The Pennsylvania State University

The Graduate School

College of Earth and Mineral Sciences

ARE LOAD FORECASTS PREDICTABLE?

AN ANALYSIS OF ELECTRICITY LOAD FORECASTS ISSUED BY THE NEW YORK INDEPENDENT SYSTEM OPERATOR

A Thesis in

Meteorology

by

Eric E. Wertz

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

May 2009

The thesis of Eric E. Wertz was reviewed and approved* by the following:

Arthur A. Small III
Associate Professor of Applied Economics and Finance
Thesis Advisor

Seth Blumsack
Assistant Professor of Energy and Mineral Engineering

George Young
Professor of Meteorology and Geo-Environmental Engineering

William Brune
Professor of Meteorology
Head of the Department of Meteorology

*Signatures are on file in the Graduate School

ABSTRACT

A statistical evaluation was conducted of electricity demand (load) forecasts for New York City issued by the New York Independent System Operator (NYISO). Analysis focused on how the NYISO refines its initial predictions progressively as the interval between forecast and observation closes. Analysis reveals that forecast adjustments exhibit a high degree of positive serial correlation: a forecast adjustment in a given direction (e.g., upward) tends to be followed the next day by a subsequent adjustment in the same direction. The NYISO's forecast refinements are thus substantially predictable, a pattern inconsistent with efficient forecasting. Load forecasters may rationally be seeking to minimize not a standard statistical measure of error, but a financial *cost* of error.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank my thesis advisor Arthur Small for offering me the opportunity to conduct research with him and for all of his support throughout the research process. I am also very appreciative of all the insights and advice that I have received from the other members of my thesis committee, Seth Blumsack and George Young. My great thanks also go out to NSF (Grant #0729413) for their financial support. I would also like to thank all of my friends and colleagues who I have interacted with here at Penn State for all their support. Lastly, I would like to thank the Pennsylvania State University Meteorology Department and College of Earth and Mineral Sciences for being a great and incredibly supportive home the past 5 years.

Chapter 1: Introduction

Forecasts can be issued about a vast array of things and are typically assumed to be the forecaster's best prediction of the realized value for whatever is being forecasted. Therein, however, lays the question of what exactly a best forecast is.

The aforementioned decision context determines whether or not a forecast is the best. For some, the best forecast could be one that has the lowest RMSE (root mean squared error) while others may say it is the one with the lowest variance or perhaps something else completely different. One aspect to consider is that some forecasters may be risk averse. That is, they issue a forecast knowing it may deviate from their expected outcome due to potential benefits a skewed forecast could bring. Such forecasts exhibit larger than necessary RMSEs and variances, causing some to discredit them because of misconceptions about the underlying purpose of the forecast. For example, a forecasting entity could always issue a daily forecast of zero tornadoes for Oklahoma City and be right virtually all the time, giving the entity a RMSE of near zero. On the contrary, another forecasting entity that does in fact forecast a tornado when the conditions are favorable will be incorrect from time to time, giving its forecasts a larger RMSE. Even with a larger RMSE, the second forecasting entity would likely be preferred to the first as it would have a better chance of predicting the danger of a tornado. Ultimately, it is unwise to categorize a forecast as poor without understanding the decision context behind it.

When determining what a "best" forecast is, two separate aspects must be studied. The first is to analyze if the forecast is doing what one would expect it to do. This constitutes looking into the RMSEs, variances, biases, and other quantitative numbers associated with the forecast. The other way would be to take those statistics, and figure out the underlying story in the decision context. Could there be a legitimate reason as to why there are substantial RMSEs, variances, and biases? What would that reason be? Forecasters sometimes respond to specific needs of their clientele so they may deviate from what they truly believe the realized value of the forecast may be. Basically, the decision context forces them to depart from what a strictly proper forecast scoring rule would require [Gneiting, 2007].

The reasons for such deviations may not be obvious to an outsider who is unfamiliar with the forecast process. In the instance of the NYISO, we will look into deviations from proper behavior within their electricity load forecasts and analyze them both quantitatively and qualitatively. Understanding the NYISO's load forecasts is important as they may have the ability to impact the energy market as well as the New York consumer that pays for electricity. While they might claim to be independent of the markets, their load forecasts are used to at least some degree when determining unit commitment and thus may have some impact on energy prices. Furthermore, by undertaking this research, it may become possible to figure out the NYISO forecasting model to some degree and offer insight in how to crack other forecasts where the elements that get plugged into the model are proprietary. If the quantity or event being forecasted has a significant bearing on any sort of financial market, getting ahead of the curve by understanding the respective forecast model could provide great monetary benefits to any individual that possesses such knowledge. Identifying forecast bias and efficiency is important as they are signs that all available forecast information is not being used. Accuracy in predicting load is also important for the ISO because it alleviates unnecessary electricity costs associated with erroneous expectations of load.

The purpose of this research is not to simply score and judge the forecasts made, but to quantify and document daily adjustments and biases to which they are subjected. In conducting the analysis, three main questions are asked:

1. Is the NYISO load forecasting system unbiased?
2. Is the NYISO forecasting system efficient?
3. Does new information arrive more rapidly as the forecast lead time shortens?

Chapter 2: Background: The NYISO's Load Forecasts

a. The New York Independent System Operator

The New York Independent System Operator is a "not-for-profit corporation regulated by the Federal Energy Regulatory Commission" that "manages New York's electricity transmission grid" [NYISO, 2009]. The NYISO further serves as a marketplace for wholesale electricity sales to various entities such as electric utility companies, load serving entities and other groups involved with energy such as demand response service providers. Because of deregulation within U.S. electricity markets around the turn of the century (beginning with experiments in California and Pennsylvania in 1996), the NYISO (along with other ISOs) have been in the forefront of overseeing such markets. In these markets, entities such as those previously mentioned can purchase and sell electricity produced by electric generators. The market holds various auctions including a day-ahead locational based marginal pricing (DA LBMP) auction for each hour of the following day. With these markets, electricity retailers can bid on power, which is more supply-side based. The NYISO oversees this entire process in an "economically efficient" manner while "foster(ing) regulatory certainty and market transparency" [NYISO, 2009].

b. The NYISO's Load Forecasts

Time-dependent load predictions (e.g. electrical load forecasts) are issued by independent system operators. The NYISO issues daily load forecasts for anticipated load at the top of each hour in megawatts, extending five days out. These forecasts are produced using linear regression models. Regression models are used instead of artificial neural nets (ANNs) because the resulting models are more "parsimonious" and tend to work better when used with hourly data [NYISO, 2001]. Such forecasts are used when assessing the value of load contracts between various utilities and generators. These load forecasts, categorized as short-term (STLF), are required to ensure system stability [Smith, 2003]. In this case, stability means keeping the grid operating smoothly and not being subject to blackouts or not having enough electricity for consumers. STLFs are also essential in the planning and scheduling of numerous operations [Chen, 2001]. In making these forecasts, the NYISO takes multiple components into account including the anticipated weather, the day of week, the time of day, economic conditions and historical data. However, the algorithms used to incorporate these variables into the forecasts are not made public and the exact process that the NYISO uses to create such forecasts is not known [NYISO, 2008].

The accuracy of load forecasts are very important for ISOs as they are crucial to the planning of their everyday operations as well as energy requirements at various times [Challa, 2005]. These load forecasts also have a strong bearing on the energy markets as well because they factor into bids participants make in their respective energy markets. On top of those incentives, inaccurate forecasts can have adverse financial impacts. When underforecasting occurs, utilities must buy reserve power on the spot market, resulting in extra costs. When overforecasting occurs, electrical generators must be compensated for startup costs of generators as well as other opportunity costs for their unused unit commitment [Hobbs, 1999]. Because the NYISO is a not-for-profit public benefit corporation regulated by the Federal Energy Regulatory Commission (FERC), one would expect no profit-driven interference in their forecasting; these forecasts are made with the interests of all players in mind.

While the NYISO does not explicitly describe the consequences of underforecasting and overforecasting, other entities such as PJM are clearer in establishing the differences. PJM is a regional transmission organization (RTO) that oversees wholesale electricity throughout portions of 13 states. Essentially, PJM is a different, larger version of the NYISO. With PJM, overforecasting load is typically more costly than underforecasting since overforecasting can lead to PJM paying "opportunity costs" to generators that are committed to run ahead of real-time, but are not actually dispatched. These opportunity cost payments can be larger than the cost of energy purchased on the spot market [PJM, 2008]. Despite their lack of transparency in those areas, the NYISO's regulations for reserve power indicate that overforecasting could be costly for utilities. For each hour, the NYISO sets a level of target reserve power (above the peak forecast) for the generators to have ready. If they set that level too high, the generators are typically compensated for commitment costs – start-up costs and any variable spinning (or "no-load") costs. Unfortunately, differences between that compensation and the cost of buying electricity on the spot market are not well defined by the NYISO.

c. Progressive Forecast Refinements

The NYISO issues 6 SRLFs for a given hour of a given day. The first forecast is issued five days in advance, while the final one is given on the day being forecasted. While the NYISO does not explicitly discuss the differences between progressive forecast refinements, it appears that they might simply run their models daily with updated weather data. We draw this conclusion because the forecasts are produced daily and the NYISO offers no other insight into other potential forecast adjustment methods. These forecast adjustments aid the NYISO in securing the commitment of electricity generators by sending a signal to the electricity market as to how much electricity supply is likely necessary for the forecasted day. For example, if the high temperature forecast for two days from now increases from 80º yesterday to 85º today, that would likely result in a positive revision of the load forecast. In turn, that sends a signal to electric generators and utilities that more commitment is necessary.

Chapter 3: A Model of the NYISO's Forecasting Process

The production of load forecasts and realizations is treated here as a discrete stochastic process y_t indexed by time t, where $t \geq 0$. The true value of y_t is revealed at time t without measurement error, and is known thereafter. Viewed from the perspective of an earlier period $s<t$, y_t is a random variable. We therefore say that the random variable y_t is *t-measurable*.

Alongside the electricity system, there exists a forecasting agency which in our case is the NYISO. Each period, the NYISO issues estimates of electricity load for each of the next several subsequent periods. We do not model explicitly the procedures by which the agency derives its forecasts (e.g., weather forecasts, economic conditions, etc.). Rather, we treat the forecast announcements as if they are generated by a set of stochastic processes that track, imperfectly, the original process y. Let K denote the maximum forecast lead time, which is also the total number of forecasts issued each period. For k = 1,…, K, let $\hat{y}_{t+k|t}$ denote the k-period-ahead forecast issued at time t. Here, $\hat{y}_{t+k|t}$ is the forecaster's point-estimate of the random variable y_{t+k} -- in other words, the electricity demand k periods hence. Although $\hat{y}_{t+k|t}$ is a prediction about the state of the world on a future date $t+k$, $\hat{y}_{t+k|t}$ is itself t-measurable. It will be convenient to adopt also the notational convention $\hat{y}_{t|t} = y_t$.

a. Bias in Forecasts

One of the important questions about a forecasting process concerns whether or not the forecasts are biased. Intuitively, an unbiased forecasting process is one that exhibits no predictable drift. Formally, following Clements (2005), a forecast $\hat{y}_{t+k|t}$ is said to be an *unbiased predictor* of y_{t+k} if $E_t[\hat{y}_{t+k|t} - y_{t+k}] = 0$. Here, $E_t[y_{t+k}] := E[y_{t+k} \mid \Omega_t]$, where Ω_t is the full set of information available. We say that the process $\hat{y}_{t+k|t}$ is an *unbiased k-period-ahead forecasting process* for y_t if $\hat{y}_{t+k|t}$ is an unbiased predictor of y_{t+k}, for all $t \geq 0$.

Associated with a given forecasting process $\hat{y}_{t+k}$, define the *error process* e_{t+k} by letting $e_{t+k|t} \equiv \hat{y}_{t+k|t} - y_{t+k}$, the error in the k-period-ahead forecast issued at time t. It is evident that the forecasting process $\hat{y}_{t+k}$ is unbiased if and only if $E_t[e_{t+k|t}] = 0$ for all k.

Let $\hat{y}_{t+k}$ be an unbiased forecasting process, and let t, $t+1$ be two forecast issue dates. The law of iterated expectations yields

$$E_t[\hat{y}_{t+k|t+1}] = E_t[E_{t+1}[y_{t+k}]] = E_t[y_{t+k}] = \hat{y}_{t+k|t} \tag{3.1}$$

Thus, for an unbiased forecasting process, each forecast announcement is an unbiased predictor of all subsequent forecasts.

To calculate the significance of the forecast bias, a t-statistic is compared to a null hypothesis that the bias is zero [Clements, 2005]. Holding k fixed, the t-statistic is established as:

$$\frac{(\frac{1}{T})\sum_{t=1}^{T} e_{t+k|t}}{\sqrt{(\frac{1}{T})(s)}} \tag{3.2}$$

where s^2 is:

$$\frac{1}{T-1}\sum_{t=1}^{T}(e_{t+k|t} - \bar{e}_{t+k|t})^2 \tag{3.3}$$

and $e_{t+k|t}$ is:

$$\sum_{j=0}^{(k-1)} u_{t+k|t+j} \tag{3.4}$$

T is the degrees of freedom while $u_{t+k|t}$ is difference between two respective forecast updates and $\bar{e}_{t+k|t}$ is the average of all those differences.

b. The Updating Process in Multi-Step Forecasts

In a typical forecasting process, predictions of near-term events have in general greater accuracy than predictions of events more remote. As time proceeds and the interval between the forecast and realization shortens, the forecaster has several opportunities to refine this estimate in light of new information. Thus, while $\hat{y}_{t+k|t}$ and $\hat{y}_{t+k|t+1}$ are both estimates of the same random variable y_{t+k}, we would expect the later estimate $\hat{y}_{t+k|t+1}$, to be more reliable.

We define an *update* (or *innovation*) to be the difference between two successive forecasts of the same event. Formally, for $k=0,\dots,K-1$, define the *updating process* $u_{t+k|t}$ by the formula $u_{t+k|t} = \hat{y}_{t+k|t} - \hat{y}_{t+k|t-1}$. By the notational convention $\hat{y}_{t|t} = y_t$, we have $u_{t|t} = y_t - \hat{y}_{t|t-1}$. Clearly $u_{t+k|t}$ is also the change in error between two successive forecasts:

$$\begin{aligned} u_{t+k|t} &= \hat{y}_{t+k|t} - \hat{y}_{t+k|t-1} \\ &= (\hat{y}_{t+k|t} - \hat{y}_{t+k}) - (\hat{y}_{t+k|t-1} - \hat{y}_{t+k}) \\ &= e_{t+k|t} - e_{t+k|t-1} \end{aligned}$$

With that, together with (3.4) above, we can see that if the forecast updating process is unbiased if and only if the corresponding error process is unbiased:

$$E[u_{t+k|t}] = E[e_{t+k|t}] - E[e_{t+k|t-1}] = 0. \tag{3.5}$$

c. Efficiency in Forecasts

An efficient forecasting process is characterized by the condition that all available information is used in each issued forecast. An implication is that the error process exhibits no serial correlation. For $k, j = 1,\ldots, K$, let $\sigma_{jk} = cov(\varepsilon_{t|t-j}, \varepsilon_{t|t-k})$. An efficient forecasting process will have the feature that $\sigma_{jk} = 0$ for all $j \neq k$.

To assess whether or not a forecast is efficient, one must check it for bias as well using the equation:

$$u_{t+k|t} = \alpha_k + v_{t+k|t} + w_{t+k|t} \tag{3.6}$$

where α_k is the bias term, $v_{t+k|t}$ is the efficiency-update term which by construction has mean zero, and $w_{t+k|t}$ is the error term that exhibits zero mean and zero correlation with $u_{t+k|t}$. Simply put, if $\alpha_k = 0$ then there is no bias in the updating process, whereas if $v_{t+k|t} = 0$, the updating process is efficient.

Efficiency is a sensible feature to expect of a forecasting system. To see why, suppose that on Tuesday a forecaster predicts Saturday's high temperature in London will be 21ºC. Yet suppose that the forecaster adds a caveat: she expects that on Wednesday she will raise her estimate of Saturday's high temperature, to 23ºC. There would something odd about this announcement. If the forecaster already possesses information that will help her to formulate the prediction she will issue on Wednesday, she should have already incorporated that information into the prediction she issues on Tuesday. More generally, if the direction of a future revision is predictable, then the forecaster is withholding potentially useful information. This practice is not consistent with the notion that a forecast announcement offers the forecaster's best estimate of future conditions.

A feature that is exhibited in some forecasting systems that can impact forecast efficiency is the *windshield wiper effect*. A forecasting system is said to possess this effect if subsequent updates constantly move in opposite directions. For example, suppose Wednesday's forecast for Sunday's high temperature is 75º. Then on Thursday, the forecast is adjusted to 80º, on Friday to 74º and on Saturday to 77º. Forecasting in such a manner has the potential to irritate users of the forecast because perpetual flip-flopping can make the forecast appear untrustworthy when being used for advanced planning. There are some that would feel this way even if it was known that the forecasting process is efficient. Because of this, some forecasters will avoid being subject to the windshield wiper effect, even if it means issuing inefficient forecasts. In instances like these, forecasters have to cater more towards their users rather than what the data actually provides.

Formally, a forecasting process is said to be *efficient* if $E_{t-1}[u_t] = 0$ for all t=1,…, T [Clements, 2005]. Forecasting efficiently requires that error terms have mean zero *conditional* on information available in the previous period. To see if forecasters are using all available information each time they issue a forecast, we would expect that the direction of one forecast update will contain no information about the direction of another, subsequent update. To do so, we use the following efficiency tests:

$$Cov(u_{t+k|t}\ ,\ u_{t+k|t+j}) = 0 \tag{3.7}$$

$$Cov(v_{t+k|t}\ ,\ v_{t+k|t+j}) = 0 \tag{3.8}$$

for $j \neq 0$.

If a forecast update provides information about the movement of a subsequent forecast, regression tests could yield more information as to how the previous forecasts are impacting the later update. Revisiting Eq. (3.6), we take the $v_{t+k|t}$ term and expand it to create a regression equation to further test efficiency in this following example:

$$v_{t+4|t+3} = (\beta_1)(u_{t+4|t}) + (\beta_2)(u_{t+4|t+1}) + (\beta_3)(u_{t+4|t+2}) + (\beta_4)(u_{t+4|t+3}) \quad (3.9)$$

where β is the correlation coefficient. To break this down in more simplistic terms, Table 1 displays what the regression equations would look like at various times in our update process.

TABLE 1. Regression Equations

Time	*Realization*
t	$v_{t+K\|t}$
$t+1$	$v_{t+K\|t+1} = (\beta_1)(u_{t+K\|t})$
$t+2$	$v_{t+K\|t+2} = (\beta_1)(u_{t+K\|t}) + (\beta_2)(u_{t+K\|t+1})$
$t+K$	$v_{t+K\|t+K} = \boldsymbol{\beta} \bullet \langle u_{t+k\|t+K} \rangle$

where $\boldsymbol{\beta}$ is a vector of all correlation coefficients and $\langle u_{t+k|t+K} \rangle$ is a vector of all efficiency-update terms. The greater the β term is, the greater the correlation between the respective forecast updates.

The way to eliminate inefficiencies in any forecasting system is to use all available information at a given time. In the instance where one is given a forecast (as opposed to creating one) with inefficiencies, any value other than zero for $E_{t-1}[u_t]$ must be compensated for. For example, assume that at time t one was to update a load forecast for time $t+K$ from 6500 MW to 6800M. Over a long period time, the forecasting process has been shown to further increase the load forecast for time $t+K$ at time $t+1$ if the revision was positive at time *t*. With that being the case, the load forecast today would be further revised upward, depending on the statistically determined correlation coefficient. If the β of date *t*'s update is 0.10 for the date $t+1$ forecast, the revision would be adjusted by another 30 MW to 6830MW (300MW adjustment * 0.10). This procedure corrects for autocorrelation: there would now be no signal from the time *t* forecast adjustment in $t+1$'s.[1]

d. Arrival of Forecast Information

In determining whether or not forecast information arrives more rapidly as the forecast lead time shortens, we use $Var(u_{t+k|t})$ as a reasonable measure of how much new information is added to the forecast between dates $t-1$ and date t where $\sigma_k^2 = Var(u_{t+k|t})$.

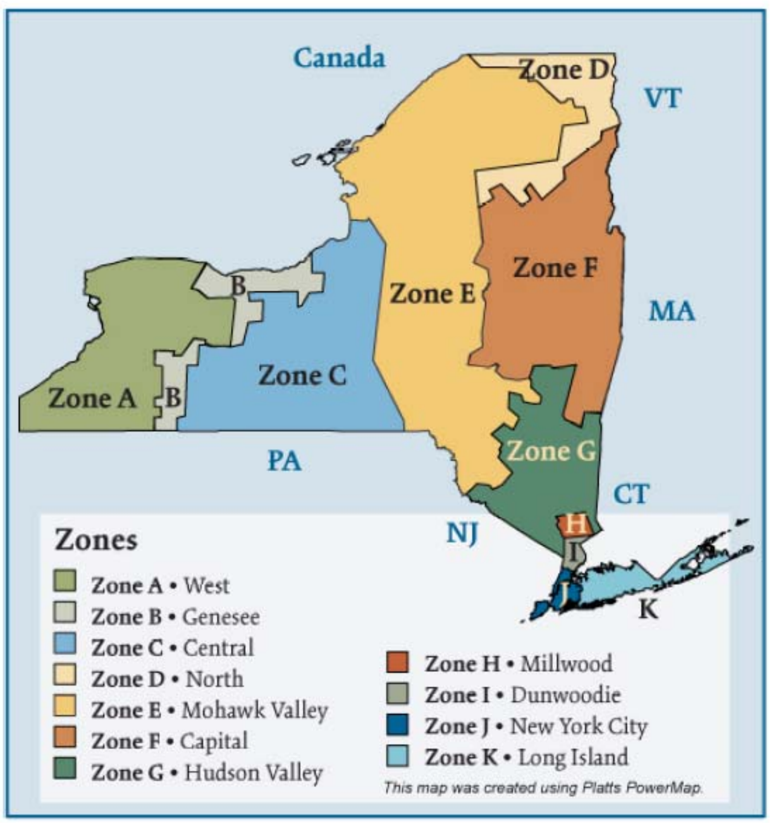
Our model of the forecast generation process may allow us to form hypotheses about the rate at which we expect new useful information to arrive, to the benefit of the forecaster. For example, for a temperature forecast, we might expect that information arrives with increasing rapidity as the forecast lead time shortens. This is because there is an increased level of uncertainty in the realized value the further away from the realization time. Forecasts are considerably more accurate immediately prior to the realization time than they are days away, indicating the majority of information arrives in the latter days. The corresponding hypothesis is that $\sigma_k \leq \sigma_j$ for all $k \leq j$, for all *t*.

[1] An item of note is that efficiency in this situation pertains solely to previous forecast updates. The forecasting system could theoretically still be inefficient, with other data such as weather forecast information containing information that could sharpen the load forecasts. We test to see that each forecast update is being made independent of all previous forecast updates

Chapter 4: Data

All data were collected from the NYISO's website (www.nyiso.com) which includes hourly, real-time electricity load data and hourly day-of through 5-day forecasts (6 forecasts in all) for the New York City region of the NYISO's power grid. This region is labeled 'Zone J' on the FERC map for NYISO zones below. Collected observations were from the top of all available hours. The period of record for the NYC region is 31 January 2005 -- 31 May 2008.

FIGURE 1. NYISO Electric Regions [FERC, 2008]



The fact that forecasts are not issued exactly at midnight introduces the tedious complication that the 24-hour forecasting cycle does not correspond exactly with the calendar day. This complication can for most purposes be safely ignored, as long as we are clear and consistent in our nomenclature. In this situation, a *period* corresponds to the twenty-four hours beginning immediately *after* the release of one set of forecast figures, up to and including the release of the next batch of figures. Once the date-t forecasts are issued (typically around 8:00 a.m.), the remaining hours of the current calendar day are treated as part of period $t+1$. This splitting of the calendar day is sensible from the vantage point of stochastic process modeling: the period-t numbers incorporate only information available at that time; events that have not yet occurred are assigned to a future period.

NYISO's maximum forecast lead time, five days, corresponds to a value K=6. Except as noted, in the discussion that follows a model period will correspond to a 24-hour forecasting cycle, not to a calendar date.

Chapter 5: Analysis and Results

If the NYISO intended to make accurate load forecasts, the covariance of the forecast updates would be minimal. That was found to be untrue; there are strong, positive covariances within the data as well as high variances and biases in the forecast changes.

a. Is the NYISO's Load Forecasting System Unbiased?

Across the entire data set, the day-of forecast averaged 26 MW lower than the realized load for a given hour, indicating a negative bias in the forecasts. During each year and all months, except for two (July and August), the NYISO's day-of forecasts averaged lower than the realized electricity load. The only time the NYISO did not underforecast on average was for daylight hours of summer months. Nearly every other time of year, the average hourly forecast was below the realized load (Table 2).

Throughout the entire period of record, the average difference between the hourly load forecast issued 5 days out and the one issued on the day of was +52 MW. Similar numbers are seen across all lead times. In June, September, and October, the total change in the NYISO's hourly load forecast averaged well over +100 MW, showing further bias in the forecast adjustments issued by the NYISO. The slight underforecasting bias can be seen in Figures 2-7.

While it is evident that these forecasts have at least some level of bias, testing to see if these biases hold any sort of statistical significance is crucial in determining their meaning in the overall evaluation of the forecast. The calculation of the t-statistic (significance) values were all far over the t-distribution critical threshold of 99%

confidence. We therefore reject the null hypothesis that the adjustments in load forecasts are unbiased (Table 3).

b. Is the NYISO's Forecasting Process Efficient?

Statistical analysis indicates positive covariance within the forecast updating process; a forecast revision in one direction (positive or negative) indicates that a subsequent revision will likely be in the same direction. For example, if a load forecast increases between days t and t+1, then that would indicate that the forecast revision between days t+1 and t+2 would also be positive. This covariance holds between nearly all forecast adjustments (Table 4). On the whole, all of the different forecast adjustments between respective days appeared to follow a normal distribution, with the mean being slightly positive (Figure 8).

Coefficients of variation associated with our covariance yield values suggesting these covariances are important in the context of the forecasting system. As seen in Table 5, some of these values were much larger than zero. Furthermore, the coefficients from our regression equations were also larger than zero, showing that previous forecast updates did have a bearing on future updates (Table 6). Regression equations were created combining our expansion of the $v_{t+K|t}$ and coupling that with Eq. (3.6). There was not a noticeable pattern to the regression coefficients other than that they were generally positive. Coefficients being positive indicate that a forecast revision will tend to move in the same direction as its previous one. Magnitudes of coefficients varied across the equations and the only regression equation (using all data) where coefficients were over 0.10 was the $u_{t+K|t+4}$ equation (forecast revision for the day-of forecast) where the coefficients were 0.19 and 0.11 for $u_{t+K|t+3}$ and $u_{t+K|t+2}$ respectively. While still small, they are not insignificant. It is noteworthy that the $u_{t+K|t+4}$ equation possessed the highest coefficients. Markets such as the day-ahead energy market have already been closed by the time that forecast is issued.

Because the four seasons produce different trends in forecast adjustments (as seen in Table 4 and Figure 9), it is important to see how regressions vary over the different seasons. While the coefficients differ across the many equations, one constant was that the $u_{t+K|t+4}$ regression equation typically yielded the highest values for coefficients with some reaching 0.28. In the end, these seasonal regressions further proved the significance in our covariance findings and that the NYISO tends to save their largest forecast revisions until the day-of forecast (Tables 7-10).

c. Does New Information Arrive More Rapidly as the Forecast Lead Time Shortens?

Statistical analysis of the data produced results that do not support the hypothesis that new information arrives more rapidly as the forecast lead time shortens. With the highest variances in forecast adjustments coming between the 5-day and 4-day forecasts, our data and hypothesis suggest that the most information is arriving between when those forecasts are made. Across all data, the hypothesis does not stand; however, it does hold for forecast adjustments made during 2005 (Figure 8). That year may be an aberration because the NYISO strays far from that pattern during all other time periods.

Variances in forecast adjustments during the summer and fall have a significant pull on that data as well. As shown in the graphs, the variance of forecast adjustments in fall, winter, and spring pale in comparison to the summer's. RMSEs in forecast changes were consistently high over the summer as well, spanning from 228 MW to 281 MW. Despite RMSEs and variances during summer being much higher, the other three seasons followed the same pattern of greatest variances of forecast adjustment occurring immediately after the initial forecast. Because of the large overall changes in forecast, the 3 months with the highest variances were expectedly the summer months of June, July, and August (Figure 9).

d. Further Observations

The accuracy of NYISO's forecasts did not improve over our period of record. Between 2005 and 2006, the average overall change in forecast and the overall difference between the initial forecast and realized load decreased by 56 MW to 17 MW and 51 MW to 41 MW respectively. The difference between the day-of forecast and realized load increased by 5 MW to 24 MW. The changes between the 2006 and 2007 forecasts differed, however. Both the average overall change in forecast and average difference between the 5-day forecast and realized load increased (by 44 MW and 24 MW, respectively) while the average difference between the day-of forecast and realized load decreased by 20 MW to an "accurate" 4 MW. NYISO flipped back and forth on improving and worsening their skill with long range and short range load forecasts over those few years. In 2008, while the data set was not complete at the time of analysis, it appeared to be on track to be the worst year yet for the NYISO's NYC forecasts with average errors currently over 100 MW through May.

During the summer months, there was a substantial difference between the initial 5-day and day-of forecast with June having the highest overall change in forecast (+157

MW) and August possessing the largest decrease (-54 MW). Substantial swings in forecast adjustments over the summer occur because changes in temperature during those months have a substantial impact on the amount of energy consumed. If the temperature forecast initially calls for a high of 70ºF but then later changes to 80ºF, the NYISO would be expecting much more energy consumption as a result of more air conditioning units being used. The opposite would occur when a forecasted high temperature is decreased from 85ºF to 80ºF. Variances of forecast adjustments are lower during other seasons such as the spring because a change in temperature forecast would have little impact on the load forecast. Even a substantial change in forecasted high temperature from 50ºF to 60ºF would not cause a notable difference in energy consumption. Amato (2005) depicts this fact by analyzing residential energy consumption as a function of temperature in Massachusetts; this relationship is reproduced as Figure 10.

TABLE 2. Forecast Changes and Differences (in MW)

	1	2	3
January	17.76	86.04	103.80
February	33.24	28.79	62.03
March	28.87	40.69	69.57
April	51.68	36.88	88.56
May	79.62	70.75	150.38
June	157.07	21.56	178.63
July	59.25	-28.00	31.25
August	-54.49	-58.30	-112.78
September	144.59	6.97	151.56
October	138.22	36.14	174.36
November	2.30	23.12	25.42
December	-26.95	24.57	-2.38
2005	73.37	18.65	92.02
2006	16.57	24.08	40.65
2007	60.44	3.82	64.26
2008	67.25	102.53	169.78
2005-2008	51.68	26.28	77.96
Winter	9.72	45.43	55.15
Spring	53.31	50.06	103.37
Summer	52.84	-22.07	30.76
Fall	97.33	22.15	119.48
Winter-low	8.81	33.85	42.66
Winter-peak	13.81	59.11	72.92
Spring-low	44.51	-16.03	28.48
Spring-peak	67.12	90.33	157.45
Summer-low	59.57	12.99	72.56
Summer-peak	66.22	-81.50	-15.28
Fall-low	72.06	-27.84	44.22
Fall-peak	120.49	29.80	150.30
Column 1:	Average Overall Change in Forecast		
Column 2:	Difference between Day of Forecast and Realized Load		
Column 3:	Average Overall Difference from Initial Forecast to Realized Load		

TABLE 3. T-Statistic Values

T-Statistic Values	
5-Day - 4-Day Adjustment	8.91
4-Day - 3-Day Adjustment	3.41
3-Day - 2-Day Adjustment	8.84
2-Day - 1-Day Adjustment	19.26
1-Day - Day-Of Adjustment	22.29
Day-Of - Realized Difference	17.39
5-Day - Realized Difference	5.95

TABLE 4. Covariance of Forecast Adjustments (in MW^2)

σ0^2						64472					
σ10	σ1^2					-674	19420				
σ20	σ21	σ2^2				524	3692	18912			
σ30	σ31	σ32	σ3^2			1119	2191	1134	18054		
σ40	σ41	σ42	σ43	σ4^2		2505	977	990	-515	17934	
σ50	σ51	σ52	σ53	σ54	σ5^2	433	1234	1805	877	347	26896

TABLE 5. Coefficients of Variation

ρ0^2						ρ0^2					
ρ10	ρ1^2					-0.02	ρ1^2				
ρ20	ρ21	ρ2^2				0.02	0.19	ρ2^2			
ρ30	ρ31	ρ32	ρ3^2			0.03	0.12	0.06	ρ3^2		
ρ40	ρ41	ρ42	ρ43	ρ4^2		0.07	0.05	0.05	-0.03	ρ4^2	
ρ50	ρ51	ρ52	ρ53	ρ54	ρ5^2	0.01	0.05	0.08	0.04	0.02	ρ5^2

TABLE 6. Regression Tests (All Data)

$u_{t+K\|t+1} = 2.65 + (0.02)(u_{t+K\|t}) + w_{t+K\|t+1}$
$u_{t+K\|t+2} = 7.13 + (-0.03)(u_{t+K\|t+1}) + (0.03)(u_{t+K\|t}) + w_{t+K\|t+2}$
$u_{t+K\|t+3} = 14.68 + (0.05)(u_{t+K\|t+2}) + (0.05)(u_{t+K\|t+1}) + (0.07)(u_{t+K\|t}) + w_{t+K\|t+3}$
$u_{t+K\|t+4} = 14.60 + (0.19)(u_{t+K\|t+3}) + (0.11)(u_{t+K\|t+2}) + (0.05)(u_{t+K\|t+1}) + (0.03)(u_{t+K\|t})$ $+ w_{t+K\|t+4}$
$u_{t+K\|t+K} = 25.54 + (-0.06)(u_{t+K\|t+4}) + (0.03)(u_{t+K\|t+3}) + (0.07)(u_{t+K\|t+2}) + (0.14)(u_{t+K\|t+1})$ $+ (0.01)(u_{t+K\|t}) + w_{t+K\|t+K}$

TABLE 7. Spring Regression Tests

$u_{t+K\|t+1} = 2.56 + (0.03)(u_{t+K\|t}) + w_{t+K\|t+1}$
$u_{t+K\|t+2} = 1.17 + (-0.04)(u_{t+K\|t+1}) + (0.12)(u_{t+K\|t}) + w_{t+K\|t+2}$
$u_{t+K\|t+3} = 14.39 + (0.10)(u_{t+K\|t+2}) + (0.10)(u_{t+K\|t+1}) + (0.21)(u_{t+K\|t}) + w_{t+K\|t+3}$
$u_{t+K\|t+4} = 18.08 + (-0.26)(u_{t+K\|t+3}) + (0.20)(u_{t+K\|t+2}) + (0.13)(u_{t+K\|t+1}) + (0.20)(u_{t+K\|t})$ $+ w_{t+K\|t+4}$
$u_{t+K\|t+K} = 51.16 + (-0.04)(u_{t+K\|t+4}) + (0.10)(u_{t+K\|t+3}) + (-0.14)(u_{t+K\|t+2}) + (-0.22)(u_{t+K\|t+1})$ $+ (-0.07)(u_{t+K\|t}) + w_{t+K\|t+K}$

TABLE 8. Winter Regression Tests

$u_{t+K\|t+1} = 2.94 + (-0.06)(u_{t+K\|t}) + w_{t+K\|t+1}$
$u_{t+K\|t+2} = 5.69 + (0.05)(u_{t+K\|t+1}) + (0.12)(u_{t+K\|t}) + w_{t+K\|t+2}$
$u_{t+K\|t+3} = 8.07 + (-0.04)(u_{t+K\|t+2}) + (0.02)(u_{t+K\|t+1}) + (0.04)(u_{t+K\|t}) + w_{t+K\|t+3}$
$u_{t+K\|t+4} = 4.47 + (-0.07)(u_{t+K\|t+3}) + (0.06)(u_{t+K\|t+2}) + (0.04)(u_{t+K\|t+1}) + (0.04)(u_{t+K\|t})$ $+ w_{t+K\|t+4}$
$u_{t+K\|t+K} = 47.29 + (-0.12)(u_{t+K\|t+4}) + (-0.03)(u_{t+K\|t+3}) + (0.02)(u_{t+K\|t+2}) + (-0.09)(u_{t+K\|t+1})$ $+ (0.19)(u_{t+K\|t}) + w_{t+K\|t+K}$

TABLE 9. Fall Regression Tests

$u_{t+K\|t+1} = 20.17 + (0.05)(u_{t+K\|t}) + w_{t+K\|t+1}$
$u_{t+K\|t+2} = 14.45 + (0.07)(u_{t+K\|t+1}) + (0.10)(u_{t+K\|t}) + w_{t+K\|t+2}$
$u_{t+K\|t+3} = 12.58 + (0.04)(u_{t+K\|t+2}) + (0.05)(u_{t+K\|t+1}) + (0.10)(u_{t+K\|t}) + w_{t+K\|t+3}$
$u_{t+K\|t+4} = 8.70 + (0.13)(u_{t+K\|t+3}) + (0.05)(u_{t+K\|t+2}) + (0)(u_{t+K\|t+1}) + (0.08)(u_{t+K\|t})$ $+ w_{t+K\|t+4}$
$u_{t+K\|t+K} = 16.68 + (0.05)(u_{t+K\|t+4}) + (-0.07)(u_{t+K\|t+3}) + (0.06)(u_{t+K\|t+2}) + (0.18)(u_{t+K\|t+1})$ $+ (0.03)(u_{t+K\|t}) + w_{t+K\|t+K}$

TABLE 10. Summer Regression Tests

$u_{t+K\|t+1} = -14.68 + (0)(u_{t+K\|t}) + w_{t+K\|t+1}$
$u_{t+K\|t+2} = 2.05 + (-0.07)(u_{t+K\|t+1}) + (-0.01)(u_{t+K\|t}) + w_{t+K\|t+2}$
$u_{t+K\|t+3} = 20.75 + (0.06)(u_{t+K\|t+2}) + (0.06)(u_{t+K\|t+1}) + (0.04)(u_{t+K\|t}) + w_{t+K\|t+3}$
$u_{t+K\|t+4} = 34.45 + (0.28)(u_{t+K\|t+3}) + (0.12)(u_{t+K\|t+2}) + (0.06)(u_{t+K\|t+1}) + (0.01)(u_{t+K\|t}) + w_{t+K\|t+4}$
$u_{t+K\|t+K} = -18.57 + (-0.05)(u_{t+K\|t+4}) + (0.05)(u_{t+K\|t+3}) + (0.10)(u_{t+K\|t+2}) + (0.17)(u_{t+K\|t+1}) + (0)(u_{t+K\|t}) + w_{t+K\|t+K}$

FIGURE 2. Histogram of Day-5 to Day-4 Forecast Adjustments (in MW)

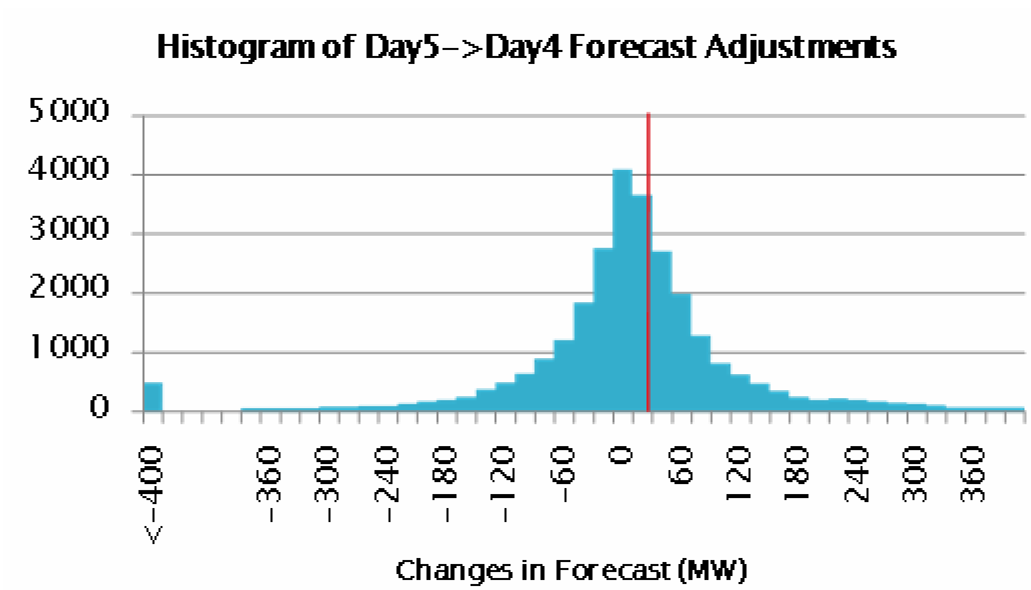


FIGURE 3. Histogram of Day-4 to Day-3 Forecast Adjustments (in MW)

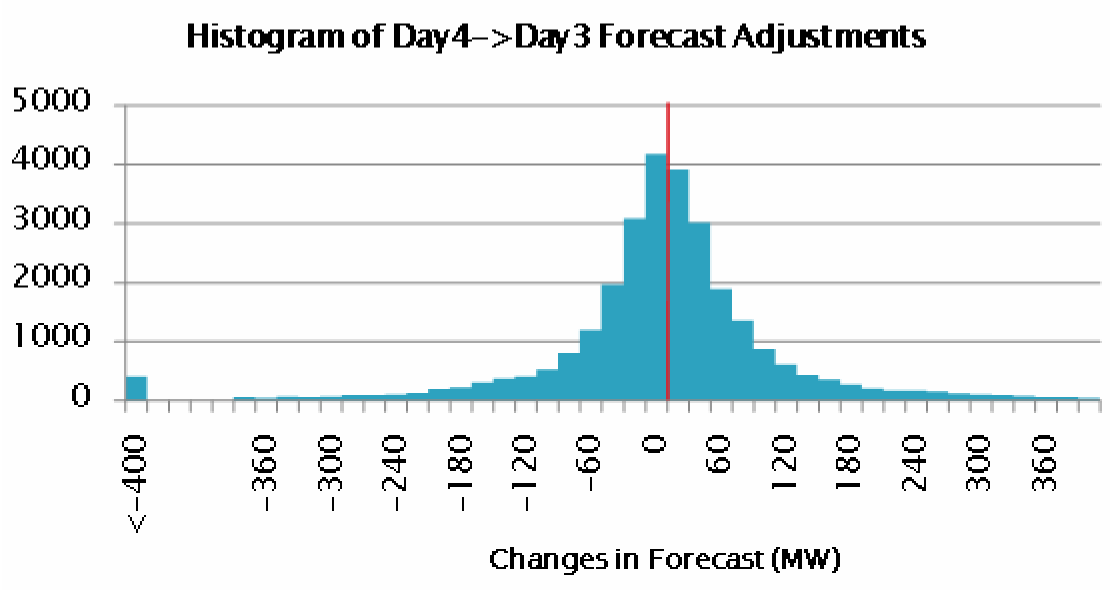


FIGURE 4. Histogram of Day-3 to Day-2 Forecast Adjustments (in MW)

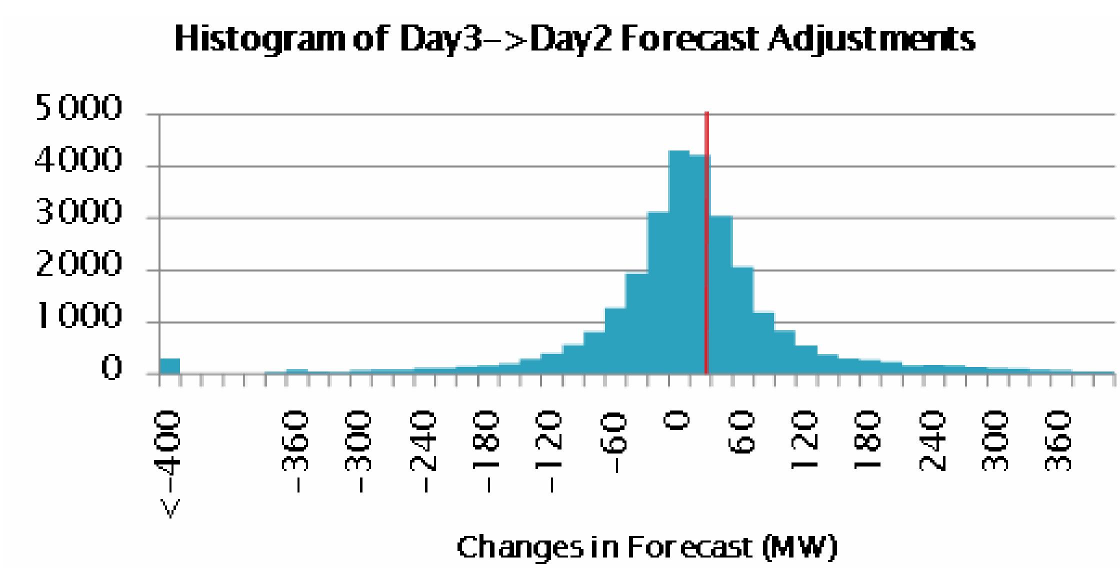


FIGURE 5. Histogram of Day-2 to Day-1 Forecast Adjustments (in MW)

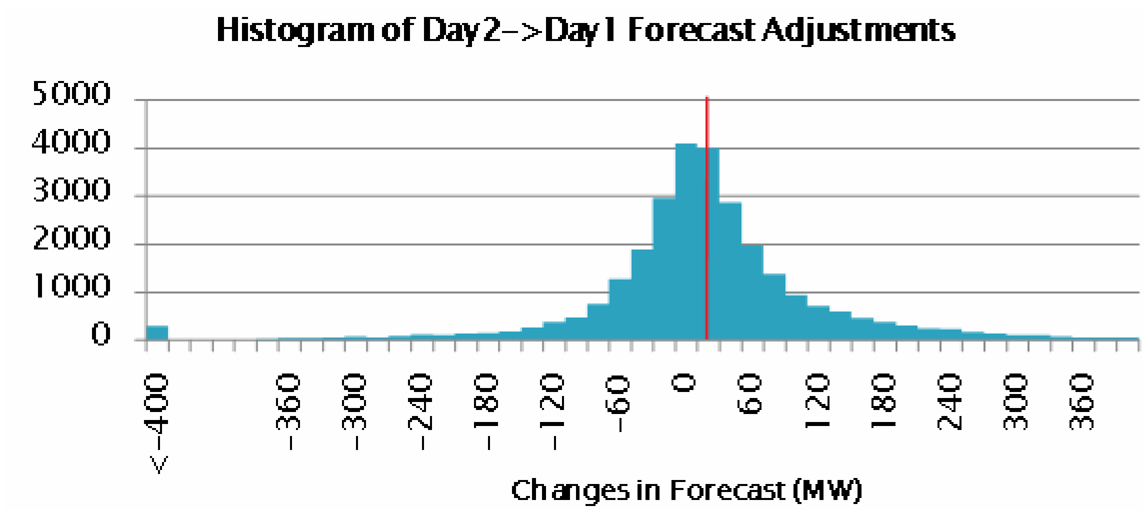


FIGURE 6. Histogram of Day-1 to Day-Of Forecast Adjustments (in MW)

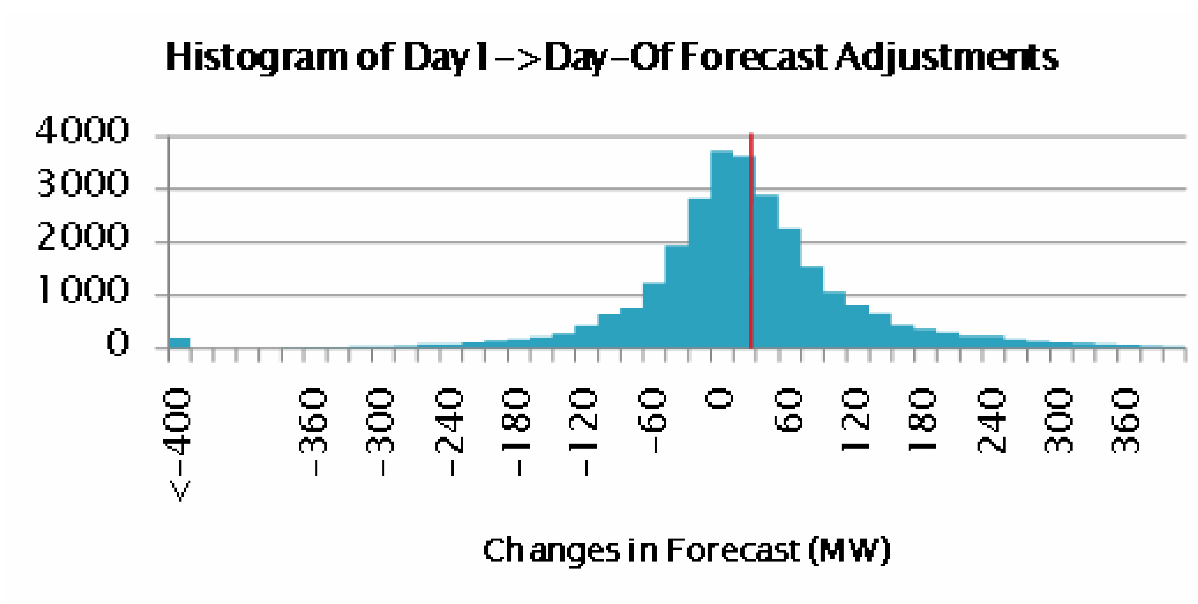


FIGURE 7. Histogram of Day-Of to Actual Load Difference (in MW)

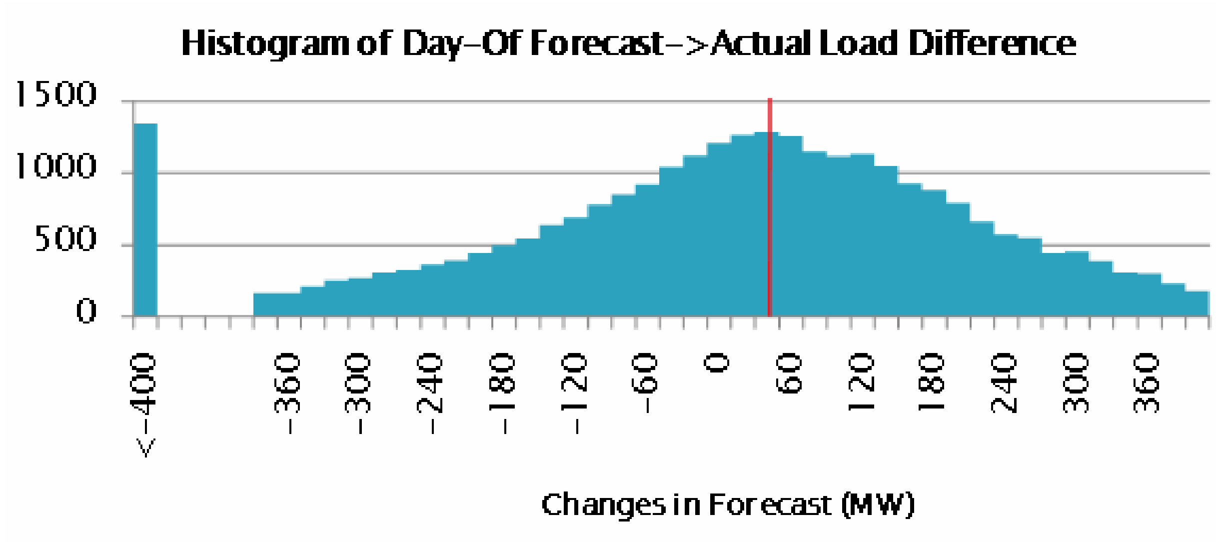


FIGURE 8. Annual Variance of Forecast Adjustments

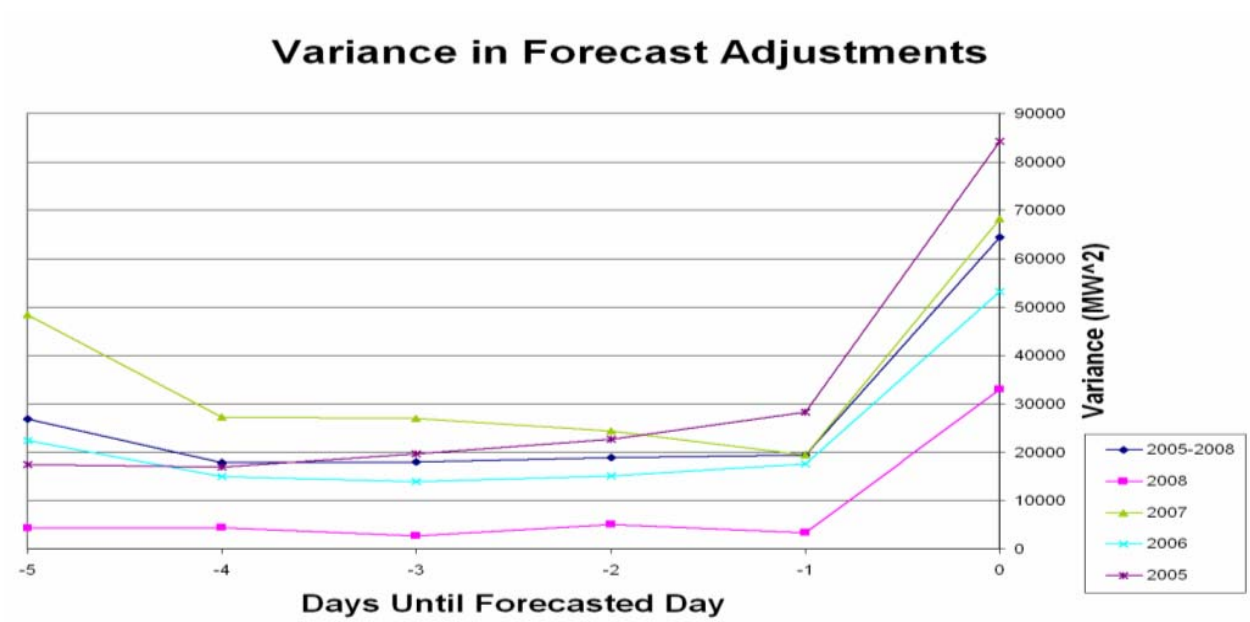


FIGURE 9. Seasonal Variance of Forecast Adjustments (in MW^2)

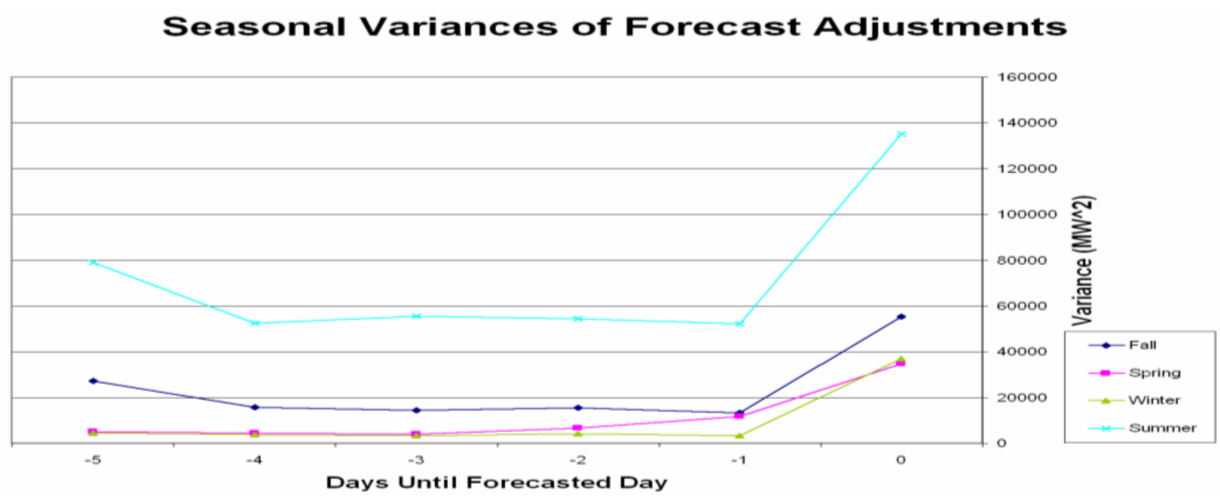
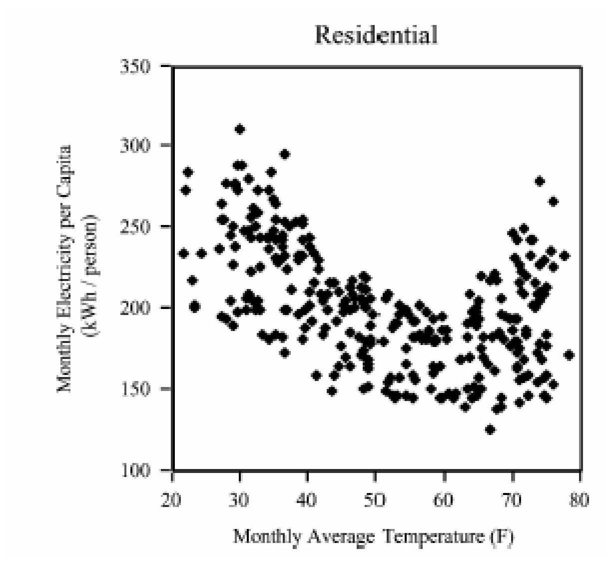


FIGURE 10. Monthly Electricity per Capita as a Function of Temperature [Amato, 2005]



Chapter 6: Conclusions

a. Possible Motivation for Underforecasting

It is possible that the NYISO underforecasts to prevent themselves from overcommiting units of electricity. If the NYISO was to issue a forecast 2-days from the realization time that ended up being +800 MW from the actual load, it is possible that they ended up requesting the commitment of units that ended up not being used. As discussed earlier in the paper, that results in opportunity and startup costs to be paid.

Once generators are committed, they are difficult to decommit. Even if they are, an opportunity cost would still have to be paid. Altalo [2004] writes about a 350-400 MW electricity generator in Boston that needs 24-hour notice to come online. On top of that, the generator must run for a few days before it can be shut down. If an ISO were to commit this generator and not use it, there would undoubtedly be an opportunity cost associated with it.

This observation leads to an understanding that it might be financially 'smarter' to underforecast load by minimal amounts. In other words, it may be optimal for the NYISO to issue a load forecast close to the expected value, but with a very low likehood of exceeding the realized load. One might call this approach "accurate underforecasting." While the aforementioned generator needs a whole day to start up, there are other smaller units that can have electricity online in hours or less. Minor underforecasting is therefore not a huge problem for the NYISO. Also, as previously stated with reference to PJM, it is typically cheaper to buy electricity on the spot market than to cover opportunity costs of over-committment. Because of this consideration, "accurate underforecasting" could potentially be an ideal strategy for the NYISO and other entities trying to guage unit commitment days in advance.

b. Further Observations

The NYISO states that they use linear regression models with parameters estimated using ordinary least squares. It can be proven, however, that such a procedure is statistically efficient. Our statistical analysis indicates that the NYISO's forecasting process is not efficient, suggesting that there is potentially some further "tweaking" of the forecasts. It is uncertain where additional adjustments may be coming from. It is possible that the weather forecasts the NYISO uses as inputs to their load forecasting models are themselves generated from a biased and inefficient forecasting process. If so, then the NYISO's load forecasting system would be inefficient in turn.

It appears that the NYISO has room for improvement with their forecasts if the intent is to issue strictly proper forecasts. High variances in forecast adjustments occur more over the summer months than during any other season, because forecast adjustments involving uncomfortably warm temperatures have a noticeable impact on the demand for electricity. For all seasons, there is a jump in variance from the day-of forecast to the realized load (Figures 2-7).

A conjecture as to why there is a large dip in the variance of forecast adjustments between the 5-day to 4-day and the 4-day to 3-day forecast is that the NYISO's initial 5-day forecast is based on climatology. It appears that four days out, the NYISO begins to look at incoming data and tweak their initial forecast more so than they adjust any of their other forecasts. This hypothesis cannot be proven given the limited data and information publicly available, but would be worth further investigation. An alternative explanation would trace the source of inefficiency back to their incoming data..

Strong autocorrelations in the data may indicate that not all new information is being used in subsequent forecast adjustments. If all new information was being used, then these foreseeable revisions would be made in earlier forecasts. Thus, the NYISO could be leaving out forecast information and making certain decisions knowing that forecasts will be changed in the following days.

Furthermore, forecast adjustments in August saw a very notable reduction from the initial forecast. This could mean that the NYISO is hedging their forecasts so as to not underforecast when the temperature could potentially get hotter than expected. As explained earlier, if the realized high temperture was 85ºF instead of an expected 80ºF, there would subsequently be more electricity consumption than expected. The NYISO does not want to have an issue revolving around not having enough power to distribute. With an average high of 81ºF [The Weather Channel, 2009] in August for New York City, this is a common predicament. Such problems could cause blackouts and miserable conditions for many.

Similar reasoning could be used to explain why June sees a substantial increase in forecast adjustments. With the average high temperature for New York City in June being 79ºF [The Weather Channel, 2009], then there are obviously days where residents of the City use their air conditioning units. The NYISO could be issuing an initial forecast, not expecting conditions to be favorable for air conditioning but end up changing that assertion as the realization time nears. Because of that, they would again not want to be caught in a situation where not nearly enough electricity is available.

Another item of note was that the spring and winter saw more negative correlation coefficients than those in the summer and fall. This is likely because the spring and winter generally do not see large variances in forecast adjustments as drastic as those in the summer and fall. While previous forecasts typically hint towards future forecasts moving in a similar manner, swings are not as drastic and are more small-scale during the spring and winter months. Because of this, the summer and fall months have much more of a pull on overall data because their variances are much larger than those during the spring and winter.

It is possible that the NYISO does not take all forecast information into account when making electricity load forecasts. Were the NYISO attempting to make accurate forecasts, biases would be accounted for. While the worst forecast errors were only approximately 2% of the realized load, the biases and variances are both notable and consistent. However, it has been exhibited in Bunn (1985) that even a very small adjustment in forecast error can have major financial impacts. Bunn found that a mere 1% increase in forecast error caused a £10 million increase in operating cost for one British electric utility company. Furthermore, it appears the NYISO has done little over time to correct the pattern. Because of this, we conclude that the NYISO may not be issuing strictly proper forecasts, but rather *risk-adjusted* forecasts for electricity load. They forecast in a manner that attempts to avoid the payment of costs that do not end up benefiting any parties.

For future work, it would be beneficial to discover exactly how much money is being lost by over-forecasting or under-forecasting, or if in fact the risk-adjusted forecasts are saving money for generators and utilities. In doing so, it will explain how much a forecast error costs the electricity generators and utilities, providing more detailed insight as to why the NYISO forecasts the way it does.

REFERENCES

Altalo, M., M. Hale, 2004: Turning weather forecasts into business forecasts: *Environmental Finance*, May 2004.

Amato, A. D., M. Ruth, P. Kirshen, J. Horwitz, 2005: Regional Energy Demand Responses to Climate Change: Methodology and Application to the Commonwealth of Massachusetts. *Climatic Change*, **71**, 175-201.

Bunn, D., E. Farmer, 1985: Economic and Operational Context of Electric Load Prediction. *Comparative Models for Electrical Load Forecasting*. John Wiley and Sons Inc.

Challa, B.P., S. Challa, R. Chakravorty, S.K. Deshpande, D. Sharma, 2005: A Novel Approach for Electrical Load Forecasting Using Distributed Sensor Networks. *3rd International Conference on Intelligent Sensing and Information Processing*, pp. 189-194.

Chen, Hong, C. A. Cañizares, A. Singh, 2001: ANN-based Short-Term Load Forecasting in Electricity Markets. *Power Engineering Society Winter Meeting*, **2**, 411-415.

Clements, Michael P., 2005: Evaluating econometric forecast of economic and financial variables. *Palgrave texts in econometrics,* Houndmills, Basingstoke, Hampshire; New York, N.Y.: Palgrave Macmillan.

Federal Energy Regulatory Commission, 2007: New York Electric Market: Overview and Focal Points. <http://www.ferc.gov/market-oversight/mkt-electric/new-york/elec-ny-reg-des.pdf>

Gneiting, T., A. E. Raftery, 2007: Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, **102**, 359-378.

Hobbs, Benjamin F., S. Jitprapaikulsarn, D. J. Maratukulam, S. Konda, K. A. Loparo, 1999: Analysis of the Value for Unit Commitment of Improved Load Forecasts. *IEEE Transactions on Power Systems*, **14**, 4, 1342-1348 .

PJM Forward Market Operations, 24 November 2008: PJM Manual 11: Scheduling Operations.

The New York Independent System Operator, 2009: About Us. <http://www.nyiso.com/public/company/about_us/index.jsp>

The New York Independent System Operator, 2008: Ancillary Services Manual, pp. 4-8 – 4-13.

The New York Independent System Operator, 2001: Day-Ahead Scheduling Manual, Attachment B, pp. 1-8.

The Weather Channel, 2009: Average Weather for New York, NY. <http://www.weather.com/outlook/homeandgarden/schoolday/wxclimatology/monthly/graph/USNY0996?from=36hr_bottomnav_schoolday>

Smith, Michael, 2003: Electricity Load and Price Forecasting Using Statistical Methods and Models. *Second Moment*®. <http://www.secondmoment.org/articles/electricity.php>