistent with systematic differences between the assumptions in the day-ahead and real-time market models. Market participants respond to these inconsistencies by rationally adjusting the purchases and sales in the day-ahead market, which tends to improve price convergence. In this case, that arbitrage results in over-scheduling within New York City and under-scheduling outside of New York City. The RTS software implemented in 2005 in real time should reduce these modeling inconsistencies. We will be reviewing the performance of the RTS systems following the summer of 2005.

1. Virtual Trading

Virtual trading was introduced in November 2001 to allow participation in the day-ahead market by entities other than LSEs and generators. The motivation was to improve arbitrage between the day-ahead and real-time markets as well as allowing flexibility for all participants in managing risk. Virtual energy sales or purchases in the day-ahead market settle in the real-time market, allowing participants to arbitrage price differences between the day-ahead and real-time markets. For example, a participant can make virtual purchases in the day-ahead market if the participant expects prices to be higher in the real-time market, and then sell the purchased energy back into the real-time market. The result of this intertemporal arbitrage would be to raise the day-ahead price slightly and decrease the real-time price slightly to improve convergence.

We analyzed the quantities of virtual load and supply that have been offered and scheduled on a monthly basis during the past two years. Figure 24 and Figure 25 show the pattern of virtual bidding in New York City and elsewhere in the State in 2004.

Figure 24: Hourly Virtual Load and Supply New York City and Long Island

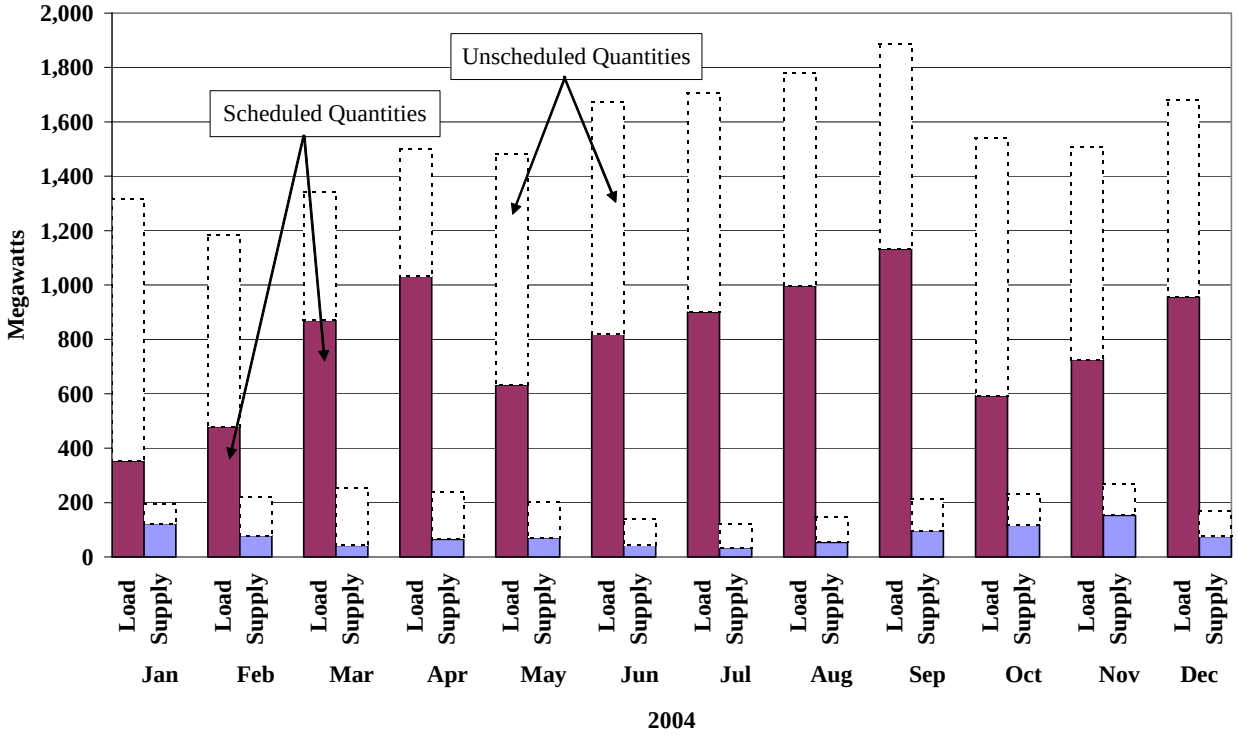
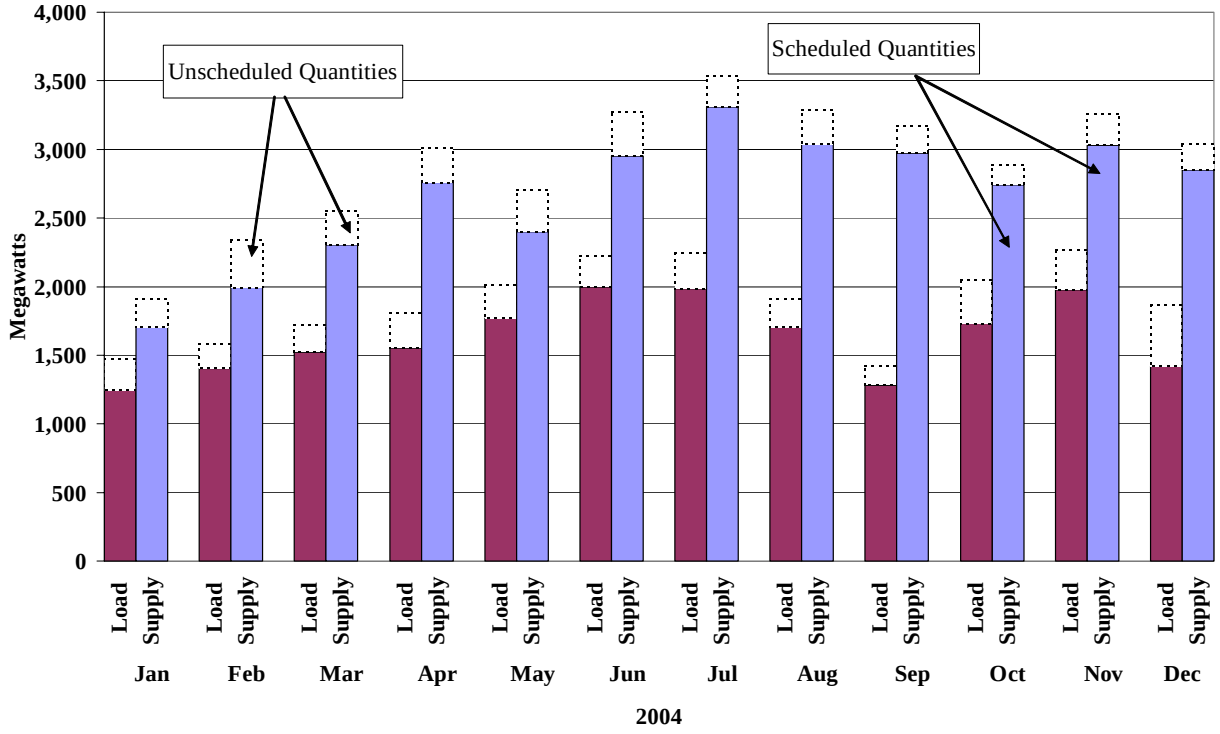


Figure 25: Hourly Virtual Load and Supply Outside New York City and Long Island



Virtual trading activity tends to be highest during the summer when real-time load is highest and prices are most volatile. Virtual supply scheduled upstate has grown substantially since the

summer of 2002, increasing to close to 2000 MW per hour by the end of 2003 and to 3000 MW per hour by the end of 2004. Growth in upstate virtual demand bidding was also substantial in 2003, but slowed considerably in 2004, settling in a range between 1500 and 2000 MW per hour. Virtual load grew much more slowly in New York City and Long Island, from an average of close to 1200 MW an hour in 2003 to 1500 MW an hour in 2004, while virtual supply remained at very low levels. 50 percent of virtual bids and offers in New York City and Long Island were scheduled, while 90 percent of virtual bids and offers in upstate New York were scheduled.

The net virtual purchases in New York City and net virtual sales upstate of New York City contribute to the overall over-scheduling in the City and under-scheduling upstate discussed in the prior section. We find that these scheduling patterns are consistent with the transmission modeling issues discussed in the next section. Importantly, these virtual trading patterns have contributed to improved convergence between the day-ahead and real-time prices.

III. MARKET OPERATIONS

Aside from operating the spot markets, a primary role of the ISO's market operations is to ensure safe and reliable grid operation. Many of the ISO's operating functions in this regard can have a substantial impact on market outcomes, especially during peak demand conditions. Operating functions that can affect the market outcomes include:

- Modeling a security-constrained transmission system in the day-ahead and real-time markets;
- Dispatching generation out-of-merit in order to resolve transmission constraints;
- Committing supplemental resources not selected by the day-ahead market;
- Dispatching reserves under peak load conditions; and
- Making real-time load curtailments and emergency out-of-market purchases.

Reliability requires that operators carry out all of these functions, but they should be done in a way that promotes efficient market pricing and behavior. This section evaluates these operating functions and examines how they impact market outcomes.

A. Transmission Congestion

Congestion can arise in both the day-ahead and real-time markets when transmission capability is not sufficient to accommodate a least-cost dispatch of generation resources. When congestion arises, both the day-ahead and real-time market software establish spot prices based on the cost of meeting load at each location, which reflects the fact that higher-cost generation may be required at locations where transmission constraints prevent the free flow of available resources. This will result in higher spot prices at these "constrained locations" than would occur in the absence of congestion. Furthermore, transmission losses greatly affect the cost of serving load in each area, and are also reflected in locational spot prices.

The day-ahead market is a forward market, facilitating financial transactions among participants that are binding in real-time. The NYISO applies congestion charges to these transactions, which are both bilateral transactions and spot transactions, by modeling anticipated congestion. Bilateral transactions are charged based on the difference between day-ahead spot prices at the

two locations (the price at the sink less the price at the source). Buyers and sellers pay congestion charges implicitly equal to the difference in prices between the locations where power is injected and withdrawn from the transmission network.

Congestion charges may be hedged in the day-ahead market by owning TCCs, which entitle the holder of the TCC to payments corresponding to the congestion charge between two locations. A TCC consists of a directional pair of points (locations or zones) and a MW value. For example, if a participant holds 150 MW of TCC rights from point A to zone B, this participant is entitled to 150 times the congestion price at zone B less the congestion price at location A. Excepting losses, a participant can perfectly hedge its bilateral contract if it owns a TCC between the same two points over which it has scheduled the bilateral contract.

In the real-time market, participants with day-ahead contracts do not pay real-time congestion charges. Only transactions that are not scheduled in the day-ahead market are assessed real-time congestion charges. As in the day-ahead market, charges for bilateral transactions are based on the difference between the locational prices at the two locations of the bilateral contract. For real-time spot market transactions, the congestion charge is paid by the purchaser through the congestion component of the LMP. There are no TCCs for real-time congestion because the real-time spot market is a balancing market where congestion charges should be zero on average.

1. Aggregated Congestion Costs

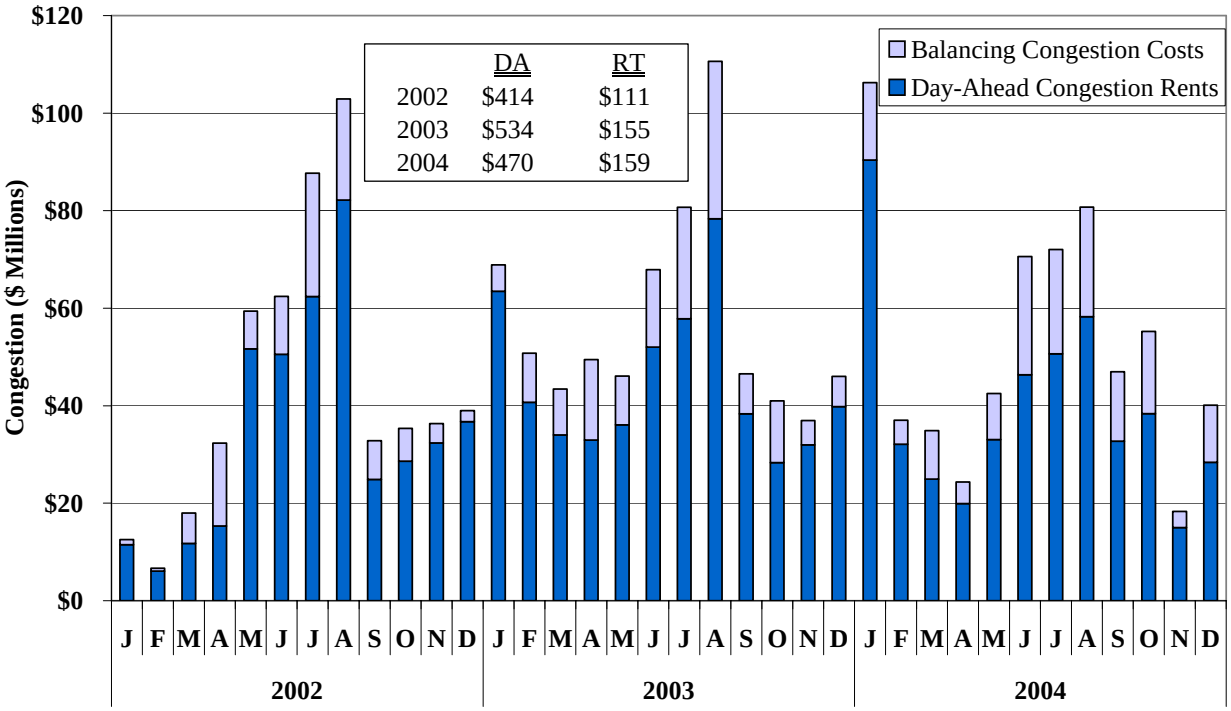
Our next analysis evaluates congestion costs in the day-ahead and real-time markets. These values are the total congestion revenues collected from participants, which include: a) the difference between the total payments by loads and the payments to generators and net imports (excluding losses), and b) the congestion costs collected from physical bilateral schedules. In an LMP system, this revenue will be equal to the marginal value of the transmission capacity (i.e., the shadow price of the transmission constraint)[8] times the amount of power flowing across the constrained interface.[9] It is important to recognize that these costs do not represent the net

[8] A shadow price is the value to the system of increasing the constraint by a very small amount (e.g., 1 MW). In this case, it would be equal to the reduction in system production costs that could be achieved by substituting lower cost resources on the unconstrained side of the interface for higher-cost resources on the constrained side of the interface.

[9] These amounts should not be expected to resemble the historical congestion values that have been

benefits of eliminating all congestion in New York, which has been estimated to be less than $100 million. Figure 26 shows the monthly congestion costs that occurred in the day-ahead and real-time markets from 2002 to 2004.

Figure 26: Monthly Congestion Expenses 2002 - 2004



Overall congestion costs increased substantially since 2001, from $310 million in 2001 to $525 million in 2002 and $688 million in 2003, before leveling off at $629 million in 2004. This increase was primarily due to the modeling of load pockets within New York City, which began in June 2002 and was one of the most significant modeling changes that have occurred in New York since the market began. This modeling improvement substantially increased the apparent congestion (because LBMPs began reflecting these constraints) and reduced uplift costs. Prior to this change, resources were re-dispatched out-of-merit and the costs were recovered through uplift charges.

calculated with methods approved by the NYISO Operating Committee in January 2004. These methods utilize a model to calculate various types of cost differences between the current system and a completely unconstrained system.

Increases in fuel prices tend to proportionately affect electricity prices. Thus, the substantial rise in natural gas prices from 2002 to 2003 contributed to the rise in congestion costs over that period. However, the effect of rising fuel prices was partly mitigated by milder summer weather in 2003 than in 2002. This is because congestion generally increases during high load periods when the transmission system is more fully utilized. The decrease in congestion costs from 2003 to 2004 is also partly due to the mildness of summer weather during 2004.

The lack of convergence between day-ahead and real-time prices has had an impact on the congestion costs shown in Figure 26. The previous sections indicated that, in 2004, day-ahead prices were higher than real-time prices in upstate New York, while day-ahead prices are lower in New York City and Long Island. Thus, the cost of congestion along the path of power flows is generally smaller in the day-ahead market than in the real-time market. This decreases the amount of revenue collected as day-ahead congestion rents relative to what it would be if price convergence were better.

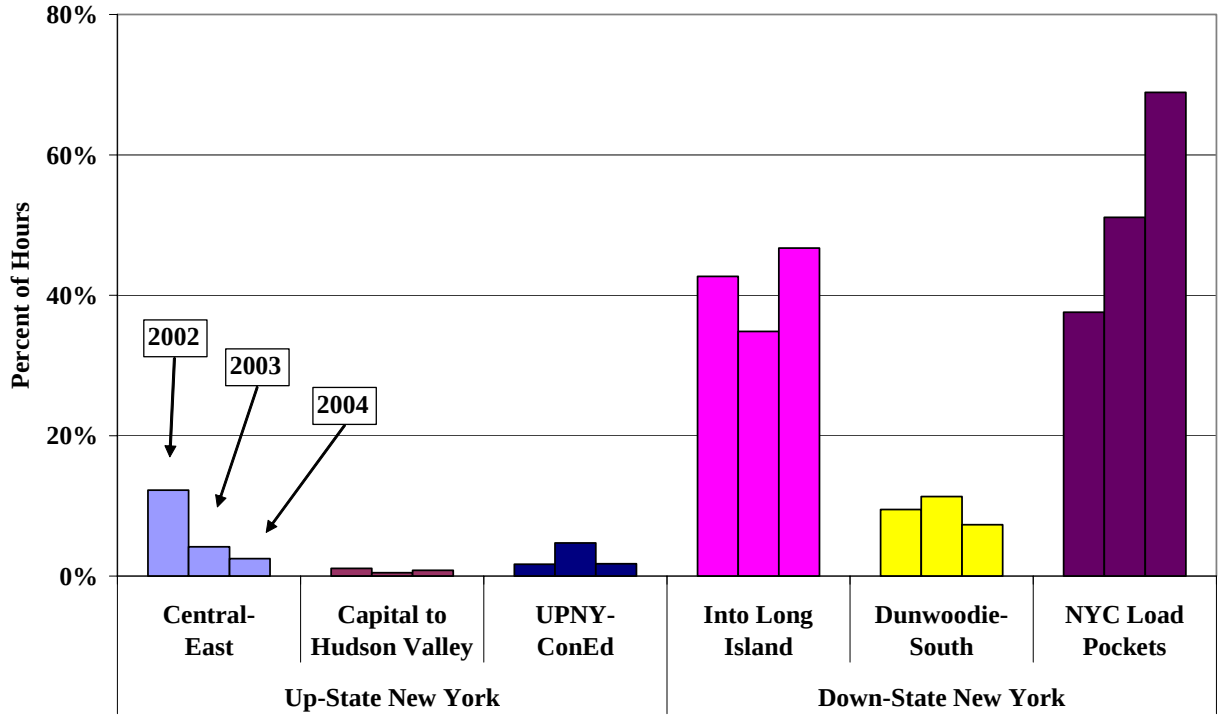
2. Major Transmission Interfaces

Supply resources in New York City and Long Island generally have higher costs than in upstate New York. The physical capability of the transmission system limits the amount of power that can be transferred from lower cost resources to load pockets in New York City and Long Island, making the economic value of major transmission interfaces considerable. Thus, is important that the transmission planning process and incentives for transmission investment lead to efficient new investment. The analyses in this sub-section summarize the value of congestion on several key interfaces in New York.

Figure 27 shows the frequency of congestion on select interfaces in upstate and downstate New York. From upstate New York, the figure includes constraints that (i) are part of the Central-East Interface, (ii) limit southward flows from the Capital region through the Hudson Valley, and (iii) make up the interface between upstate New York and the Con Ed transmission area. From downstate New York, the figure includes (i) transmission constraints from upstate New York into Long Island, (ii) the Dunwoodie-South constraint that limits flows from upstate New York into New York City, and (iii) the group of constraints that limit flows within New York City.

This analysis excludes constraints within Western New York and also within the Long Island zone.

Figure 27: Frequency of Real-Time Congestion on Major Interfaces 2002 - 2004



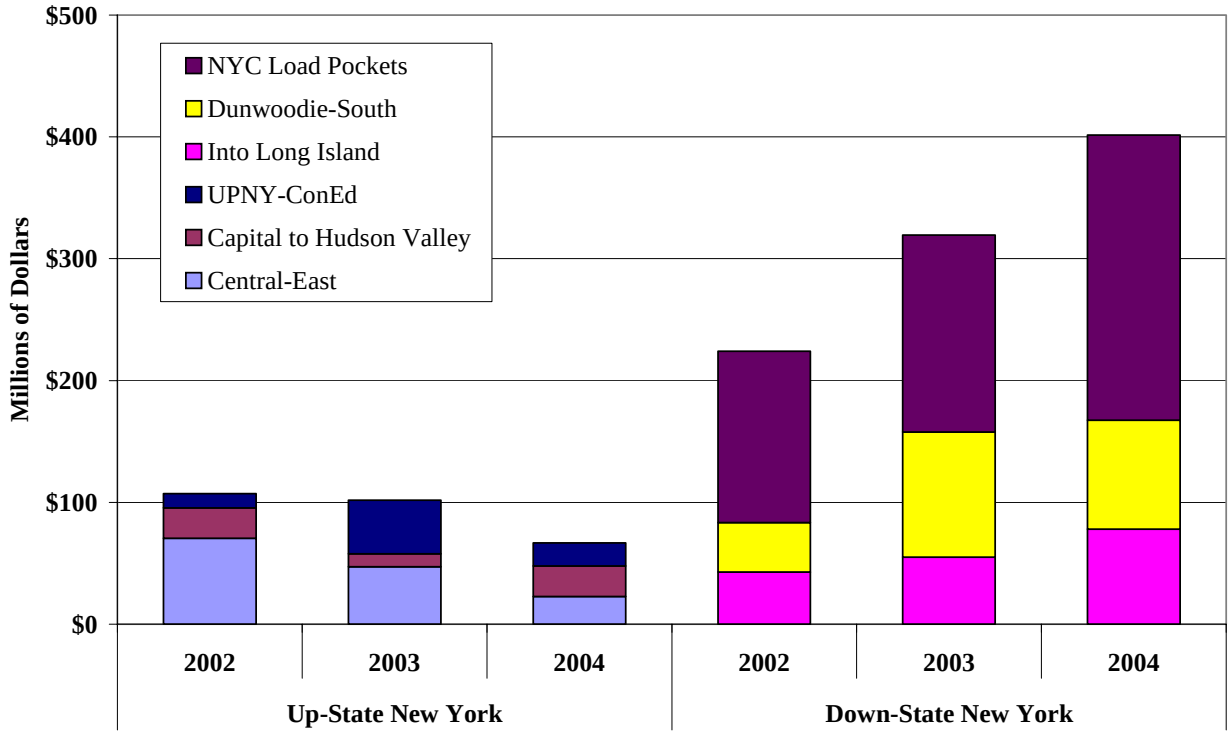
The results of Figure 27 show the preponderance of congestion occurs into and within downstate areas. Furthermore, congestion into New York City load pockets has increased substantially over the three years while congestion on the Central-East interface has grown less frequent. There are three main factors that influence the trends in congestion shown above. First, load pocket modeling was introduced to New York City in June 2002. The NYC Load Pockets were constrained during more than 60 percent of the intervals in 2002 after load pocket modeling was introduced (which is comparable to the results for 2003 and 2004).

Second, there have been significant transmission outages that have affected congestion patterns. The Central-East interface experienced large outages during the spring of 2002 which contributed to the frequency of congestion in that year. There was also significant congestion on the Dunwoodie-South interface early in 2003 as a result of maintenance work, resulting in lower import levels to New York City. Frequently the need for more New York City generation to resolve the Dunwoodie-South constraint was met with generation in load pockets, which reduced the frequency of congestion within New York City.

Third, generating capacity additions and other changes in supply have also influenced congestion patterns. The Athens plant in the Capital region began operation during 2004 along with a substantial amount of new generation in New England in 2003 and 2004, which have together reduced the flows over the Central-East interface. In addition, imports from Hydro-Quebec, have decreased substantially since 2002, reducing the loadings on the Central-East interface. Finally, more than 500 MW of peaking capacity was added to Long Island in 2002 and 2003, which contributed to the reduced congestion into Long Island.

Figure 28 measures the approximate value of congestion in real-time for the interfaces shown in the previous figure. For this analysis, the value of congestion is measured as the shadow price[10] of the interface in the real-time market multiplied by the flow.

Figure 28: Value of Real-Time Congestion on Major Interfaces 2002 - 2004



The figure shows that the introduction of load pocket modeling greatly increased the value of congestion modeled in the day-ahead and real-time markets. Furthermore, there were changes in the market in 2003 and 2004 that helped shift congestion from upstate to downstate areas. For

10 See note 8, above for a definition of shadow prices.

instance, the Athens plant has generally helped reduce congestion on the Central-East interface, but increased congestion slightly on the interfaces into the Hudson Valley and New York City. In 2004, the value of the upstate transmission interfaces was approximately $70 million, while the value of the downstate interfaces totaled $400 million.[11] This suggests that the most likely areas of the system for economic upgrades are in the downstate areas. However, a thorough analysis of costs and benefits of individual investment options would be necessary to determine whether and where transmission investment would be cost-effective.

3. Transmission Congestion and TCCs

This sub-section evaluates congestion levels in the day-ahead and real-time markets relative to the outcomes in the Transmission Congestion Contract ("TCC") market. Market participants that purchase TCCs are hedged against unexpectedly high congestion costs in the day-ahead.[12] In a well-function system, the value of congestion revealed in the TCC market, in the day-ahead market, and in the real-time market should converge. In this section, we examine several indicators of convergence between these markets:

- Day-ahead Congestion Revenue Shortfalls: Revenues collected by the NYISO from congestion in the day-ahead market compared with payments by the NYISO to the holders of TCCs;
- Balancing Congestion Revenue Shortfalls: Congestion revenues collected from buyers in the real-time market are not sufficient to cover congestion payments by the NYISO to sellers;
- Price Convergence Between TCCs and Day-ahead Market: Prices paid for TCCs should be comparable to congestion prices in the day-ahead market that determine payments to TCC holders.

The NYISO conducts auctions to sell the TCCs to market participants. In order to determine the maximum quantity of TCCs that can be sold in a TCC Auction, the transmission system must be modeled to ensure that the TCCs are simultaneously feasible. The NYISO uses a power flow model that includes an assumed configuration of the transmission system. The simultaneous feasibility condition requires that the TCCs awarded be simultaneously are feasible in a

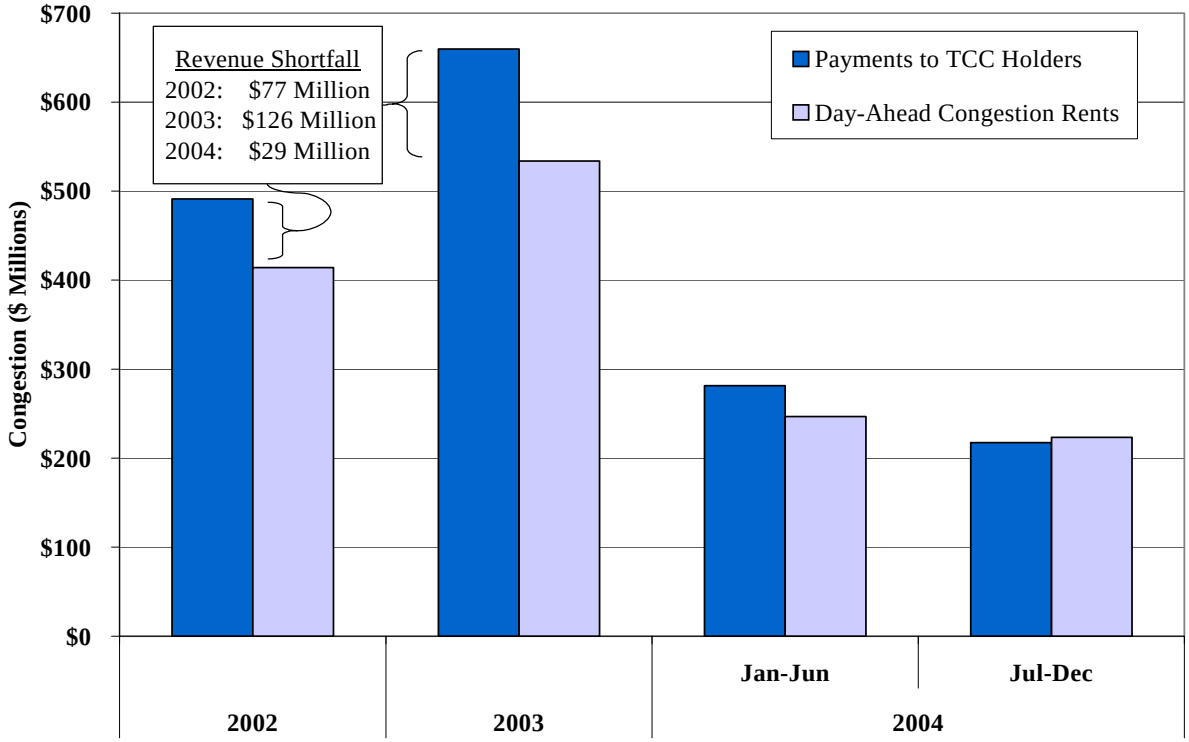
[11] These totals will not equal the total congestion costs in 2004 because it values congestion based only on the real-time market results and does not include all transmission interfaces and facilities.

[12] The NYISO administers both longer-term forward TCC Auctions, in which 6-month and 1-year TCCs are sold, and monthly Reconfiguration Auctions to allow participants to buy and sell shorter duration TCCs.

contingency constrained economic dispatch of the NYISO transmission system. If this condition is satisfied, the congestion rents collected should be sufficient to fully fund all TCCs.

If transmission outages occur that were not modeled in the TCC auction, then the TCCs may not be feasible and, thus, the congestion rents may be insufficient to meet the TCC obligations. To fully fund TCCs under these conditions, the congestion rent shortfall is charged to transmission owners and passed through to final customers through the transmission owners' service charge. Because these charges are "socialized," they do not provide efficient incentives to minimize the congestion effects of transmission outages. To evaluate the shortfall amounts over the past three years, Figure 29 shows day-ahead congestion costs and TCC payments.

Figure 29: Day-Ahead Congestion Costs and TCC Payments 2002-2004



The figure shows that congestion revenues were substantially lower than payments to TCC holders until mid-way through 2004. This occurred because the transmission capability assumed in the TCC auction generally exceeded the capability available in the day-ahead market. The pattern of consistent congestion revenue shortfalls was eliminated in 2004 when the NYISO took several actions. First, a large share of the shortfall was due to excess TCCs mistakenly sold from upstate New York to New York City. The excess TCCs were repurchased in July 2004.

Second, on December 15, 2003, the FERC approved NYISO's proposal to employ cost-causation principles in assigning responsibility for TCC revenue shortfalls and surpluses to transmission owners. The NYISO now assesses shortfall costs resulting from maintenance to individual transmission owners. This encourages transmission owners ("TOs") to schedule outages in a manner that minimizes their market impact.

Third, the NYISO also implemented two mechanisms to reduce congestion rent shortfalls by allowing TOs to retain transmission capacity by converting up to 5 percent of transmission capacity into six-month TCCs, which would not be available in TCC Auctions. The first mechanism would permit TOs that hold Existing Transmission Capacity for Native Load ("ETCNL")[13] to reserve a limited amount of this capacity. Under the second mechanism, all TOs would be permitted to reserve a limited portion of the residual transmission capacity[14] between contiguous pairs of load zones. Congestion payments for the reserved TCCs will help to offset the TOs' share of a Congestion Rent Shortfall. The FERC approved these measures, subject to minor changes, effective February 2, 2004. The figure shows that together, these provisions eliminated the congestion revenue shortfalls after June 2004.

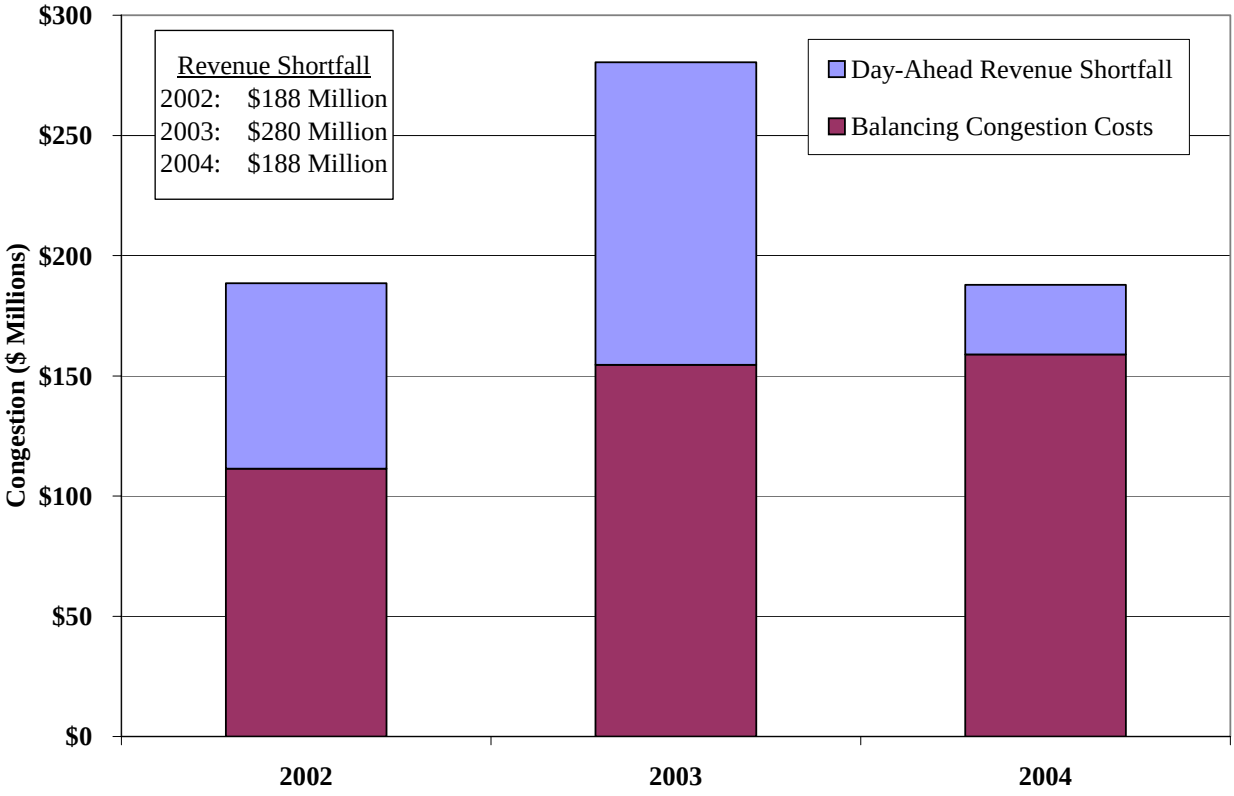
The next analysis summarizes the additional congestion revenue shortfalls incurred in the real-time market (balancing congestion costs). One cause of balancing congestion costs are reductions in transmission limits between the day-ahead and real-time markets. In this case, the ISO must purchase additional generation in the constrained area and sell back energy in the unconstrained area (i.e., purchase counter-flow to offset the day-ahead schedule). The cost of this re-dispatch is collected from loads through uplift charges. In addition, differences in the modeling of transmission losses can result in balancing congestion costs. If transmission capability and losses assumed in the day-ahead market are generally comparable to the physical characteristic in real-time, the magnitude and direction of these counter-flows should be distributed randomly and should sum to zero over time. However, as Figure 30 shows, the

13 TOs were allocated ETCNLs to facilitate the transition to locational marginal pricing.

14 Once ETCNLs and grandfathered transmission rights are accounted for, the NYISO sells any remaining transmission capacity as Residual TCCs.

balancing congestion costs have been positive and increasing over time, while day-ahead shortfalls were largely addressed in 2004.

Figure 30: Day-Ahead Congestion Revenue Shortfalls and Real Time Congestion 2002 - 2004

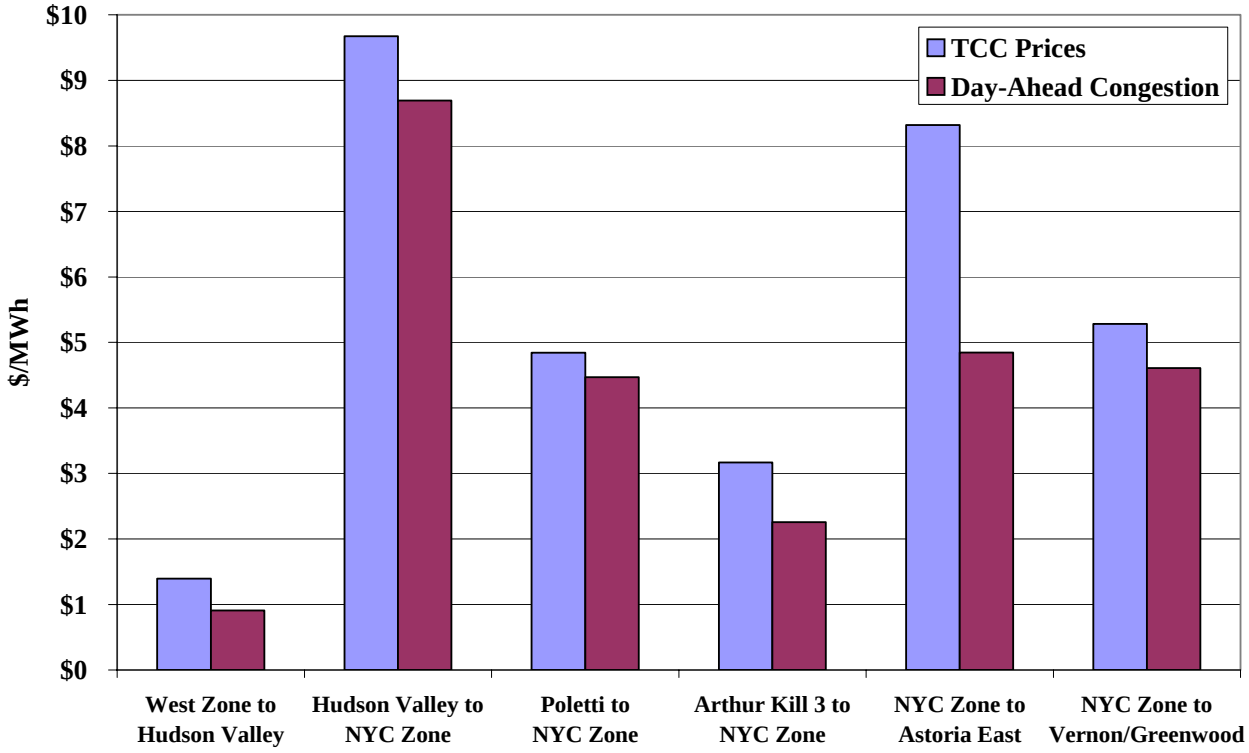


The NYISO implemented its new RTS software in February 2005, which includes considerable improvements over the previous software. The RTS market model is similar to the day-ahead market model, which should improve the consistency of the assumptions and results of the two markets.

Our final analysis in this area is designed to evaluate whether the TCC prices that have emerged from the NYISO's markets converge with the outcomes in the day-ahead energy market. TCCs provide an entitlement to the holder for the day-ahead congestion between two points. Hence, in a well-functioning market, the price for the TCC should reflect a reasonable expectation of the day-ahead congestion.

To evaluate this, Figure 31 compares the auction prices from the auction of 6-month TCCs during the summer capability period for 2004 to the day-ahead congestion that actually occurred during the period.

Figure 31: TCC Prices and Day-Ahead Congestion
May to October 2004

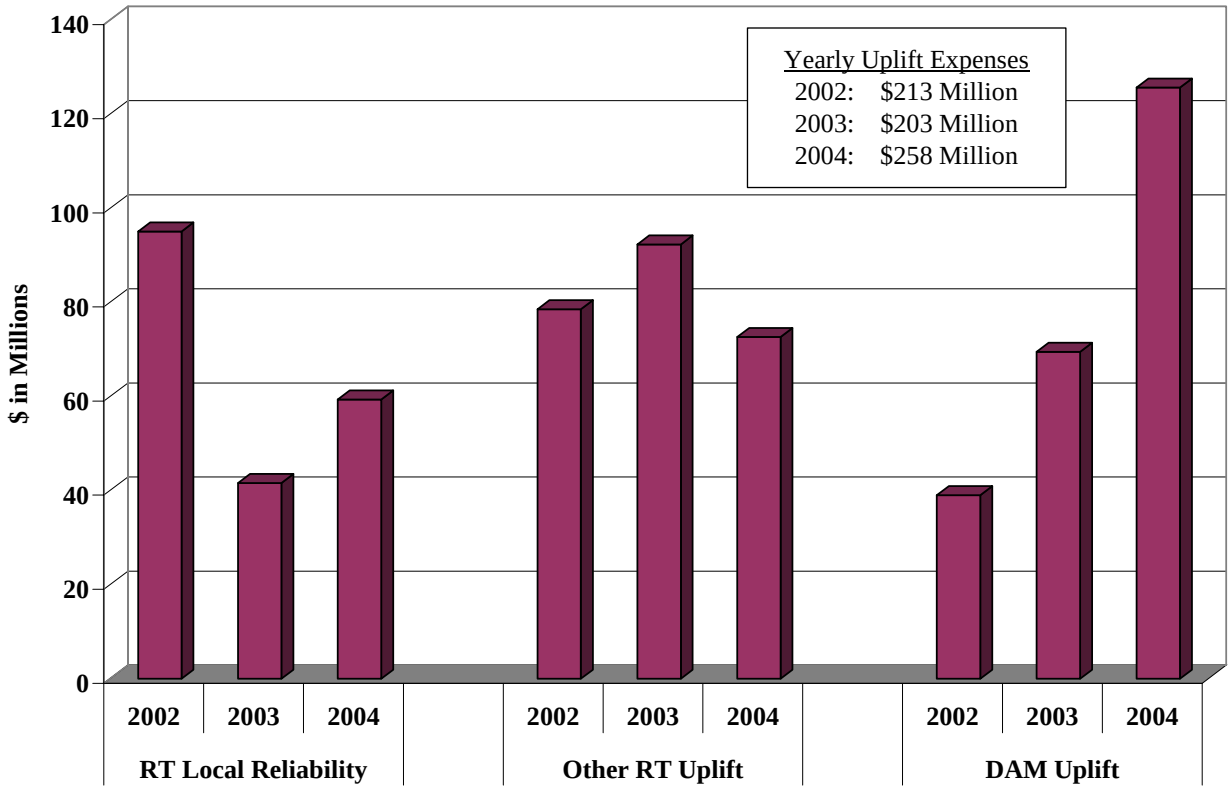


The results of this analysis show that the TCC prices have reflected the value of the day-ahead congestion relatively accurately, with TCC prices slightly exceeding actual congestion. This premium is not unexpected, since a TCC is essentially an insurance policy against unexpected congestion, and risk-averse purchasers will pay a premium to avoid risk. Furthermore, actual congestion was lower than expected due to unseasonably mild conditions that prevailed during the summer of 2004, which generally reduced congestion. The largest difference between congestion in the TCC market and the day-ahead market that is shown in the figure above was related to the Astoria East load pocket, where the value of the TCC in the six-month auction exceeded day-ahead congestion by approximately 70 percent. This seems to be related to the significant real-time price premium in Astoria East which was discussed in Section I of this report.

B. Uplift and Out-of-Merit Commitment/Dispatch

In this section of the report, we evaluate patterns of uplift and out-of-merit actions that occurred in 2004. This evaluation is an important component of our overall assessment of the performance of the NYISO's markets because it indicates the extent to which the markets satisfy New York's operational requirements. The first analysis presented in Figure 32 shows the trends in uplift costs over the past three years.

Figure 32: Day-Ahead and Real-Time Uplift Expenses 2002 - 2004



The incidence of uplift for real-time local reliability fell sharply after 2001 due to the introduction of load pocket modeling, which reduced the need for out-of-merit dispatch. Reduced uplift for out-of-merit generation to manage congestion in the New York City load pockets is now reflected in the congestion component of the spot market price. The high 2002 uplift costs for real-time local reliability were largely due to out-of-merit dispatch early in the year before load pocket modeling. Changes to the BME in 2002 to more accurately schedule

units and imports for the real-time market helped reduce other uplift associated with the real-time market, but the effect of these changes was partly offset by higher fuel costs.

Day-ahead market uplift has tripled since 2002. This is uplift paid to units committed by SCUC that do not recoup their as-bid costs from the day-ahead clearing prices. This category of uplift largely stems from the local reliability pass of SCUC, which commits generators out-of-merit in New York City to protect against second contingencies. These supplemental commitments by SCUC have a tendency to decrease day-ahead prices. As a result of lower prices, large amounts of DAM uplift are paid to generators committed before the local reliability pass in the form of Bid Production Cost Guarantees. Only uplift paid to units committed in the local reliability pass is allocated to the local area, while the majority of DAM uplift is assessed market-wide.

There are several factors that help explain the substantial increase in day-ahead market uplift from 2002 to 2004. First, there has been a modest rise in the quantity of local reliability commitments, particularly in 2004. Second, the increase in fuel prices after 2002 has led to higher production costs for units receiving uplift payments. Third, the real-time price premium in New York City indicates that the day-ahead price is slightly understated. Lower day-ahead prices tend to increase the amount of production costs that must be recovered through uplift payments.

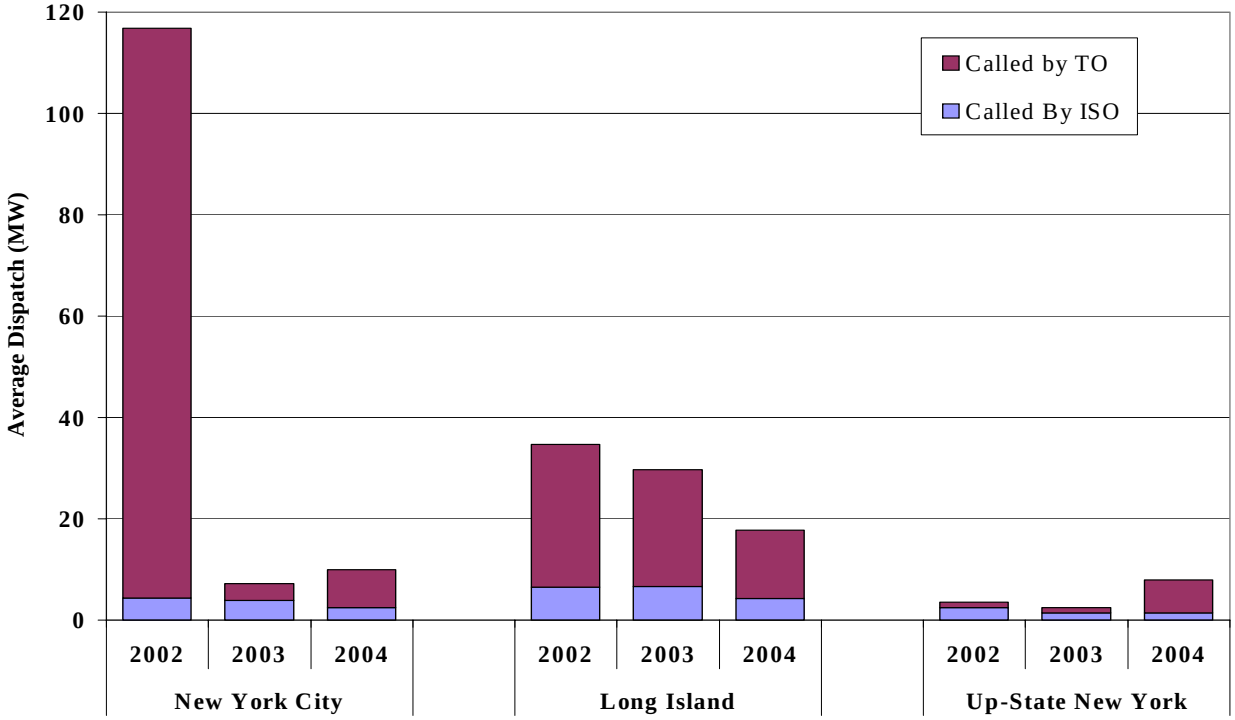
1. Real-Time Out-of-Merit Dispatch

A resource is out-of-merit ("OOM") when it is dispatched by the ISO even though its energy offer exceeds the price at its location. This can be caused by the physical parameters of the unit (e.g., minimum run-time that requires the unit to run after it has become uneconomic) or by operator action. OOM actions are generally taken to ensure reliability and resolve congestion. Actions to ensure reliability in the day-ahead market to ensure enough capacity is committed for the real-time market results in OOM commitment, as discussed in the next subsection. OOM dispatch in real-time can also be used to manage network constraints that are not included in the model.

OOM actions tend to depress spot market prices, particularly during peak demand conditions when prices are most sensitive to small changes in the quantity of load or supply. This is because OOM units are ineligible to set prices and when they are added to the supply stack, the

result is to supplant higher-offer units on the margin and depress prices, causing a divergence between the spot price and the actual marginal cost of meeting load.[15] The use of OOM units to maintain reliability also creates a need to make supplemental payments to the OOM units because the spot price is not sufficient to pay the OOM units' offer costs. The costs of these payments are recovered through uplift charges. Figure 33 shows the average OOM dispatch quantities from 2002 to 2004.

Figure 33: Average Out-of-Merit Dispatch Quantities 2002 - 2004



Note: August 2003 blackout hours excluded.

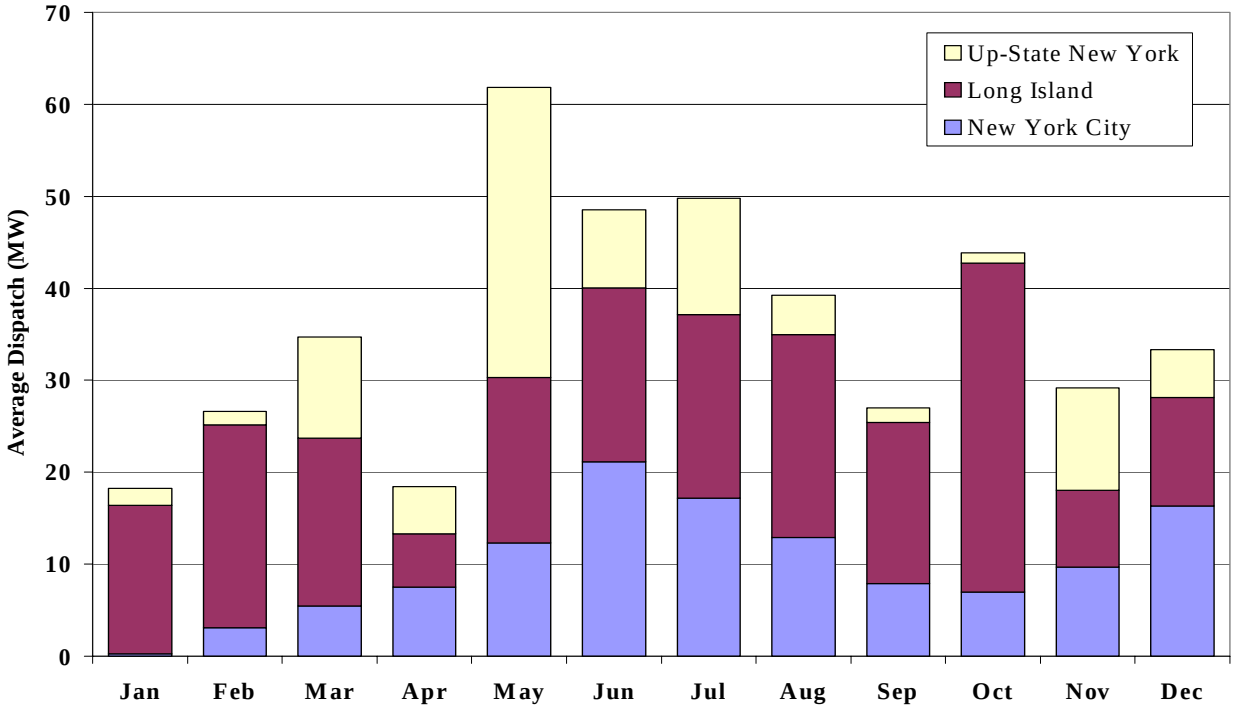
Prior to changes in the modeling load pockets in New York City, OOM dispatch in New York City accounted for approximately 80 percent of all resources dispatched OOM in the real-time market. OOM quantities fell by more than two-thirds from 2002 to 2003, primarily due to the introduction of load pocket modeling and improvements in the commitment of gas turbines in the real-time market. Because this demand for OOM dispatch has been substantially eliminated,

15 While OOM resources are ineligible to set energy prices, in many cases these resources turn out to be economic (i.e., in merit). Only units that are economically OOM will affect prices.

Long Island units now account for about half of all OOM dispatches. OOM dispatch quantities are generally very low across the state.

Figure 34 shows the incidence of OOM dispatch by month for New York City, Long Island, and the rest of the state. This figure shows that although the OOM dispatch levels were highest during the summer months, the average quantity of OOM dispatched was less than 65 MW.

Figure 34: Average Out-of-Merit Dispatch Quantities 2004



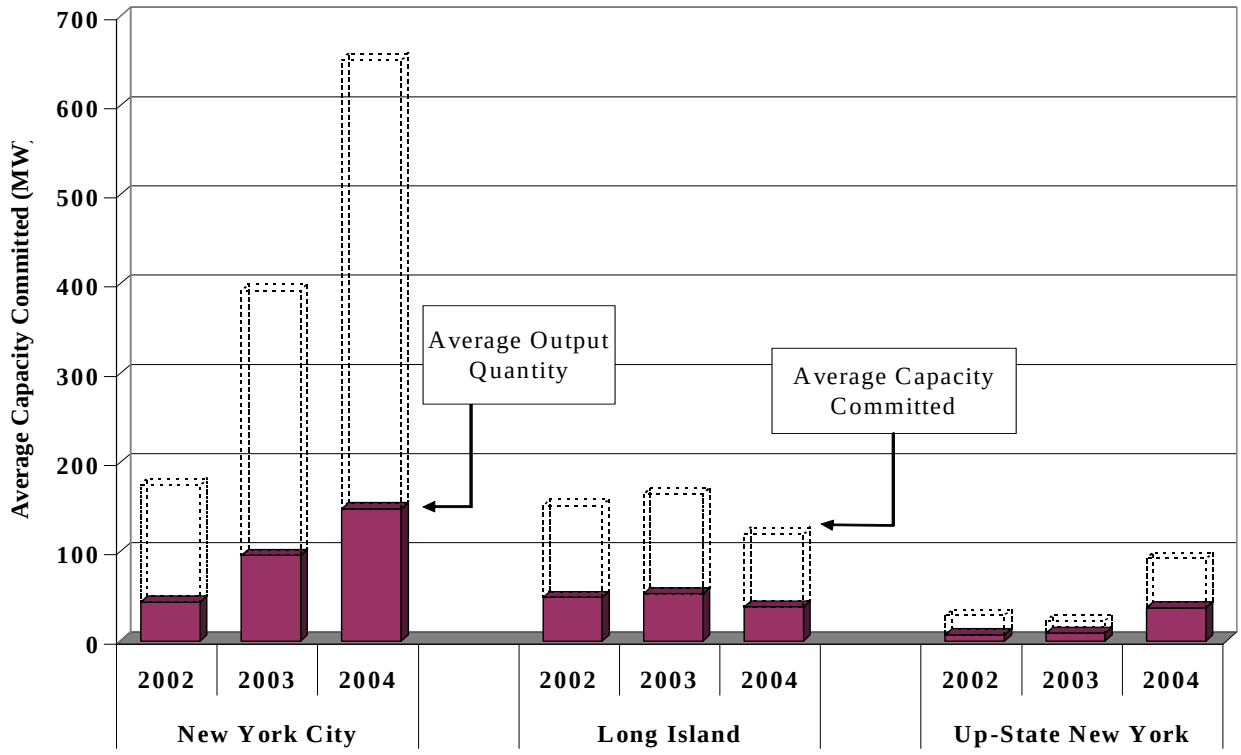
2. OOM Commitment

There are two types of OOM commitment: local reliability commitment by the day-ahead model and the Supplemental Resource Evaluation ("SRE") commitment. Day-ahead local reliability commitment is a form of out-of-merit commitment that takes place during the day-ahead market process, as opposed to the SRE that occurs after the day-ahead market closes. The day-ahead local reliability commitment is an element of the SCUC market process whereby some units that are not committed economically may be committed to meet certain specific reliability requirements, particularly second-contingency requirements in New York City.

The SRE is a process by which the ISO commits additional resources after the day-ahead market closes in order to meet reliability requirements. This may occur when the day-ahead market assumptions are modified after the market has closed (e.g., operators expect loads to be higher than the day-ahead forecast). SRE commitment may also be necessary whenever the ISO needs to manage a reliability requirement not included in the day-ahead model.

Our first analysis in this section is of the SRE commitments. Figure 35 shows the quantity of SRE commitments made from 2002 to 2004 in New York City, Long Island, and upstate New York.

Figure 35: Supplemental Resource Evaluation 2002- 2004



When the operators undertake SRE commitments these actions are logged and reported on the NYISO website. Such supplemental commitments do not directly affect the day-ahead prices, but instead make additional resources available in real-time, and, therefore, may reduce real-time prices as a result of additional units operating at their minimum generation levels. Because of the potential for price distortion as a result of these actions, it is important to evaluate the SRE process and its impact.

Figure 35 shows that most of the SRE commitments occur in New York City and on Long Island. Improvements in day-ahead modeling and commitment have reduced the quantity of SRE actions outside of New York City and Long Island since 2001, but the average quantity of capacity committed through SRE in New York City has increased three-fold since 2002.

One reason for the SREs in New York City is nitrous oxides (NOx) emission limits that require certain baseload units to turn-on in order for gas turbines to operate. The SRE commitments in the City are generally made to satisfy the generators' NOx requirements, which restrict the average emissions (per MWh of output) from a generator's portfolio. Because gas turbines emit NOx at a much higher rate per MWh generated, each supplier must have a steam unit committed to provide the capability to dispatch the gas turbines, if necessary (to keep average emissions below permitted levels). Hence, certain steam units in the City are committed through the SRE process when they are not committed by SCUC. More frequent SRE actions were required in 2003 and 2004 to meet the NOx requirements due to lower day-ahead market commitments. Since SREs are ordinarily called by individual transmission operators, the uplift associated with them constitutes a large share of RT Local Reliability Uplift, and is allocated to the local area.

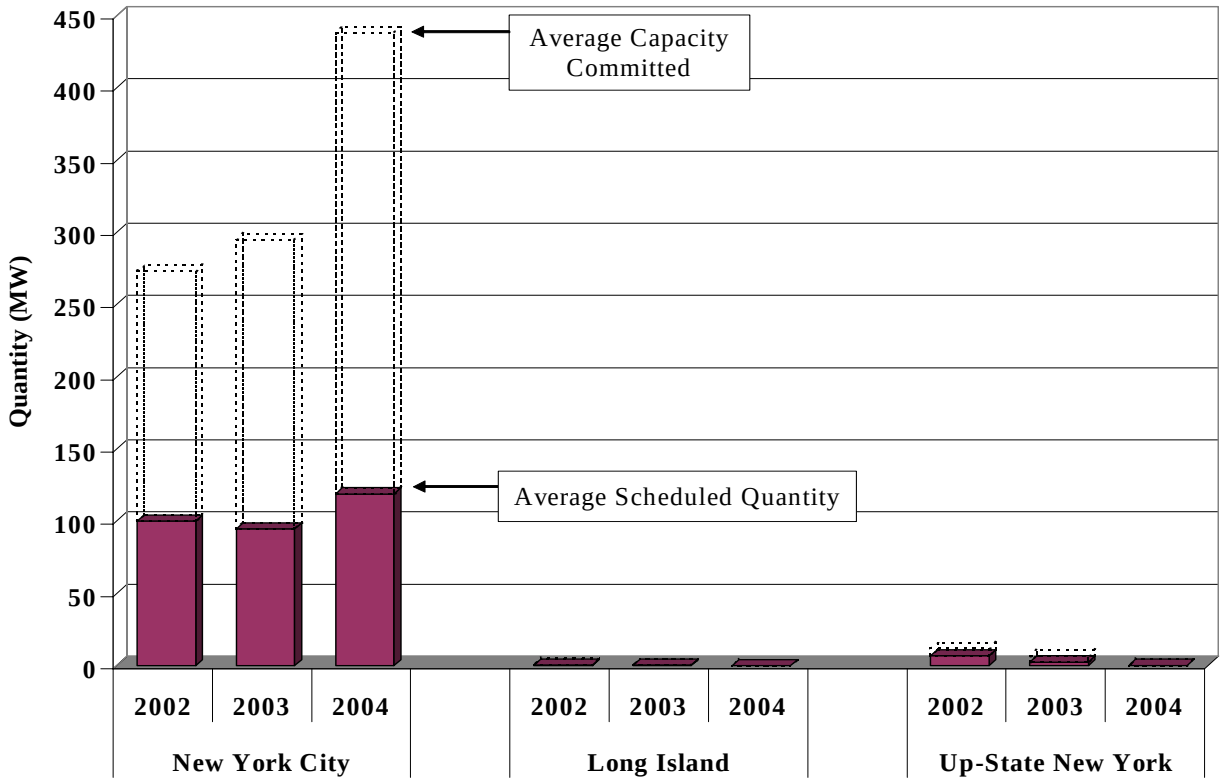
Figure 35 also shows that most of the units committed through the SRE process are dispatched at close to their minimum generation levels (i.e., 25 to 35 percent of the maximum capacity). Hence, although more than 400 MW of capacity is committed in the City, only a little more than 100 MW of additional energy is produced due to these commitments on average. This reduces the impact of these commitments on the NYISO energy markets.

The next analysis focuses on commitments made in the day-ahead market (i.e., by SCUC) to meet local reliability requirements. Figure 36 shows the average capacity committed in the day-ahead market for local reliability and the day-ahead scheduled quantity. These commitments tend to reduce prices from levels that would result from a purely economic dispatch; and can increase uplift incurred to make guarantee payments to other generators that will not cover their as-bid costs at the reduced price levels.

The average capacity committed for local reliability was approximately 440 MW in 2004, which is a 50 percent increase from 2003. Virtually all of the local reliability commitments made by SCUC involved three units in New York City. These units received average day-ahead

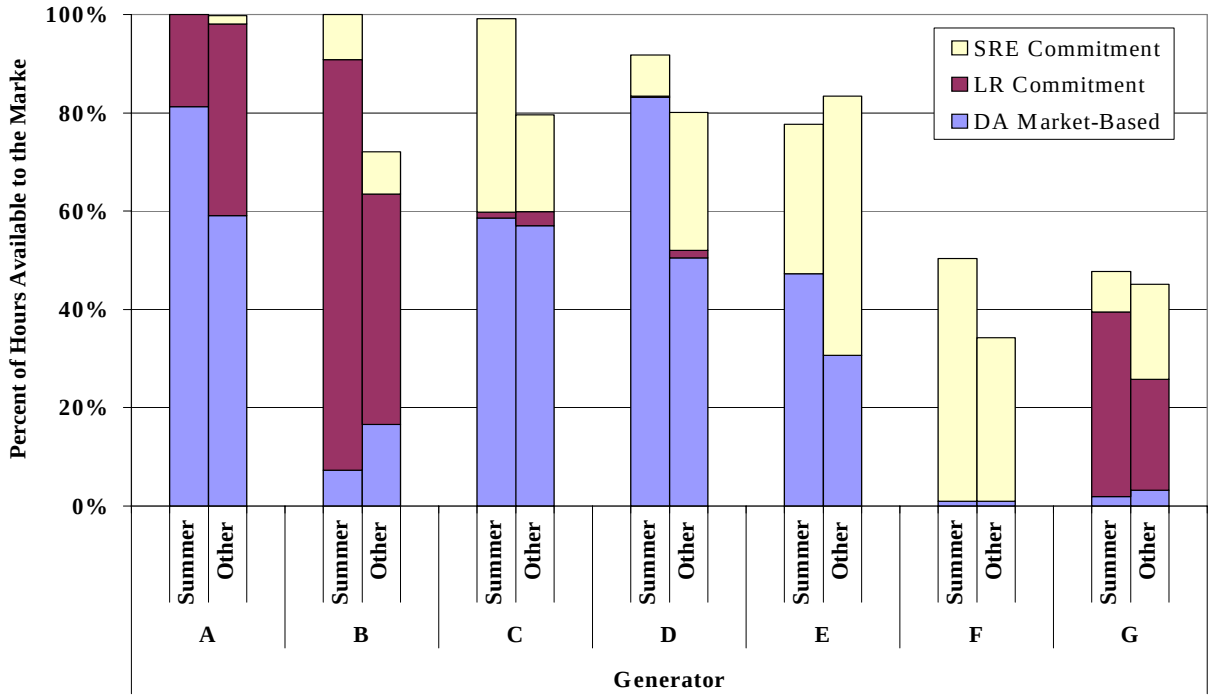
schedules of nearly 120 MW, indicating they are generally scheduled at their minimum generation level. This is the quantity of energy that will affect the day-ahead prices.

Figure 36: SCUC Local Reliability Pass Commitment
June 2002 – December 2004



In our final analysis of the OOM commitment, we evaluate the frequency of these commitments at the individual unit level. Figure 37 shows seven units that were frequently committed for local reliability by the day-ahead model or through the SRE process. The values shown are the hours that each unit is committed as a percent of the hours that the unit is available (i.e., not on outage) in summer (June to August) and non-summer days. The units in the figure accounted for more than 52% of the SREs and 93% of local reliability commitments by SCUC. Five of these units are in NYC and two are on Long Island. Three of these units analyzed were needed almost every day in the summer.

Figure 37: Units Most Frequently Committed for Local Reliability and SRE 2004



Note: DA Market Based included periods when the unit is committed economically in the day-ahead market.

The figure above shows that these seven units were committed for local reliability in a large share of the hours when they were not otherwise committed economically. For instance, during the summer, Generator B was committed economically in the day-ahead market in less than 10 percent of hours. However, this generator was committed in 90 percent of the remaining hours by the local reliability pass of the day-ahead market model. Furthermore, it was committed through the SRE process in all of the remaining hours. Based on this analysis, it seems clear that certain generators are predictably needed for local reliability, and will be committed for reliability whenever they are not economically committed.

It would be more efficient for these units to be committed before or within the economic pass of day-ahead market model. Committing additional units after the economic pass of the day-ahead market model causes some economically committed units to no longer be economic. This leads to excess capacity, depressed clearing prices, and additional uplift.

3. Out-of-Merit Dispatch and Commitment -- Conclusions

Out-of-merit dispatch and commitment have significant market effects. Primarily, they inefficiently reduce prices in both the day-ahead market and real-time market. When this occurs in a constrained area, it will inefficiently dampen the apparent congestion into the area. OOM commitments also may increase uplift payments as units committed economically will be less likely to recover their full bid production costs in the spot market.

SRE commitments are generally made to satisfy certain reliability requirements. Supplemental commitments have a number of significant market effects:

- Inefficiently reducing prices in the day-ahead and real-time markets;
- When they occur in a constrained area, they will inefficiently dampen the apparent congestion into the area; and
- Increasing uplift as units committed economically will be less likely to recover their full offer production costs.

Local reliability commitments increased in 2004 because the resources needed in New York City were committed less frequently on an economic basis. In the long-run, it would be superior to include local reliability constraints into the initial economic commitment pass of SCUC. In the short-run, we recommend that the ISO consider the feasibility and benefits of allowing operators to pre-commit units needed for NOx compliance and other predictable conditions.

C. Market Operations under Shortage Conditions

When the system is in shortage (that is, when available capacity is not sufficient to meet both energy and reserve requirements), the ISO may take a number of operating actions to satisfy its operating requirements. First, the NYISO can ask for load reductions from SCR and EDRP resources. EDRP loads that curtail in real time on two hours notice are paid the higher of $500/MWh or the real-time clearing price. While response is voluntary for the EDRP resources, SCRs are loads that *must* curtail within two hours after having been notified day ahead. The SCRs may sell capacity in the ICAP market as supply resources in exchange for accepting this curtailment obligation. When these actions are needed to meet certain reserve requirements, they will generally set energy prices at $500 per MWh. Second, the ISO may curtail exports from capacity resources or purchase emergency power from neighboring control areas. Unlike the

demand response programs, these actions do not contribute to setting energy prices during shortages.

Despite these actions, shortages can occur that require the ISO to relax its reserve requirements so the system can meet energy needs. When reserves are released and dispatched for energy, the reserve market has effectively become the marginal supplier of energy and the energy price should reflect the value of the reserves compromised. The economic value of ten-minute reserves (and hence the energy price during shortages) is implicitly established by the $1000 NYISO bid cap.

The NYISO submitted and the FERC accepted, effective June, 2003, a scarcity pricing proposal, called "Reserve Shortage Pricing", which sets the LBMP at $1000/MWh when a 10-minute reserve shortage persists and a short-term response will not immediately remedy the situation. The Real-Time energy price during scarcity conditions will be the higher of the LBMP set by the SCD, the price set under Reserve Shortage Pricing (if activated), or the price set pursuant to the pricing rules for SCR and EDRP. However, due to the relatively mild weather in the summer of 2004, the surplus in generating capability outside of New York City, and increased imports from New England, there were no shortages in 2004 and these pricing provisions were not triggered. The lack of shortage conditions in the last two years has dampened the economic signals that govern new investment.

A more sophisticated approach to shortage pricing utilizing reserve demand curves has been implemented as part of RTS in February 2005. The implementation of reserve demand curves and other changes in RTS has replaced the Reserve Shortage Pricing provisions. The reserve demand curves are fully integrated with the market software – they are included in both the day-ahead and real-time market models, ensuring that the commitment decisions made in the day-ahead market, the scheduling of external transactions and off-dispatch generation, and the dispatch of resources in real time are all consistent with the economic value of reserves. Hence, the reserve demand curves provide a more efficient means to set prices during shortage conditions. The reserve demand curves have been designed to emulate the current operating requirements and reflect the implicit value of the operating reserves based largely on the $1000 bid cap.

IV. CAPACITY MARKET

A. Background

This section assesses the design and competitive performance of the capacity market. The capacity market is intended to ensure that sufficient capacity is available to meet New York's electricity demands reliably. This market provides economic signals that supplement the signals provided by the NYISO's energy and operating reserve markets.

The NYISO implemented a change to the design of its capacity market at the end of 2001. Since that time, LSEs have been required to purchase Unforced Capacity ("UCAP") rather than Installed Capacity ("ICAP"). The difference is that UCAP is adjusted to reflect forced outages. Thus, an unreliable unit with a high probability of a forced outage would not be able to sell as much UCAP as a reliable unit of the same installed capacity. For example, a unit with 100 MW of nameplate capacity and a forced outage probability of seven percent would be able to sell 93 MW of UCAP. This creates a mechanism that attaches an explicit value to investments in reliability and gives suppliers a strong incentive to maintain their units for reliable performance.

The New York Reliability Council has recommended certain installed capacity margins for the NYISO in order to achieve NERC's one-day-in-ten-years outage standard. Since these recommendations are stipulated in terms of ICAP, the NYISO uses a control area-wide forced outage rate to convert this recommendation into UCAP terms. Likewise, suppliers sell capacity from each of their units on a similarly adjusted basis. An LSE could contract for capacity, self-schedule, or rely on the deficiency auction to fulfill its UCAP requirements. Capacity that is "self-scheduled" corresponds to capacity owned by an entity with a capacity obligation or purchased through a bilateral contract. All requirements must be satisfied at the conclusion of the spot market. All other auctions are voluntary forward markets.

Starting in June 2003, the New York state-wide ICAP purchases were no longer fixed at 118 percent of peak load. Instead, it will vary depending on the market price for ICAP, which is determined using an ICAP Demand Curve in the spot capacity auction that occurs each month. Thus, the ICAP Demand Curve replaced what was effectively a vertical demand curve with a sloped demand curve. In addition, the fixed deficiency charge was replaced with a variable charge equal to the ICAP price that results from the spot auction. For the state-wide capacity

requirement, the ICAP Demand Curve was set so that at a capacity of 118 percent of peak load (or the UCAP equivalent in the UCAP deficiency auction), the demand price would be set equal to the annualized cost of a new peaking unit. The demand price would reach zero at 132 percent of peak load, and rise to a maximum of twice the annualized cost of the new peaking unit if capacity declines below the 118 percent.

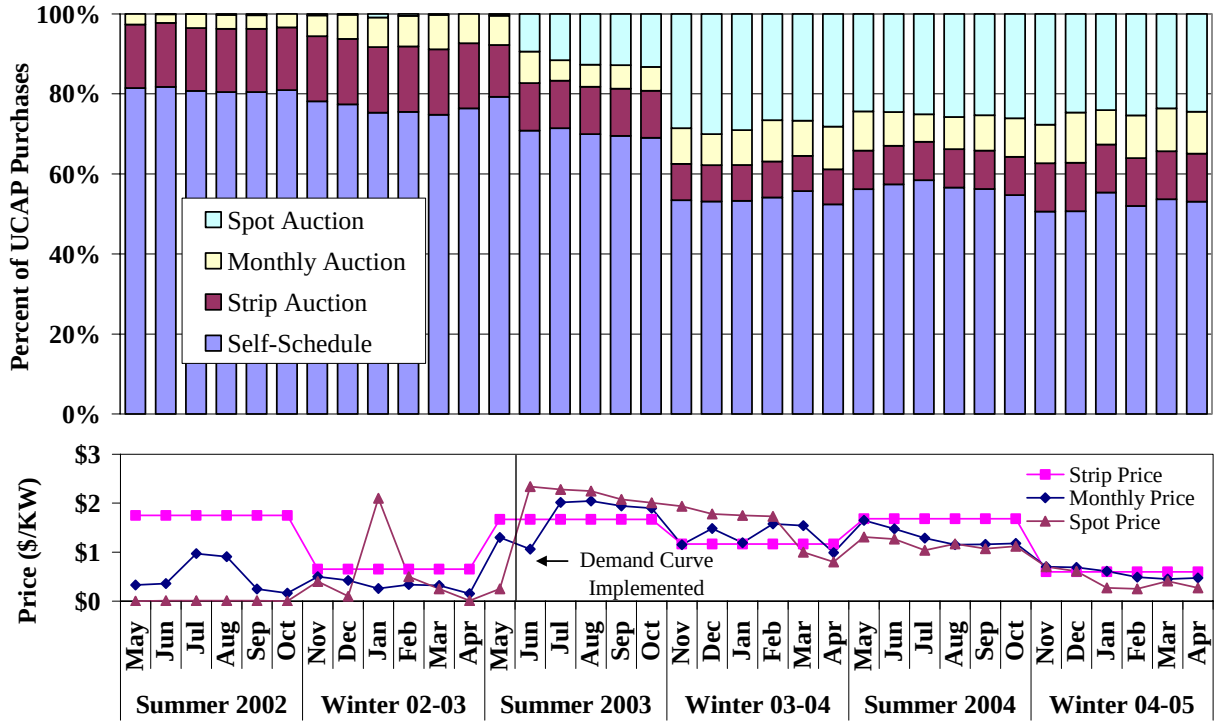
The ICAP Demand Curves for Long Island and New York City work in a similar manner, but they are adapted to the specific requirements for native generation in those areas. The ICAP Demand Curve for Long Island goes from the annualized cost of a peaking unit at 99 percent of peak load to zero at 117 percent of peak load. The ICAP Demand Curve for New York City goes from the annualized cost of a peaking unit at 80 percent of peak load to zero at 94 percent of peak load. In the unlikely event that the sales of ICAP in New York City were to exceed 94 percent, the New York City UCAP price would be equal to the UCAP price in the rest of New York State.

Monthly UCAP spot market auctions replaced LSE bids in deficiency procurement auctions. The ICAP Demand Curve and the results of the monthly UCAP supply (or bid) auction define the amount of Installed Capacity each LSE must obtain for the following month. The aggregate UCAP requirement and the associated UCAP price are established at the point where the supply curve of offers crosses the ICAP Demand Curve. All ICAP resources accepted in the auction, including resources offered by LSEs, are paid the applicable market-clearing UCAP price, and all LSEs pay the applicable market-clearing UCAP price for their UCAP requirement.

B. Capacity Market Results in 2004

To evaluate the impact of the ICAP Demand Curve on the capacity market we looked at the two six-month capability periods before the capacity demand curve was implemented and the four capability periods since implementation. Figure 38 shows UCAP prices in the “rest-of-state” area (i.e., the capacity requirements of the state after the local requirements of New York City and Long Island are satisfied). It also shows the proportion of UCAP self-scheduled and purchased in the various UCAP auctions.

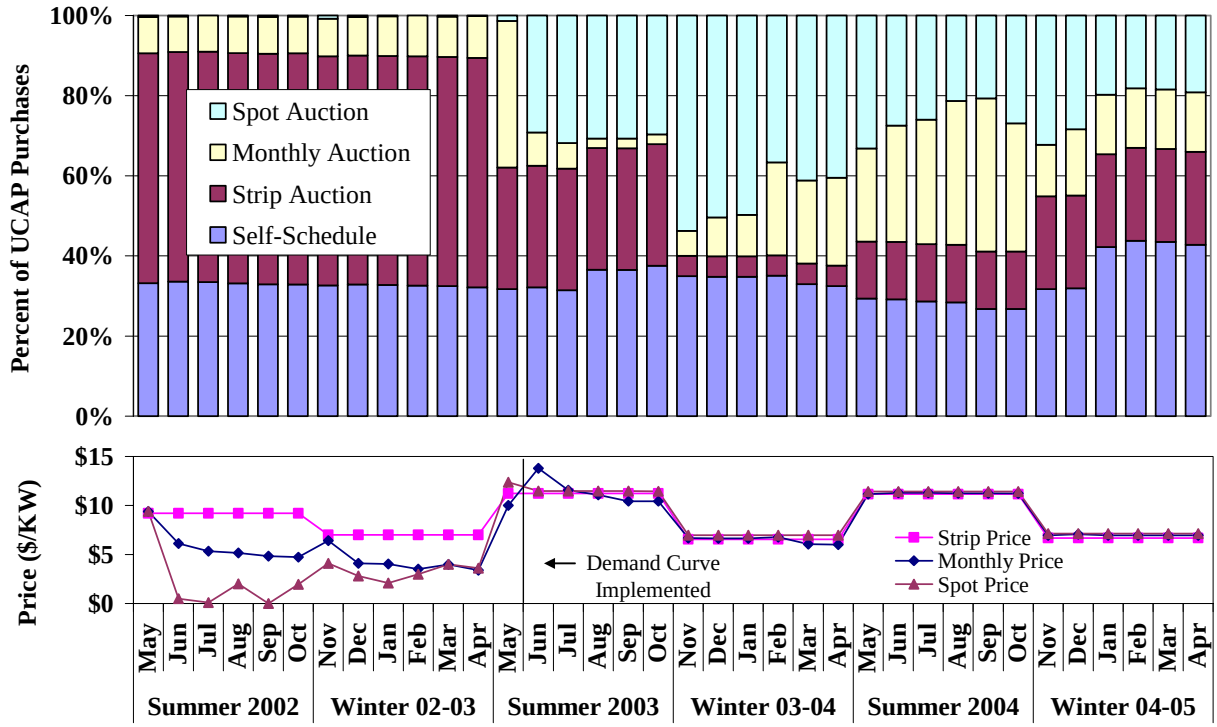
Figure 38: Unforced Capacity Market – Rest of State



This figure shows that the capacity demand curve stabilized the capacity prices and substantially improved the consistency of prices in the strip, monthly, and spot auctions. The capacity demand curve also caused a larger share of the capacity to be sold in the spot auction, where previously the small volumes purchased had contributed to erratic prices in this auction. The increase in spot procurements corresponds to a reduction in self-schedules. This is not a concern because it indicates that the spot purchases are largely displacing short-term bilateral purchases.

Overall, the capacity prices were not substantially higher following the implementation of the demand curve. Capacity prices in the strip auction, where most capacity is sold or self-scheduled, decreased slightly in the summer 2003 from the prior year and increased slightly in the winter 2003-2004 from the prior year. Capacity prices in the summer 2004 strip auction remained stable, though shorter term prices declined, reflecting mild summer weather. Prices in the winter of 2004-2005 declined significantly to pre-demand curve levels. Figure 39 provides similar data for New York City.

Figure 39: Unforced Capacity Market – New York City

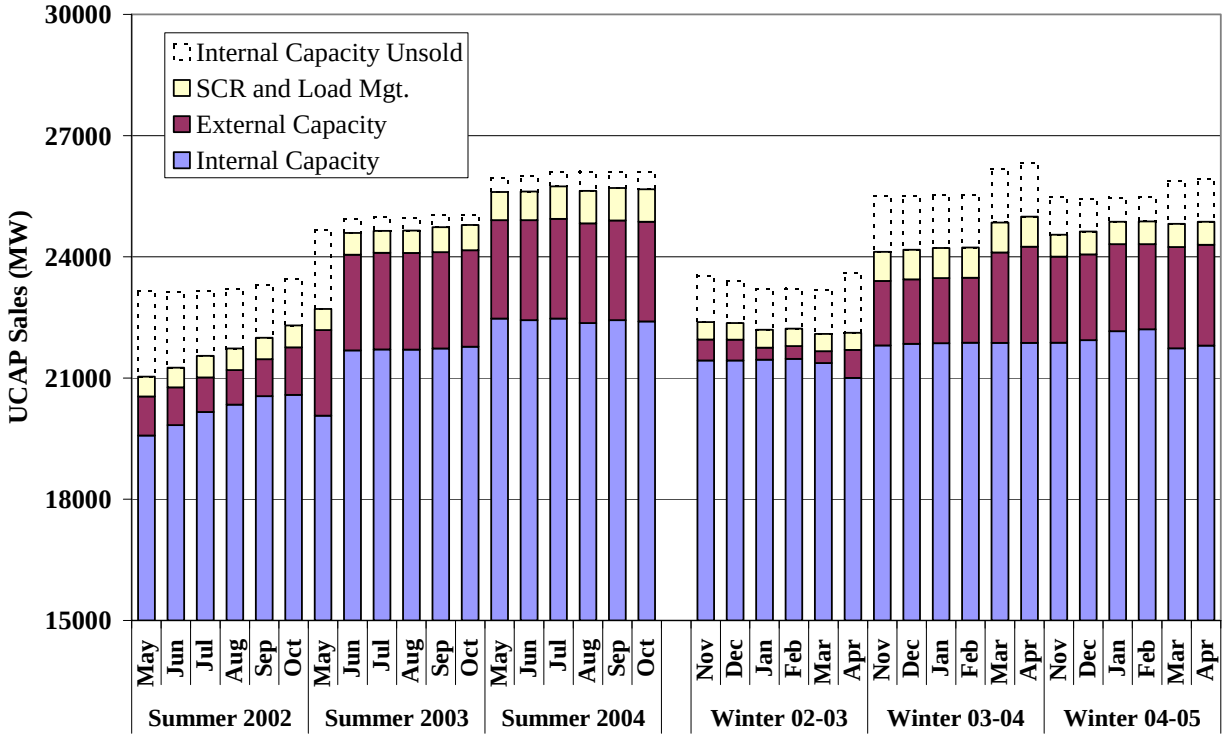


As in the upstate capacity markets, this figure shows that prices in the three auctions converged following the implementation of the demand curves. Prices were higher in summer 2003 and 2004 as the City's capacity level was at its minimum required level and purchases in the spot auction displaced purchases in the strip auction. Initially after the implementation of the demand curve, a larger share of purchases was made in the spot auction, with a lower volume purchased in the strip and monthly auctions. The portion of UCAP purchased in the spot auction rose to more than 50 percent in November 2003, then gradually decreased to 20 percent in recent months. This reflects the desire to hedge against spot capacity costs and is encouraged by the tight convergence in prices in the various markets.

One of the reasons for implementation of a capacity demand curve was to minimize the uncertainty surrounding the capacity market. The convergence and stabilization of UCAP prices is an expected and positive development. The economic signals sent by the capacity market will not have the desired effect in guiding new investment if the signals are subject to substantial uncertainty over the longer-run, causing investors to discount the capacity market signals.

The following analyses in Figure 40 and Figure 41 show the results of the capacity market over the past six capability periods (from May 2002 to April 2005). These figures show the source of UCAP supplies and quantity purchased before and after the implementation of the capacity demand curve. The amounts shown in this figure include all capacity sold by New York capacity suppliers into the New York capacity market. The hollow portion of each bar represents the in-State capacity not sold in any market.

Figure 40: UCAP Sales – Rest of State



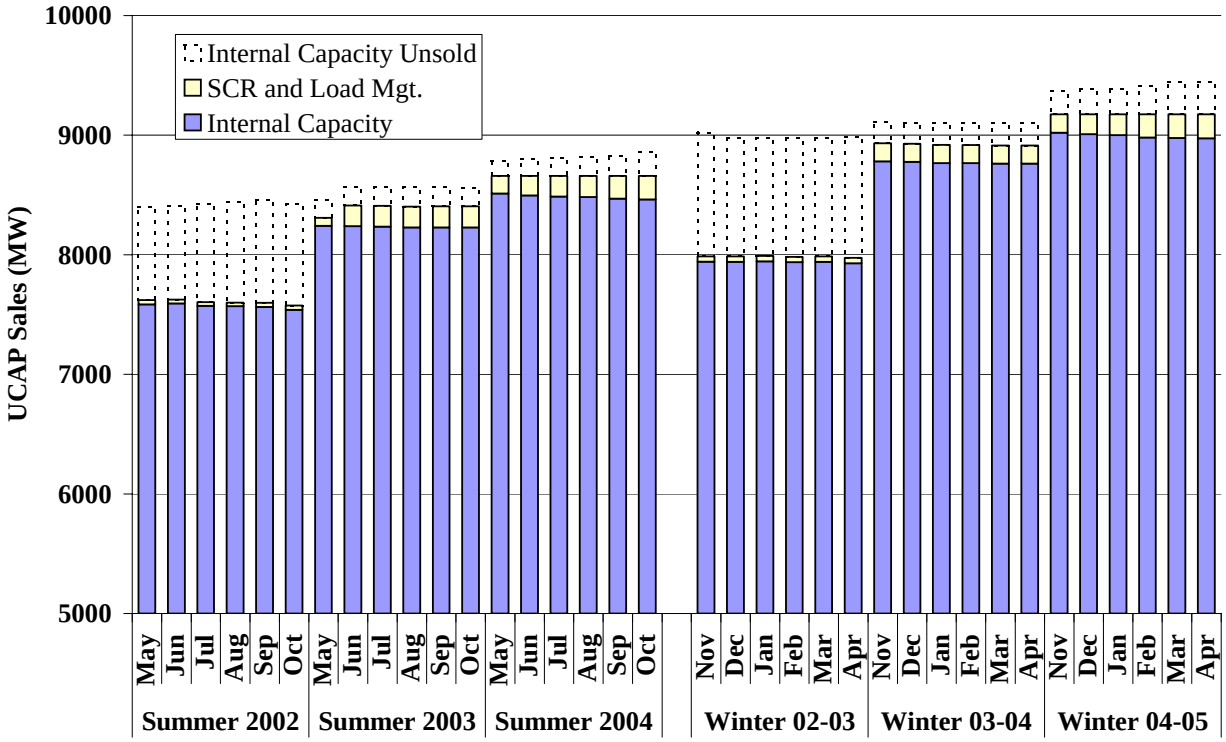
In New York State, the capacity demand curve contributed to larger purchases of state-wide capacity. The capacity demand curve resulted in additional purchases in the summer 2003 of 2200 to 2500 MW. A few hundred MW of additional capacity was purchased in the summer of 2004, due in part to the additional capacity from the Athens plant that began operation in May 2004. In the winter, the demand curve resulted in higher purchases ranging from 2500 to 3300 MW. There have also been 1000 to 1200 MW of additional UCAP purchases in New York City and Long Island during the winter capability periods that meet the state-wide requirement. These additional purchases are not shown in Figure 40. In general, the additional winter purchases are due to the higher unit ratings during the winter months that increase available UCAP supplies.

The figure also shows that most of the capacity requirement is satisfied by internal generation, although external suppliers (in the rest-of-state area) and alternative capacity suppliers (including special case resources and load management) each provide a significant amount of capacity in this market. A substantial share of the additional UCAP in all seasons came from sources external to the NYISO after the implementation of the capacity demand curve.

Figure 41 shows that capacity purchases in New York City increased significantly in 2003 and 2004. The increased UCAP purchases over the last two years are primarily due to increased peak load requirements in the City rather than the demand curve. This is because virtually all of the capacity in the City was sold, i.e., much less NYC capacity went unsold than in the rest-of-state area.

Finally, it is important to note that revenues from the capacity market play a critical role in the conclusion that the economic signals in New York City would support new investment. This is an important result because NYC capacity levels are close to the minimum required to maintain reliability, although a substantial amount of capacity was added at the Ravenswood plant in May 2004.

Figure 41: UCAP Sales – New York City



V. EXTERNAL TRANSACTIONS

This section evaluates the extent to which prices have been efficiently arbitraged between New York and adjacent regions by analyzing the price differences between the markets and the utilization of the interfaces. Although several market design improvements have been made in recent years to improve the efficiency of flows between adjacent markets, the interfaces are still not fully-utilized. There are additional changes that should be made to improve the efficient price convergence at these "seams" between New York and the adjacent markets.

In particular, we encourage the NYISO to continue working with ISO New England to develop the external scheduling provisions to enable the two markets to realize many of the benefits of a larger control area. PJM and the Midwest ISO recently implemented the Joint Operating Agreement ("JOA") to coordinate congestion in the two markets. Under the JOA, the dispatch software of each market incorporates transmission constraint information from the other market in real-time, allowing for more efficient congestion management and pricing in the two markets. The JOA could serve as a model for future coordination between New York and adjacent markets. However, in the near term, it is reasonable for New York to focus on implementing external scheduling provisions with New England to improve the price convergence between the markets.

Price convergence occurs when the energy prices at the border are equal in the absence of transmission congestion. In real-time, it has proven difficult for the adjacent markets to achieve price convergence by relying on transactions scheduled by market participants. Uncertainty, imperfect information, and offer submittal lead times limit the ability of participants to capitalize on real-time arbitrage opportunities. This failure of real-time arbitrage gives rise to market inefficiencies that could be remedied if the ISOs were to coordinate interchange to reduce or eliminate the price differences.

A. Interchange between New York and Other Markets

The performance of the wholesale electricity markets depends not only on the efficient utilization of the internal resources, but also the efficient utilization of the transmission interfaces between New York and other areas. Absent transmission constraints, trading should occur between neighboring markets to cause prices to converge. When the interfaces are efficiently

utilized, one would expect that the hourly prices in adjacent areas would not differ greatly except when the interface capability is fully used (the interface constraint is binding). In other words, when prices are higher in New England than in New York, exports to New England should continue until the interface is fully scheduled or until prices have converged and no economically-viable exports remain.

The series of scatter plots/charts in Figure 42 show the hourly difference in real-time prices between New York and neighboring markets relative to net exports during hours when transmission constraints are not binding.

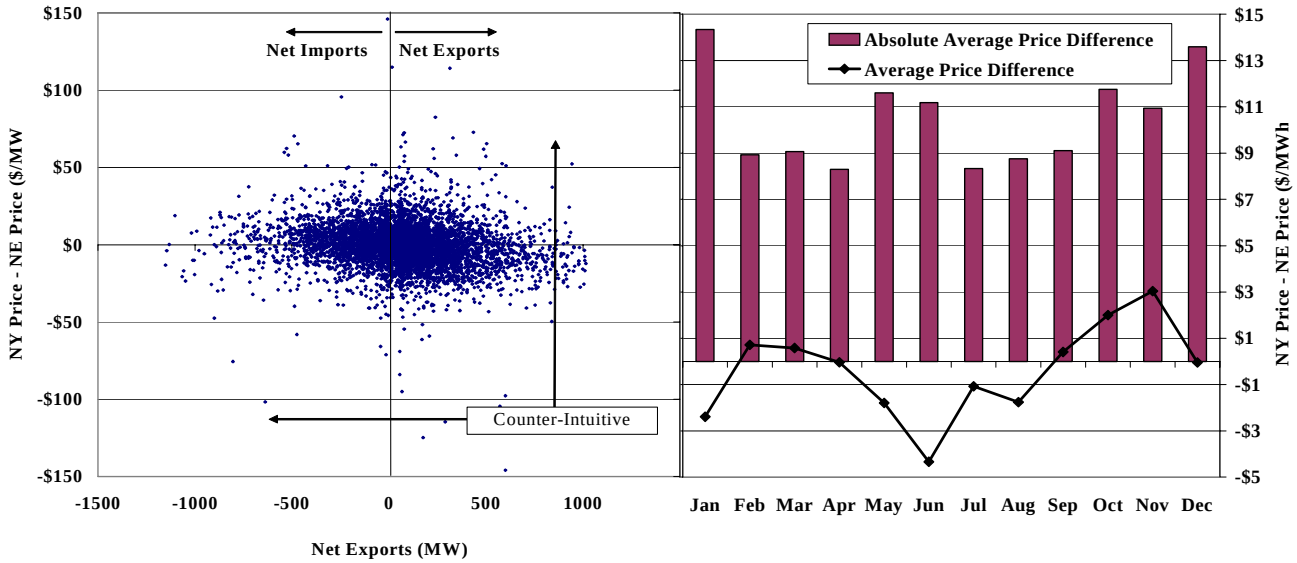
On the left side of the figures:

- The price differences plotted against the left axis are always computed by subtracting the external price from the New York price (i.e., positive price differences mean prices are higher inside New York). The top half of each scatter diagram, therefore, reflects hours when the price in New York was higher than the price in the neighboring region.
- The net exports are shown on the x-axis with positive values reflecting net exports from New York and negative values representing net imports.
- Two "counter-intuitive" quadrants are shown where power is scheduled from the higher priced market to the lower priced market.

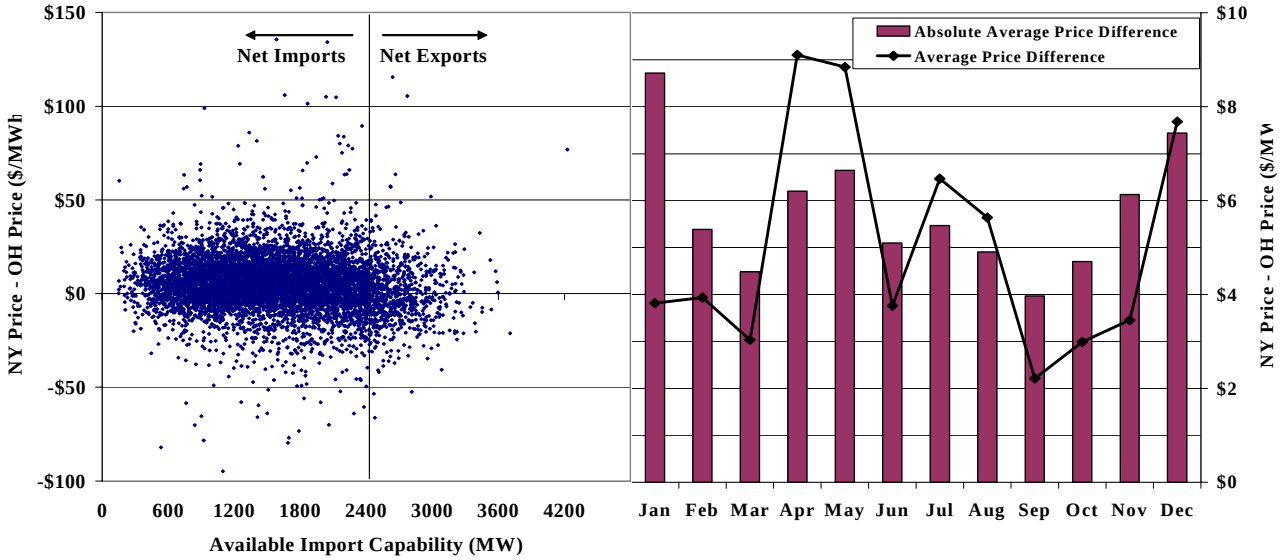
On the right side of these figures, the monthly average price differences between New York and the adjacent market are shown.

If transactions were scheduled efficiently between regions, it is expected that the points in each of the charts would be relatively closely clustered around the horizontal line at $0 – indicating little or no price difference between New York and the adjacent region in the absence of a physical transmission constraint (quantities of imports or exports can vary widely, but without transmission constraints power flows should continue in one direction or another until prices differences were arbitraged away). Moreover, one would not expect net exports to occur when the New York price substantially exceeds the price in the neighboring region. Likewise, one would not expect net imports to occur when the New York price is substantially less than the price in a neighboring region.

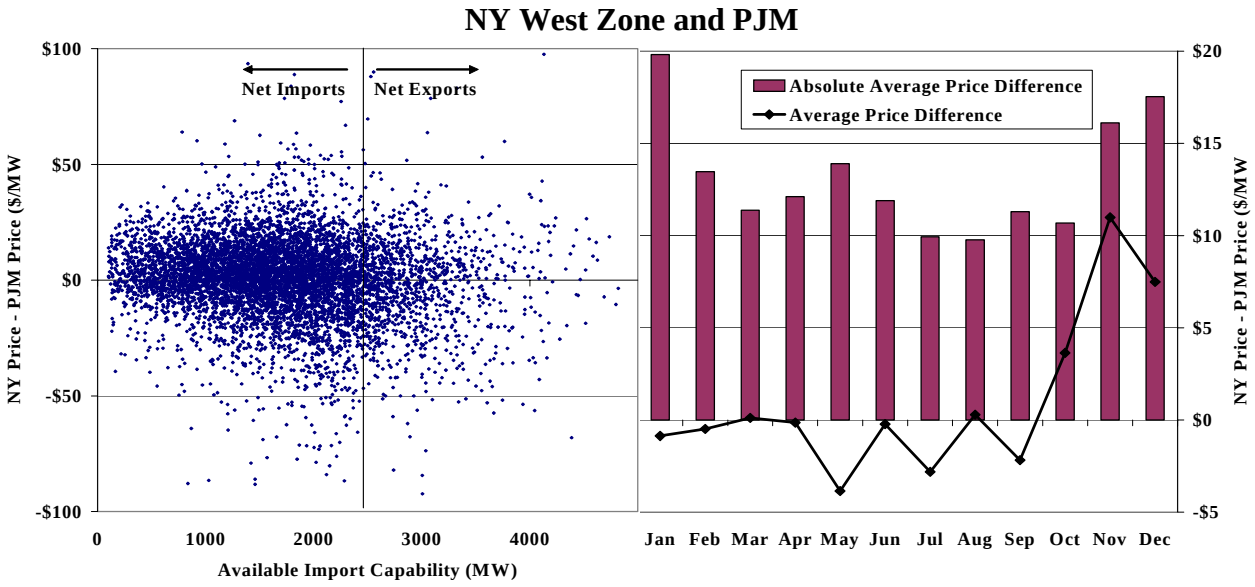
Figure 42: Real Time Prices and Interface Schedules
Eastern NY and New England



West NY and Ontario



* Price difference measured in US dollars



These figures show the real-time markets continue to not be arbitraged efficiently by participants. The dispersion in prices during unconstrained hours is shown to be considerable. In a significant number of hours for each interface, power is scheduled from the high-priced market to the lower-priced market. These results are similar to results presented in prior years.

Several factors prevent real-time prices from being fully arbitraged between New York and adjacent regions. First, market participants do not operate with perfect foresight of future market conditions at the time that transaction bids must be submitted. Without explicit coordination between the markets by the ISOs, complete arbitrage will not be possible. Second, differences in scheduling procedures and timing in the markets serve as barriers to full arbitrage. Third, there are substantial transmission fees and other transaction costs associated with scheduling imports and exports that diminish the returns from arbitrage. Participants would not be expected to schedule additional power between regions unless they expect a price difference greater than these costs. Last, risks associated with curtailment and congestion will reduce participants' incentives to engage in external transactions at small price differences.

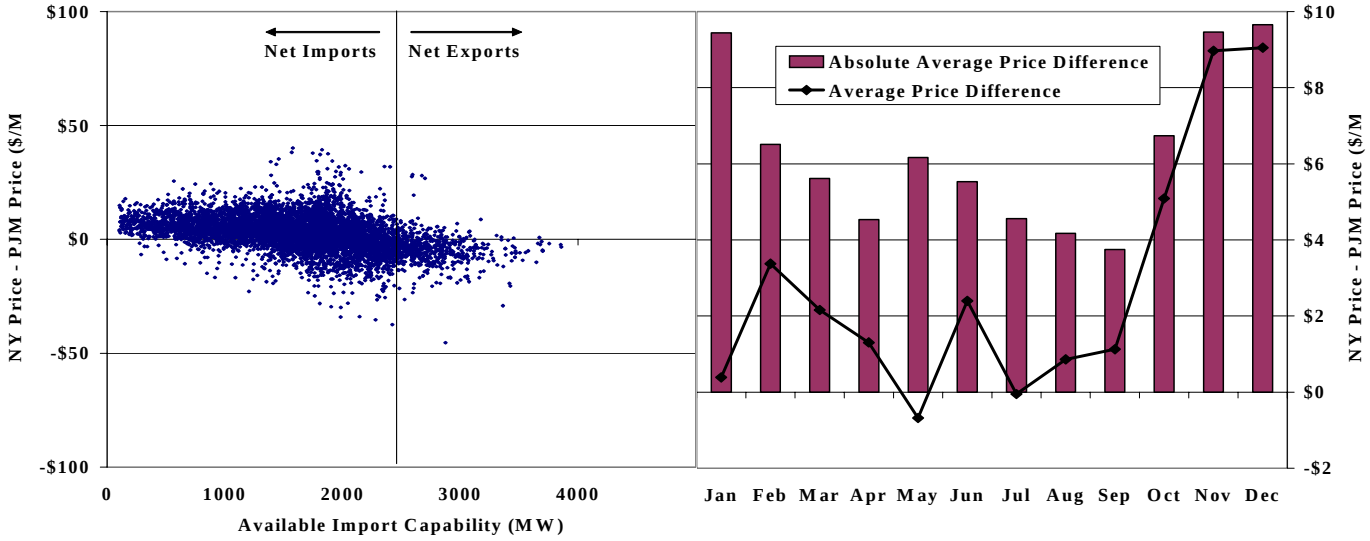
Uncertainty, imperfect information, and required offer lead times limit the ability of participants to capitalize on real-time arbitrage opportunities. These results reinforce the importance of intrahour transaction scheduling provisions that are being developed to improve real-time

interchange between New York and New England. This will be particularly important when the capacity surpluses in the Northeast are eliminated – when optimizing the flow between areas will have larger economic and reliability consequences.

Fees assessed to transactions between control areas tend to inhibit convergence. In 2005, export fees between New York and New England were eliminated, which will help improve the arbitrage of the adjacent markets. However, exports from New England scheduled after the day-ahead market continue to be allocated substantial uplift charges associated with regulation service and certain types of supplemental commitment.

This figure shows that although the average price difference for the New England and PJM interfaces is relatively low, dispersion of the prices in the absence of congestion is substantial. The figure shows that the typical price difference between New York and these adjacent markets ranged from $8 per MWh to $20 per MWh on a monthly average basis. This indicates that significant seams issues remain that continue to prevent efficient interchange between the market areas. Figure 43 presents a similar analysis for the price differences between PJM and New York in the day ahead market.

Figure 43: Day Ahead Prices and Interface Schedules
NY West Zone and PJM

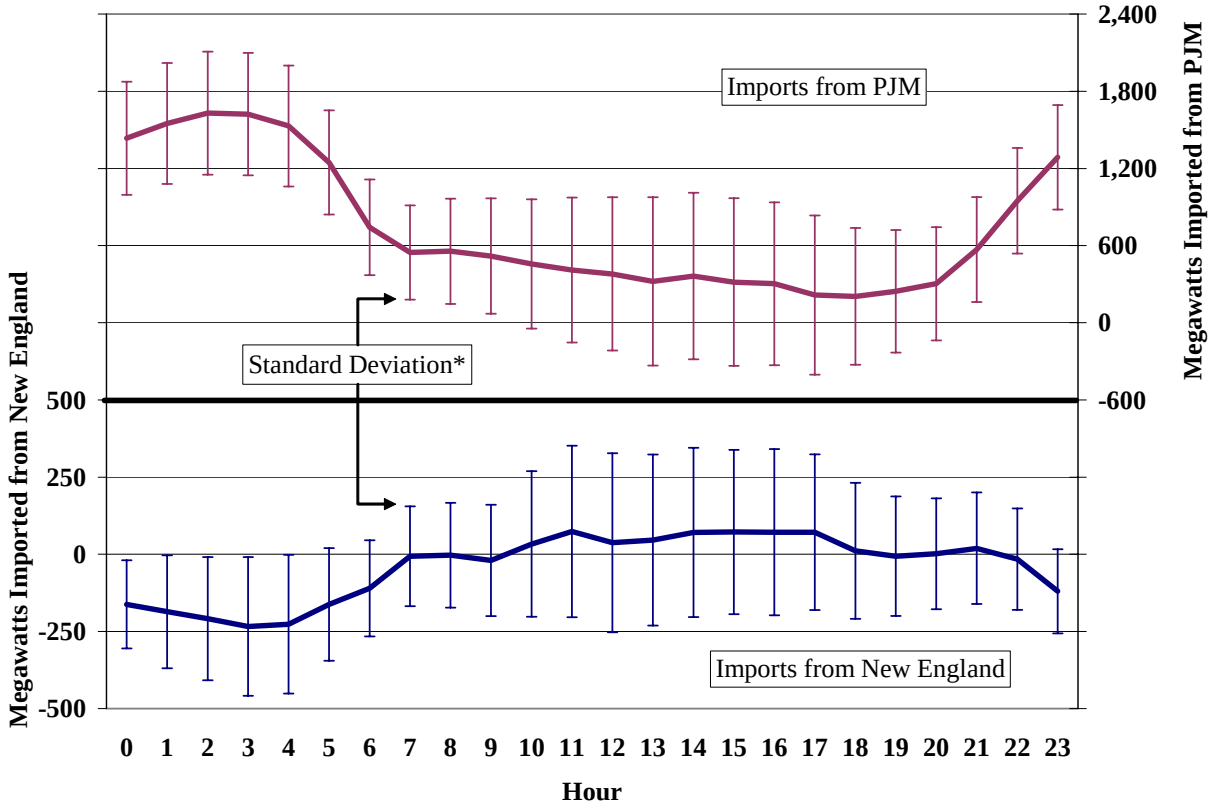


This figure shows that prices are not efficiently arbitraged day-ahead, though the reduced volatility in prices in the day-ahead markets contributes to a tighter dispersion of the prices. The monthly standard deviations of the price difference are much lower in the day-ahead market than in the real-time market.

B. Scheduled Interchange by Hour of Day

We also examined the temporal pattern of imports and exports to and from the New York markets. Figure 44 and Figure 45 show how real-time imports vary across an average day over each external interface.

Figure 44: Average Net Imports from LMP Markets by Hour of Day Weekdays 2004



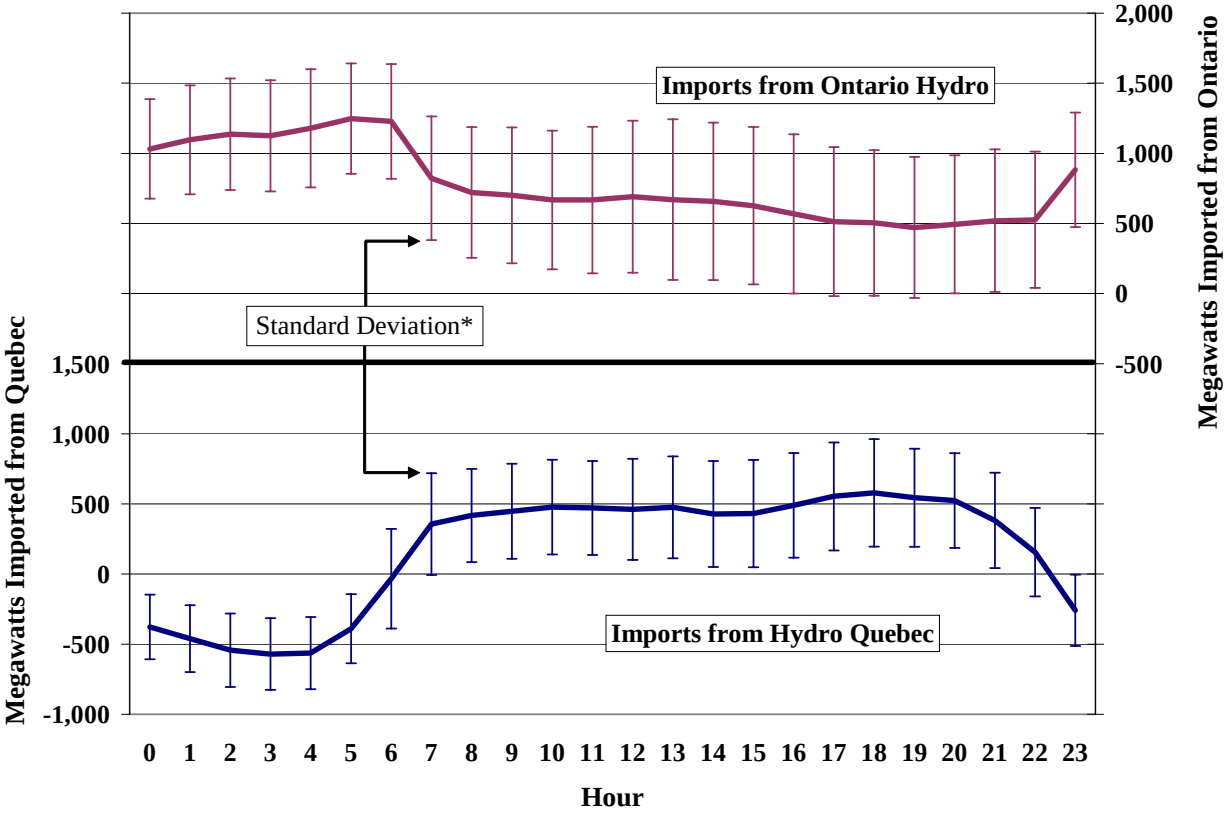
Imports from PJM are highest during the night-time hours, while New York is a net exporter to New England during this period. During the day, New York imports from both regions. Though PJM exports a smaller quantity to New York during the day than at night, it is still much larger than supply obtained from New England. Although the interface capability is smaller and trading activity is lower with New England than with PJM, trading with New England is more

economically significant because New England exports serve the congested Eastern New York area. However, in the overwhelming majority of instances, only a small portion of the interface capability is being used, even in hours where there are substantial price differences.

Figure 45 shows the transactions with Canada. Hydro-Quebec is a net importer at night and exporter during the day from New York. New York typically receives 500 MW of imports from Ontario during the day and nearly 1000 MW at night. This is a significant increase from 2003.

The figure also shows that there is a substantial change in the average interchange in hours 6 and 22. The change in schedules is consistent with schedules made to support longer-term bilateral agreements (rather than arbitrage of hourly prices). Many of these schedules tend to be insensitive to real-time prices and contribute to the price divergence.

Figure 45: Average Net Imports from Canada by Hour of Day Weekdays 2004



C. Conclusions and Recommendations

Over the past several years, modeling improvements and rule changes have led to substantial declines in price differences between control areas during non-transmission-constrained hours. While the external transaction scheduling process is functioning properly, significant price differences remain between markets in hours when no congestion is present. The economic consequences of these issues has been minimized over the past two years because there have been no instances of shortages. The economic effects of the seams issues are the largest when one market experiences a shortage that could have been avoided if the external interfaces were fully utilized. This has not occurred over the past two years because each of the markets currently has a surplus of generating capacity and peak loads have been moderated by mild summer weather.

These results reinforce the importance of addressing remaining seams issues. We continue to encourage New York and New England to develop and implement new scheduling procedures, such as "intrahour transaction scheduling". Intrahour transaction scheduling is a process that would allow the physical interchange to be adjusted within an hour when prices diverge at the interface between the two markets. These adjustments would ensure that the interchange levels are efficient, eliminating the price distortions and other inefficiencies caused by poor market arbitrage. This will lead to less volatility and more predictability in the New York to New England prices. Likewise, we recommend that the NYISO work with PJM to eliminate export fees and improve scheduling procedures.

VI. ANCILLARY SERVICES

A. Background

The NYISO operates ancillary services markets in conjunction with the day-ahead and real-time energy markets. These include three operating reserve markets and a regulation market. This section reviews the competitive performance of these markets in 2004. This section also summarizes the modifications that were introduced under RTS to improve the performance of these markets in February 2005.

New York procures three types of operating reserves: ten-minute spinning reserves, ten-minute total reserves (can be spinning or non-synchronous reserves), and 30-minute reserves. Ten-minute spinning reserves are held on generating units that are on-line and can provide additional output within 10 minutes. Ten-minute total reserves can be supplied by ten-minute spinning resources or ten-minute non-spinning resources, which are typically gas turbines that are not on-line but can be turned on and be producing within 10 minutes. 30-minute reserves may be supplied by any unit that can be ramped up in 30-minutes or that can be on-line and be producing within 30 minutes.

New York also purchases regulation services, necessary for the continuous balancing of resources (generation and NY Control Area interchange) with load and to assist in maintaining scheduled interconnection frequency at 60 Hz. This service is accomplished by committing on-line generators whose output is raised or lowered (predominately through the use of Automatic Generation Control) as necessary to follow moment-by-moment changes in load.

During 2004, the NYISO received availability offers from each generator that indicated the minimum price they are willing to accept to provide each reserve product. The marginal cost of procuring reserves includes both the availability offers and the opportunity costs in other markets (i.e., holding economic resources out of the energy market is part of the cost of maintaining operating reserves). Both of these costs are considered in the simultaneous optimization of the reserve designation and energy dispatch. However, reserve prices are set in each market by the highest-accepted availability offer – while opportunity cost payments are made to the providers of regulation and spinning reserves in the real-time market and to the providers of ten-minute

non-spinning reserves in the day-ahead market. Currently, the NYISO operates only a day-ahead market for reserves, although it reallocates the reserves hourly during the operating day.

In each hour, the New York ISO purchases approximately 1800 MW of operating reserves. Of this 1800 MW, at least 1200 MW must be ten-minute reserves (at least 600 MW must be spinning reserves and the balance may be either spinning or non-spinning). Consequently, the NYISO may purchase up to 600 MW of 30-minute reserves. There is no limit on how much spinning reserves is purchased – all 1200 MW of total ten-minute reserves (indeed, all 1800 MW of the total operating reserves) could be spinning reserves. Hence, ten-minute spinning reserves are the highest-valued reserve while 30-minute reserves are the lowest-valued reserve.

The reserves markets are cleared simultaneously with the energy market to minimize total bid-production costs. In this process, the price for lower-valued reserves typically clears below the price for higher-valued reserves. The simultaneous auction design ensures that lower-valued reserves will never be priced above higher-valued reserves. This is because a surplus of the higher-valued product could always substitute for the lower-valued product, leading the two products to have the same price.

The procurement of reserves is also subject to locational requirements to ensure that they will be fully available to respond to possible system contingencies. Because of the Central-East Interface, maintaining reliability requires that a substantial portion of the reserves be procured in Eastern New York. Likewise, the interface between Long Island and the rest of New York has resulted in a requirement that specified amounts of operating reserves must be purchased from generating units on Long Island.

For total ten-minute reserves (spinning and non-spinning) 1000 MW must be purchased east of the Central-East constraint, including at least 300 MW of 10 minute spinning reserves. Prior to 2002, the eastern requirement was 1200 MW. However, it was lowered to 1000 MW after the NYISO and ISO-NE entered into a reserve-sharing agreement. The locational reserve requirements for Long Island oblige the NYISO to designate at least 60 MW of ten-minute spinning, 120 MW of total ten-minute, and 540 MW of total reserves (ten-minute and 30-minute) on Long Island.

The NYISO sets prices for reserves that can vary for Western New York, Eastern New York, and Long Island when the locational reserve requirements are binding. This change allows reserve prices to be set by the marginal reserve supplier to satisfy each of these locational reserve requirements. The primary result of this locational pricing is that higher prices for the ten-minute reserves will emerge in the East when the locational requirements are binding.

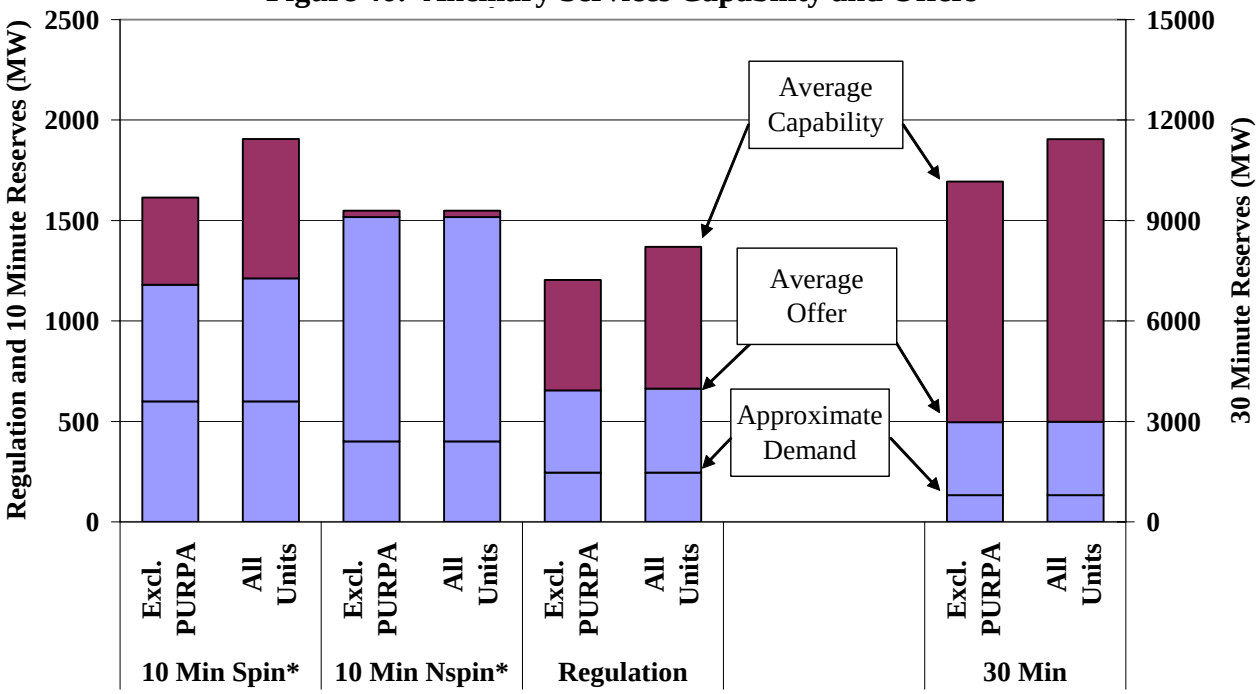
Regulation capability can be purchased from anywhere within the New York Control Area. The NYISO purchased 275 MW of regulation during high-ramp hours and 200 MW during low-ramp hours in 2004. The amount of regulating capability a generating resource may sell is equal to the amount of output it can produce within 5 minutes (ramp rate per minute times 5). In addition, to qualify as a regulating unit, the unit must be able to receive and respond to a continual dispatch signal and have the ability to ramp at a rate of 1 percent of the unit's total capability per hour.

B. Offer Patterns

Our findings in previous analyses in New York have indicated that a substantial portion of the capability of certain services is not offered in the day-ahead ancillary services markets, particularly for 30-minute reserves and regulation. Offering into the ancillary services markets is not mandatory in the day-ahead market, with the exception the ten-minute non-spinning reserves in Eastern New York. This section reassesses the ancillary services offer patterns to determine whether participation in this market has improved.

Figure 46 summarizes the average levels of capacity, offers to supply, and demand for all three day-ahead reserves products as well as demand for the day-ahead regulation service. Because of the nature of the locational requirements, ten-minute reserves are shown only for the region east of the Central-East Interface. In addition, the results of this analysis are shown with and without the PURPA units because a large portion of this capacity may be contractually limited from supplying the reserves markets.

Figure 46: Ancillary Services Capability and Offers



*Eastern side of the Central-East Interface only

With the exception of the 10-minute non-synchronous resources, which remain subject to the mandatory offer requirement, a substantial portion of the capability of the other ancillary services was not offered in the day-ahead markets. However, ancillary services markets are generally not tight because offers to supply typically exceed approximate demand:

- For 30 minute reserves, offers typically exceed approximate demand by 280 percent.
- For regulation, offers exceed approximate demand in ramping hours by 120 percent.
- Offers for total 10-minute reserves east of the Central-East interface, typically exceed demand by 170 percent.
- For 10-minute spinning reserves, offers typically exceed approximate demand by 100 percent – but this ignores the fact that some 10-minute spinning reserves can be purchased in the West.

Participation in the regulation and 30-minute reserves markets remains relatively poor. The average quantity of regulation being offered to the market is approximately one-half of the total capability, and the average quantity of 30-minute operating reserves being offered is less than one third of the total capability. Generally, this is not a significant concern given the excess reserve and regulation capability that is available. However, since these markets are jointly

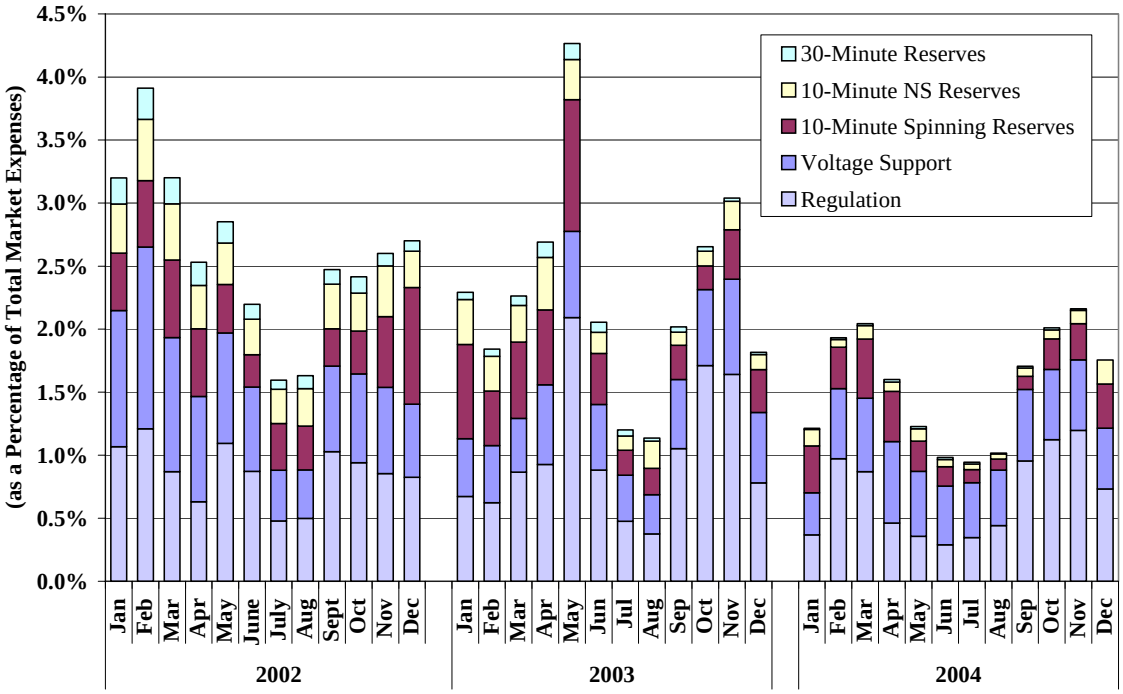
optimized and the same resources are offered in multiple markets, under peak load conditions, energy and other ancillary services markets can bid resources away from a given service resulting in relatively tight conditions in the day-ahead ancillary services markets.

During 2004, the NYISO did not pay lost opportunity cost payments to generators for providing 30-minute reserves or regulation when it would have been more profitable for them to produce the energy. This can provide a disincentive for generators that might otherwise make offers to these markets, and may help explain why the portion of capacity offered in the regulation and 30-minute reserve markets is particularly low. This problem has been addressed under the RTS software implemented in February 2005. Under the new market software, ancillary services prices will incorporate opportunity costs so that generators will never be harmed by being selected to provide ancillary services rather than energy. We will evaluate the performance of the new market software following the summer of 2005.

C. Ancillary Services Expenses

Figure 47 shows the ancillary services expenses, which include expenses for regulation, voltage support, and various operating reserves. These expenses tend to be smaller as a percent of total market expenses in the summer because loads and energy prices are higher in the summer.

Figure 47: Ancillary Services Costs 2002 - 2004



Ancillary services costs declined slightly as a percentage of total market expenses from close to 2.5 percent in 2002 to roughly 1.5 percent in 2004. During this timeframe, total ancillary services expenses decreased by $15 million to approximately $94 million in 2004, after almost reaching a total of almost $130 million in 2003 due to higher fuel costs. Reduced ancillary services expenditures were primarily due to a $19 million reduction in the cost of operating reserves from 2002 to 2004.

Declines in operating reserves costs since 2001 can be attributed to three market design changes. First, the reserve-sharing agreement implemented in March 2002 with ISO-NE permitted a reduction in the ten-minute reserve requirement for the East (from 1200 MW to 1000 MW), although the state-wide requirement is still 1200 MW. Second, locational ancillary services prices for Long Island, Eastern New York (excluding Long Island), and Western New York, implemented in October 2001, limited the impact of reserve shortages in constrained areas on state-wide reserve prices. Third, changes in April 2002 to the BME model to recognize latent 30-minute reserves on un-dispatched portions of on-line resources, resources that are available to the real-time model for energy but did not submit a 30-minute reserves availability bid, prevent the BME model from setting irrationally high prices for reserves when plenty of 30-minute capability is available.

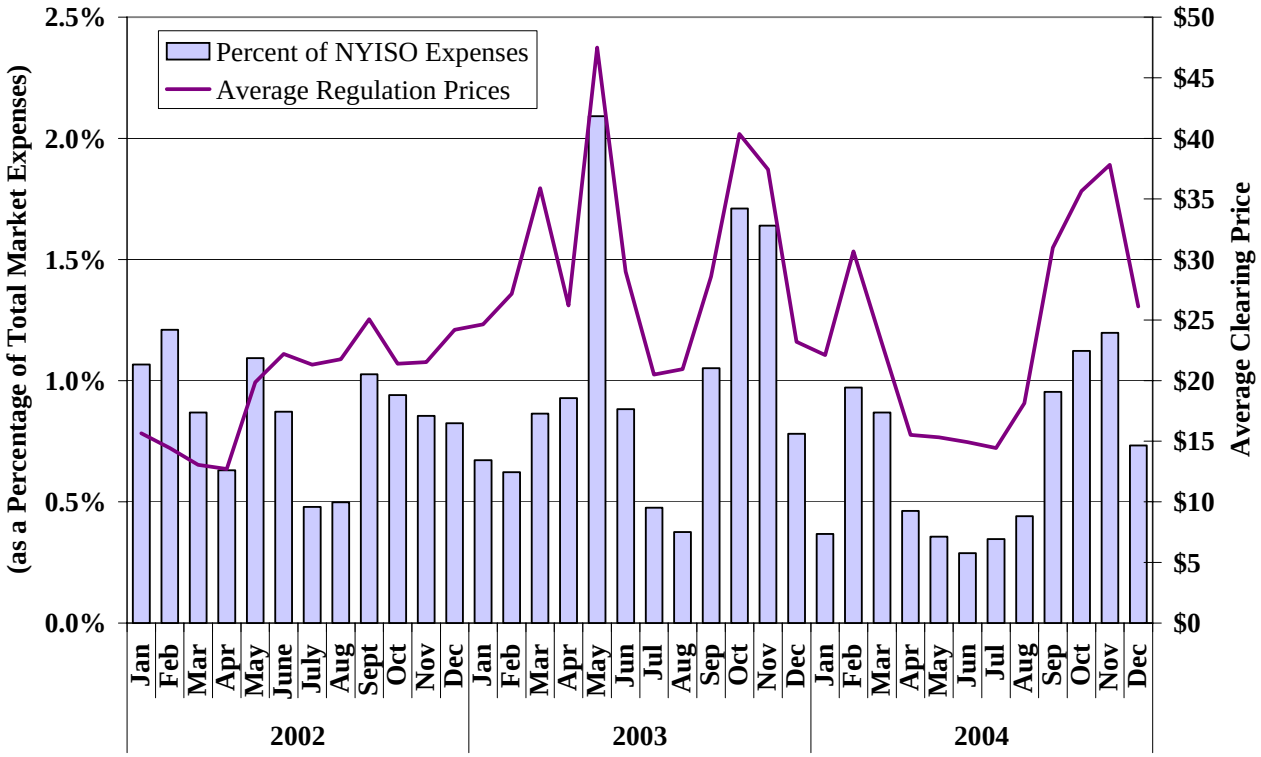
There was some concern that lifting the $2.52 bid cap for 10-minute non-spinning reserves would lead to higher prices for operating reserves. The bid cap was imposed in the spring of 2000 as a remedy for uncompetitive outcomes that were occurring in the markets for 10-minute reserves. However, since the bid cap was lifted, the costs of both 10-minute reserves products have dropped substantially, confirming that the bid cap is no longer necessary.

We expect further improvements with the implementation of RTS, as the multi-settlement system for reserve procurement eliminates additional costs incurred in today's market. More efficient pricing of reserves during shortage conditions is likely to increase total reserve costs, despite cost reductions due to other RTS improvements. This is an important feature of the RTS operating reserves markets because it provides the necessary economic signals to attract and retain resources that are primarily needed to meet the NYISO's reserve requirements, such as gas turbines. The changes under RTS are discussed at the end of this section in greater detail.

D. Regulation Market

This subsection focuses on the regulation market, which is the only market-based ancillary service that is not a type of operating reserve. Figure 48 shows the average price for regulation service from 2002 through 2004, as well as the share of the total market expenses that are accounted for by regulation.

Figure 48: Average Clearing Price and Expenses for Regulation Procurement



Regulation prices have increased from 2002 levels. One reason for the increase was modeling changes in SCUC and BME, initiated in May 2002, to recognize that units' minimum generation level may limit the range in which a unit can regulate down. This reduced the supply available on some units, particularly off-peak. Previously, units could be scheduled with unequal amounts of up-regulation and down-regulation, whereas now units must be scheduled for equal amounts. This constraint on assigning regulation will not exist after the implementation of RTS.

The second factor that contributed to the rise in regulation prices is higher fuel prices that increase opportunity costs to provide regulation and raise regulation prices, though the impact is much smaller than the effect of higher fuel costs on the energy market. Fuel price increases that

increase opportunity costs to provide regulation contributed to higher prices in 2003 and 2004. Overall, regulation costs still remain a relatively low portion of the total electricity market expenses for the NYISO (slightly more than 1 percent).

E. Changes in Reserve Markets

The implementation of RTS in 2005 will lead to major changes in the markets for reserves and regulation. The co-optimization of energy and ancillary services in real-time will enhance market efficiency. The multi-settlement system for the reserve and regulation markets will eliminate additional costs due to re-optimization or procurement of replacement services in real time. Under the multi-settlement system, real-time ancillary services schedules will be settled against the day-ahead schedules. Since suppliers are liable for the real-time cost of reserves that they schedule day ahead, they will have an incentive to be available in real time and to perform when called.

Reserve market clearing prices will be set in both the day-ahead and real-time markets on a locational basis using the shadow prices of the reserve constraints out of the SCUC and RTS models. Both day-ahead and real-time clearing prices of ancillary services will cover the lost opportunity cost of the marginal supplier (i.e., the supplier with the lowest energy bid and, thus, the highest opportunity cost). This is intended to give price incentives to the lowest-cost reserve providers to provide reserves rather than energy, and eliminate the need for separate lost opportunity cost payments currently recovered through uplift charges.

Regulation suppliers will submit availability offers for both the day-ahead and real-time regulation markets, while availability offers for reserves may only be submitted in the day-ahead reserve markets. For resources offering into the energy market, real-time availability offers for reserves are fixed at $0, reflecting the fact that these resources are already available. Hence, all "On-dispatch" and Self-Committed Flexible resources (including eligible demand side resources) that submit energy offers will be considered for reserve scheduling in real time. ICAP resources that offer into the day-ahead energy market, but do not submit an availability offer for reserves will have default availability offers of $0.

In the RTS design, the current reserve shortage pricing provisions have been superseded by the reserve demand curve. There will no longer be special energy pricing rules invoked when there

is a persistent 10-minute reserve shortage. Instead, the demand curves establish an economic value for reserves that will be reflected in energy prices at times when the energy market must bid scarce resources away from the reserve markets. Locational reserves prices are based on the shadow prices of reserves constraints, but operating reserves purchases will be reduced when necessary to prevent the shadow prices from exceeding the prices set forth by the demand curves.[16] The demand curve values have been set at levels that are consistent with the actions normally taken by the NYISO operators in reserve shortage conditions. This should ensure greater consistency between prices and the operation of the system, and better reflect the economic value of reliability. A reserve demand curve has been applied to each of the nine reserve constraints in the New York Control Area. The reserve demand curves have been applied consistently in the day-ahead and real-time markets.

16 The total value of a reserve in a location will be the sum of the reserve demand curve values for each reserve requirement constraint that the reserve contributes to relieving. In other words, because reserves should generally be substituted to maintain the highest quality reserve, the total value of a specific reserve type will generally include the sum of the demand curve values of the lower quality reserves.

VII. DEMAND RESPONSE PROGRAMS

The New York ISO has some of the most effective demand response programs in the country. There are currently three demand response programs in New York State:

- Day-Ahead Demand Response Program (DADRP) – This program schedules physical demand reductions for the following day, allowing resources with curtailable load to offer into the day-ahead market like any supply resource. If the offer clears in the day-ahead market, the resource must curtail its load in accordance with the accepted offers and is paid day-ahead clearing price for each MW of curtailed load.

- Special Case Resources (SCR) – These are loads that must curtail within two hours. They are called when operators forecast a reserve deficiency. These resources may sell capacity in the capacity market corresponding to their commitment to curtail load.

- Emergency Demand Response Program (EDRP) – The emergency demand response program pays loads that curtail on two hours notice the higher of $500/MWh or the real-time clearing price. SCRs receive this payment as well.

The EDRP and SCR programs have been effective in achieving actual load reductions during peak conditions. The total registered quantity of more than 1700 MW is larger than most comparable programs in other ISOs.

The success of these programs is largely due to incentives provided by the programs. EDRP participants are paid the higher of $500/MWh or the LBMP for voluntary load reductions (i.e., they have no obligation to respond), which is the only source of revenue for the EDRP resources. SCR resources can sell their curtailable load in the capacity market in exchange for an obligation to respond when called. SCR participants are paid the higher of a strike price that they bid (limited to be less than $500/MWh) and the LBMP.[17]

This payment structure satisfies two critical objectives. First, it results in payments to participants that are close to or exceed $500/MWh, which allow them to be paid an amount that covers their marginal value of consumption during peak periods. Hence, it would provide an adequate incentive for loads to respond, even though most are served under regulated or

[17] The NYISO will provide a 24-hour notice if it anticipates a need to make curtailments under the SCR program to meet reserve requirements. These curtailments may or may not ultimately be called. However, there is a two-hour notice given when the NYISO determines that the load should be curtailed. EDRP also provides the NYISO with resources to meet potential reserve shortfalls. These curtailable load resources are given two-hours notice prior to being asked to curtail.

otherwise fixed rates that cause them not to incur the wholesale price of electricity.[18] Second, during times when EDRP and SCR are the marginal sources of supply in the market that allow the system to satisfy its reserve requirements, the LBMP typically will be set at $500/MWh. This price is in a range that is consistent with the marginal value of reserves to the system. Hence, these payments and the associated pricing provisions contribute to efficient pricing during shortage (or near-shortage) conditions.

The EDRP and the SCR programs can contribute substantial demand-side resources to the market. Special Case Resources are qualified to sell into the capacity market, and by adding to the total supply, help reduce capacity prices. In 2004, the quantity of SCR/ICAP subscribers that sold capacity were 175 MW in NYC, 98 MW in Long Island, and 707 MW in upstate New York. The total UCAP sales from SCRs has increased 30 percent from 2003. EDRP and SCRs were not utilized during real-time in 2004 due to mild load conditions and good resource availability.

The day-ahead demand response program has provided considerably less potential demand reduction than the EDRP and SCR programs. There were 2818 hours with day-ahead demand response bids. The average quantity bid was approximately 2 MW per hour, and the average quantity scheduled was less than half a megawatt. There were 222 hours when day-ahead demand response bids amounted to 10 MW or more, with a high of 17 MW, and these bids were accepted in 132 hours. The hours with these large bids primarily occurred around holidays such as New Year's Day, Thanksgiving, and Christmas week. The low participation may be due to the alternatives available for demand to bid in the markets (virtual trading and price-capped load bidding).

[18] While the average regulated rate paid by load is much lower than $500/MWh, the value of power at peak times is typically much higher than the average. Therefore, in the absence of the NYISO's payments for EDRP and SCR load reductions, load that is interrupted would save only the regulated rate. This rate does not reflect the marginal system cost of serving the load as embodied in the wholesale LBMPs.