Forecast Model Structures of the ISO New England Long-Run Energy and Seasonal Peak Load Forecasts

for the 2015 CELT Report and 2015 Regional System Plan

Contents

Peak Load and Energy Forecast Modeling Procedures 2015

1. Introduction

ISO-New England forecasts annual energy and monthly and seasonal peak loads for the region and for each of the six New England states. Creating the forecast is a three-stage process: First, annual net energy for load (NEL) is modeled as a function of economic drivers. Second, daily peak loads are modeled as a function of weather and annualized monthly NEL, which represents the underlying economic/demographic processes that influence peak load growth. Finally, the models are used to forecast annual energy and monthly and seasonal peak loads.

2. Energy Forecasts

As in years past, the 2015 energy forecast is produced from annual models of total energy consumption in the ISO-NE control area and the states. The goal is to specify econometric models with a sound economic structure that capture the underlying relationships among the input variables, to predict electricity consumption as accurately as possible. The forecast includes exogenous estimates of the impact of new Federal Electric Appliance Standards (2013 going forward), energy efficiency resources, and solar energy, which cannot be captured sufficiently within an econometric framework[1].

2.1. The econometric model

The energy forecasting models have the same fundamental structure, with variations across states. The basic theoretical model is as follows:

$$Energy_t = f(Energy_{t-1}, Economy_t, EnergyPrice_t, Weather_t, X_t)\text{, where:}$$

Energy = Annual Net Energy for Load, net of Passive Demand Resources (NEL_PDR)

Economy = Economic activity, represented by Gross Regional/State Product adjusted for inflation (RGSP), or by Personal Income adjusted for inflation (RPI). Economic data are from Moody's Economy.Com. Gross domestic product and personal income are adjusted for inflation by Moody's price deflator series.

EnergyPrice = Average Annual Price of Energy adjusted for inflation (RPER). The coefficient on the price of energy is not highly significant in most of the models. While theoretically interesting, it is not very relevant in this construction, because:

(1) NEL_PDR is aggregated across all sectors (residential, commercial, and industrial.) Different factors affect the demand for energy in each sector, and it is not surprising that an average energy price cannot capture the unique behavior of individuals in each sector.

[1]Econometric models are estimated on historical data. Federal Electric Appliance Standards, energy efficiency, and solar energy are too new to be accounted for adequately in a historical model

(2) Price is averaged over the year. It seems reasonable to expect an average *ex post* annual price to have minimal impact on annual aggregate energy demand.

(3) Nevertheless, RPER is retained in most equations[2].

Weather = Weather is represented by two variables: Annual Heating Degree Days (HDD) and Annual Cooling Degree Days (CDD). While both variables are included in most models, one or both may be considered statistically “insignificant” in some cases[3,4].

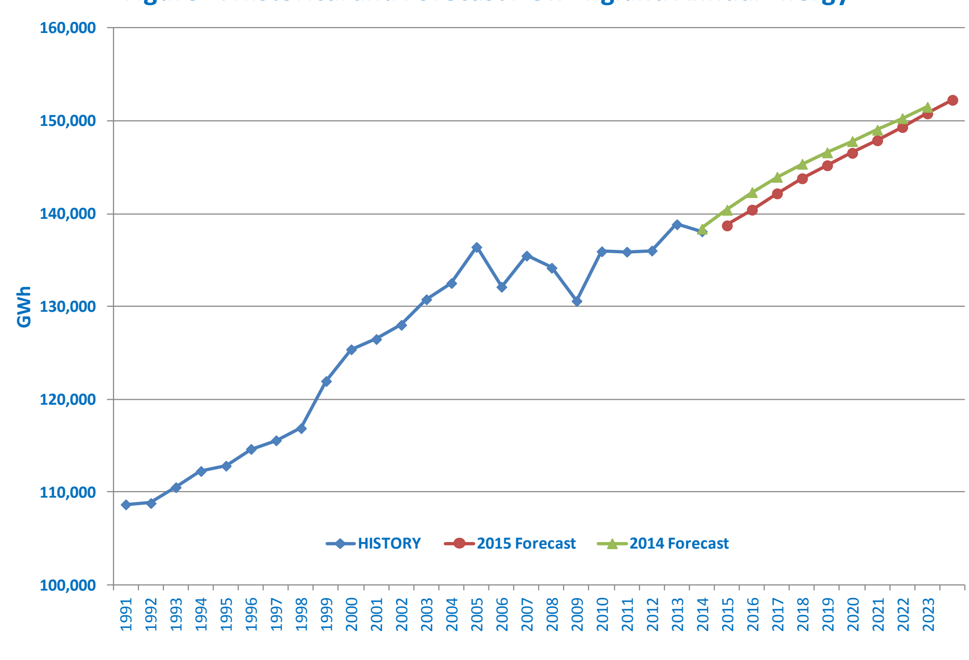
X = Unobservable variables that affect energy demand. Binary variables for specific years are included in most models. The relevant variables are determined by examining the residuals: (observed NEL_PDR$_t$ – modeled NEL_PDR$_t$). Large outliers are addressed by including a dummy variable for that year.

Sample Period: 1990-2014

3. Observations Regarding Historical and Projected Energy Consumption

Historical annual energy demand in the ISO-NE control area from 1991-2014 is depicted in Figure 1, together with a comparison of the 2014 and 2015 forecasts. The historical series shows a fairly steady upward trend through 2005. The year-to-year change became more volatile thereafter.

Figure 1: Historical and Forecast New England Annual Energy



[2] Vermont is the exception, where the coefficient has very little statistical significance and is close to zero.

[3] As discussed in footnote 9, variables may be retained in an estimating equation because they are important *a priori*.

[4] Vermont again is the exception. The coefficients on both CDD and HDD were virtually zero.

The 2015 forecast projects weather-normalized energy demand to grow slightly slower than expected in the 2014 forecast. Increasing demand response, energy efficiency resources and solar resources, together with related public policies, amplify the uncertainty around the energy forecast.

The long-run energy growth rate is forecasted to be about 1.0% in New England. The 2015 energy forecast incorporates the annual energy savings expected from the introduction of Federal Appliance Efficiency Standards in 2013, which reduces the forecast of electricity demand in 2015 by 1696 GWh. The impact of the Standards is projected to increase gradually and amount to about 2500 GWh by the end of the forecast horizon.

4. Transition: Annual Energy to Annualized Monthly Energy

As input into the peak load models, historical annualized NEL is calculated for each month by summing monthly NEL over the previous 12-month period, including the current month. As a peak load forecast driver, the annual energy forecasts are disaggregated into annualized monthly values, as follows:

(1) The annual energy forecast is prepared.

(2) Weather normalized annualized monthly energy is calculated for the latest historical year (2014).

(3) Using the forecasting model, annual energy is estimated for the latest historical year (2014) (WNEL2014).

(4) Annual growth rates are calculated:

$$G_t = \frac{NEL_t}{WNEL_{2014}}, \text{ t=2015,…,2024}$$

(5) G_t is applied to annualized monthly energy:

$NEL\_MA_{t,m} = NEL\_MA_{t-1,m} * G_t$,
where t=year and m=month

(6) $NEL\_MA_{t,m}$ serves as the energy driver for the monthly and seasonal peak models.

5. Peak Loads

5.1. Peak Load Forecast Distributions

Peak load forecast distributions are created by combining output from the daily peak load models with energy forecasts and weekly distributions of weather variables over 35 years. Heating degree days (HDD) correspond to the heating season (October-April), while the weighted temperature-humidity index (WTHI) is used for the cooling season (May-September).

5.1.1. Peak Load Distributions and Weather

The expected weather associated with the seasonal peak is considered to be the 50th percentile of the pertinent week's historical weather distribution. The monthly peak load is expected to occur at the weather associated with the 20th percentile of the pertinent week's weather distribution. The pertinent week is the week of the month or season with the most extreme weather. For transmission planning purposes, peak load distributions are developed for each week of the forecast horizon.

5.2. Daily Peak Load Models

Econometric models of daily peak loads are estimated for the New England region and each of the six New England states, for each season and nine months (July and the summer peak model coincide, and there is one winter model for December/January.) Altogether, there are 11 models for each area, for a total of 77 daily peak load models. While the models share a common theoretical basis, they are individually adjusted for the unique characteristics of the region/state and the sample period.

Fundamental Drivers. NEL (converted to annualized monthly energy) and weather variables comprise the foundation of the peak load models. Weather is the predominant observable cause of day-to-day variations in peak load, and also differentiates seasons. Energy is the base load, and represents underlying economic and demographic drivers.

Dummy Variables. The sample period comprises *all* days of the week, including holidays and weekends, while monthly/seasonal peak loads occur only on non-holiday weekdays. Including all days in the sample increases the sample size and reduces the number of “gaps” in the data. Significant gaps already exist because separate models are estimated for each month and season. Dummy variables account for holidays and weekends.

Sample Size. The sample period needs to span enough years to capture significant variation in the weather; i.e. an abnormally warm or cool year cannot be allowed to unduly influence model predictions. The sample period also must be short enough to expect reasonable stability in the relationship between peak loads and the regressors over the period of analysis.[5]

Peak Load Model. The basic peak load model is a nonlinear function of energy and weather, expressed as:

$$Peak\ Load_d = f(NEL\_MA_{t,m}, W_d, D_{w,d}, D_{h,d}, X_d) \quad \text{where:}$$

$Peak\ Load_d$ = Peak Load on day *d*

$NEL\_MA_{t,m}$ = Annualized monthly Net Energy for Load for month *m* in year *t* (see page 1-2 for details).

W_d = Weather at the peak load hour on day *d*

= $(WTHI_d-55)^2$ for the months May-September

= $(65-db_d)^2$ for the months October-April, with some differences in base temperature in April and October.

[5] Analysis suggests 2000-2001 as a reasonable point for the beginning of the sample period for most models. For a few models, the sample stretches back into the late 1990s.

$WTHI_d$= 3-day weighted temperature-humidity index (THI) measured at the hour of the daily peak loads:

$$WTHI_d = \left\{\frac{[10*THI_d + 5*THI_{d-1} + 2*THI_{d-2}]}{17}\right\} - 55 \text{ , and}$$

$$THI_d = 0.5 * DryBulbTemp_d + 0.3 * DewPointTemp_d + 15$$

$D_{w,d}$ = Dummy variable: 1 if day *d* is a weekend, 0 otherwise.

$D_{h,d}$ = Dummy variable: 1 if day *d* is a holiday, 0 otherwise. Holidays take precedence; if day *d* is both a weekend day and a holiday, $D_{w,d} = 0$ and $D_{h,d} = 1$.

X_d = Vector of other (unobservable) variables explaining daily peak loads

While Energy and Weather variables explain most of the trend and variation in Peak Load, there are many other largely unknowable factors (*X*) that can be included in the model only by proxy, if at all.

The basic non-linear estimating equation with autoregressive error structure is specified as:

$\text{Peak Load}_d = b_0 + b_1 * NEL\_MA_{t,m} + b_2 * W_d + b_3 * D_{w,d} + b_4 * D_{h,d} + \hat{e}_d$

$\hat{e}_d$ is the error term (residual), which follows an autoregressive process:
$\hat{e}_d = f(e_{d-1}, e_{d-2}, \ldots , e_{d-n})$

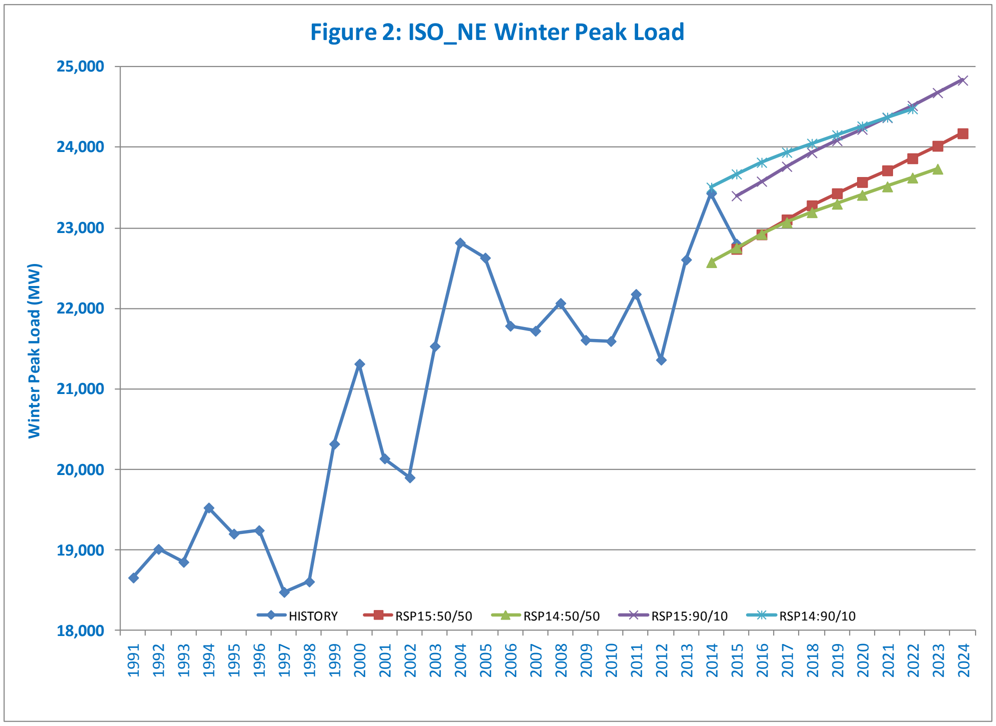
6. Observations Regarding the Projected Winter Peak

The 2015 forecast projects modest growth over the next 10-year period. The long-run 50/50 winter peak load growth rate is projected to be 0.7%.

The slow economic recovery has had a less severe impact on the winter peak load than on the summer peak load, because the winter peak is determined mainly by residential load. As such, it occurs late in the afternoon (6 P.M.) and has less direct exposure to industrial and commercial demand.

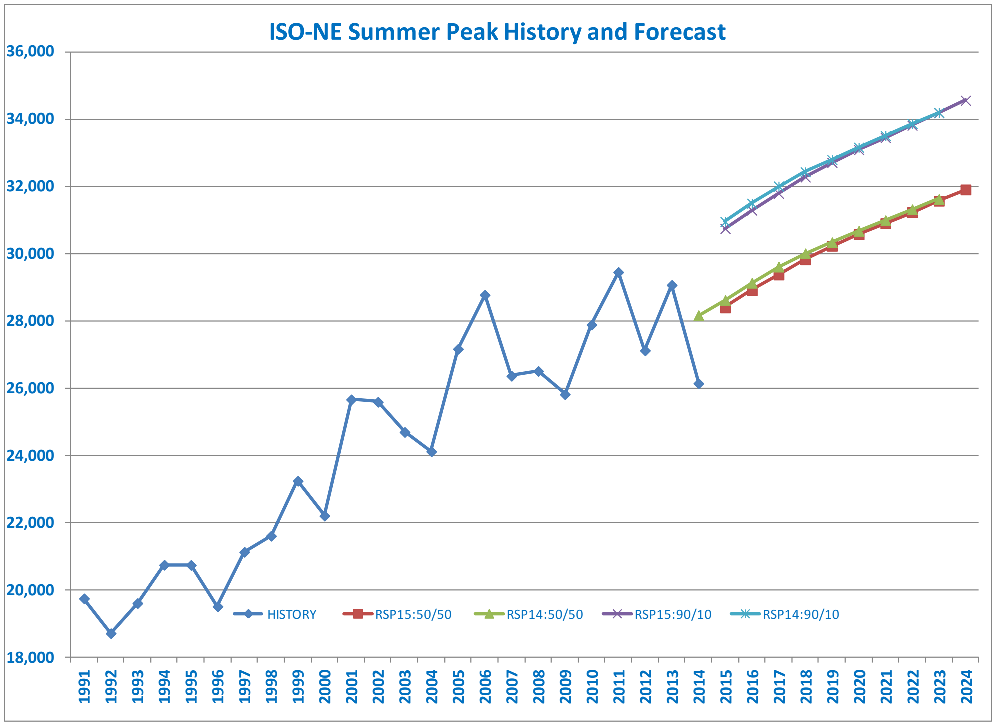
The base load, represented by the energy forecast, is the major driver of the winter peak load forecast, since the non-weather-sensitive part of the load accounts for 85% of the winter peak.

Figure 2 shows history, and compares the 50/50 and 90/10 winter peak forecasts for the 2014 and 2015 cycles for the ISO-NE control area.



7. Observations Regarding the Projected Summer Peak

Figure 3 shows history, and 50/50 and 90/10 forecasts for the 2014 and 2015 cycles for the ISO-NE control area. As this chart illustrates, the 2015 summer peak load forecast is very similar to the 2014 forecast. The 2015 forecast projects an average growth rate for the 50/50 peak of 1.3 percent over the next 10-years, the same as the 2014 forecast. The trajectories of both forecasts are mainly influenced by the energy driver, which in turn depends on economic drivers. An increasing penetration of air-conditioners is also reflected in the peak load forecast. The non-weather-sensitive part of the load accounts for 46% of the summer peak. Weather sensitive load accounts for 54%.


ISO-NE Summer Peak History and Forecast
36,000
34,000
32,000
30,000
28,000
26,000
24,000
22,000
20,000
18,000
HISTORY
RSP15:50/50
RSP14:50/50
RSP15:90/10
RSP14:90/10
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024

Appendix A: Forecasting Methodology, Evaluation and Testing[6]

A.1. Peak Load Forecasting Methodology, Evaluation and Testing

The process for developing econometric-based peak load forecasting models is discussed in this appendix. Developing the equations used to forecast seasonal and monthly peak loads is an iterative process comprised of the following steps:

(1) Informed by past years' models, a nonlinear econometric model with an autoregressive error structure is specified and parameters are estimated.

(2) The residuals are examined for extreme outliers.

(3) The residuals are examined to determine if they exhibit any trends.

(4) Influential observations are identified.

(5) Proxy variables that might help explain the trends and influences in the residuals are evaluated.

(6) Influential observations suspected of biasing the coefficient estimators are removed from the sample.

(7) The model is tested for the existence of heteroscedastic error terms, and corrections are made if warranted.

(8) Statistical tests for goodness-of-fit and significance of the regressors are evaluated.

(9) The autoregressive error process is then further identified.

(10) Steps 2 through 9 are repeated as necessary.

The modeling process begins with analysis of last year's models because forecasts should not change radically from one year to the next, in the absence of critical events. This suggests that changes in the models should be minimal. The changes from RSP2014 to RSP2015 derive primarily from adding a year of data to the sample period.

A.1.1. Identification of the Autoregressive error structure.

The following steps help to identify the autoregressive error process.

(1) Experience over the years has shown that the error terms in the daily peak load models follow a process of *at least* AR(1). The first step, then, is to specify a first-order autoregressive model.

(2) Serial correlation in the errors may be evidence of problems with the model specification. Before testing for higher-order serial correlation, the residuals from the AR(1) model are examined for extreme outliers. To the extent possible, the reason for large outliers is determined. Very large residuals can often be explained by severe weather events or

[6] Refer to the document on the ISO-New England website for details, "Regional and State Energy and Peak Model Details," at http://www.iso-ne.com/trans/celt/fsct_detail/2014/rsp_14_peak_models.xls.

abnormalities on the electric system, and these anomalous observations must be eliminated from the sample to prevent biasing the estimators[7]. In other cases, dummy variables may be introduced as proxies for unobservable variables.

(3) The Durbin-Watson test evaluates the model for the presence of first-order autocorrelation. Since the errors are known to follow a process of AR(1) or greater, other evaluation techniques are used to assess the degree of serial correlation.

(4) After accounting for large outliers, trends in the residuals, and possible heteroscedasticity, the AR(1) model is re-estimated and the correlogram is examined. The correlogram shows autocorrelation and partial autocorrelation functions, and reports “Q-statistics.” The Q_statistic at lag *k* tests the null hypothesis that there is no autocorrelation up to order *k*. Identifying the appropriate autoregressive structure can lead to more accurate inferences about the parameter estimates. In some cases, autoregressive terms are left in the estimating equation even though they may not be “highly” significant statistically, if the adjustment to the error structure improves overall model performance.

(5) After reviewing the correlogram, the Breusch-Godfrey Lagrange multiplier test for general, high-order autoregressive errors is performed. The highest order AR process that might describe the serial correlation is selected.

(6) Based on this analysis, the model is re-specified with the higher-order AR process.

(7) This process is repeated each time the equation is modified.

In most cases, the final correlogram and Breusch-Godfrey tests fail to reject the hypothesis of further serial correlation in the residuals. In some cases, it was not possible to achieve this result[8].

A.1.2. Heteroscedasticity

In the presence of heteroscedastic error terms, the parameter estimators are unbiased but inefficient, which affects statistical tests of significance. Heteroscedasticity primarily affects cross-section data, but can be present in time-series data as well. The peak load models are strictly time series, so the likelihood of truly heteroscedastic error terms is reduced. More important, heteroscedasticity in these models may arise from the presence of outliers, or because the model is not correctly specified for other reasons. It can also result from skewness in the distribution of regressors in the model, which may be present in some periods in the weather data.

To address the types of specification issues that these tests might be reflecting, the residuals are examined for influential observations and for identifiable patterns in the variation of the residuals. To the extent possible, proxy variables such as year dummies are introduced to represent omitted variables

[7] Ordinary Least Squares regression equations fit an “average” model. An influential observation is one that carries too much weight, pulling the average in its direction. This biases the regression line in the direction of the influential observation.

[8] The presence of autocorrelation does not affect the estimated coefficients. Rather, it biases the statistical tests, so that inferences about the significance of the estimators may be inaccurate. However, in all models the important drivers exhibit very high t-statistics, indicating that the coefficients are significant with a high level of confidence.

and correct for within-year anomalies. If the model shows improvement, the proxy variables may be left in the model even if they do not seem particularly significant.[9]

(1) Tests for heteroscedasticity. The Arch test with lag(2) and the White test are used to assess the presence of heteroscedasticity. Since the White homoscedasticity test assumes that the errors are both homoscedastic and independent of the explanatory variables, this statistic is generally used to test for model misspecification as well. Failure of any one of these conditions could lead to a significant test statistic.

(2) Correction for heteroscedasticity. When tests suggest that heteroscedasticity is present after specifying the model as completely as possible, it is re-estimated with heteroscedasticity and autocorrelation of unknown form [(Newey-West) HAC] errors and covariance. When possible, residuals are pre-whitened, with one lag greater than the largest autoregressive term in the model. The Newey-West automatic bandwidth with truncation is used for pre-whitening. This procedure enables more accurate statistical inferences.

A.1.3. Influential Observations

(1) The residuals are examined for obvious patterns. In some instances, specific holidays or surrounding days need to be accounted for separately, for example, Christmas and Christmas Eve. Saturdays and Sundays may need to be specified individually, rather than combined into weekends. The effect of weekends can vary by year, in which case interactive year/weekend dummy variables are called for. The residuals may show patterns within a particular year, suggesting that a dummy variable for that year might improve the model's properties.

(2) Influential observations are identified. Specific observations can have a large influence on model estimation. Since standard regression techniques produce "average" results, influential observations can cause biased estimators[10]. There are several methods for identifying influential observations.

   a. Leverage Plots. Leverage plots are the multivariate equivalent of a univariate plot of an explanatory variable against the residuals. The leverage plots of residuals against weather are analyzed to detect observations with greater than average "leverage."

[9] The common "rule of thumb" that a t-statistic>2 indicates statistical significance is an approximation based on the t-distribution, which depends on the sample size for interpretation. A one-tailed t-test is appropriate for variables whose sign is known *a priori*. In this case, a t-statistic of 2 implies that the null hypothesis, $\beta=0$, can be rejected with about 97.5% confidence. A lower value for the t-test does *not* mean the coefficient is not statistically significant, but that the null hypothesis can be rejected only at lower confidence levels. If the t-statistic is below our accepted tolerance level, we fail to reject the null hypothesis. Because the value of a t-statistic may be less than two does not mean that the variable has no explanatory power.

[10] Ordinary least squares regression minimizes the sum of squared errors (observed dependent variable minus predicted dependent variable.) An influential observation can distort the regression line (or plane), biasing the results towards the influential observation. A discussion of this issue can be found in most introductory econometrics tests.

b. Influence Statistics. Four panels of influence statistics are evaluated: RStudent (similar to the standardized residuals), DFFits (similar to Cook's Distance), the Covariance Ratio, and the "Hat" matrix. Observations that are significant by more than one measure are then focused on. It is often not possible to determine why an observation is influential and correct for it, and observations considered sufficiently extreme are eliminated from the sample[11,12].

A.1.4. Goodness-of-Fit and Statistical Significance.

It is important for the model to fit the historical data as closely as possible. Forecasts from an econometric model assume that historical relationships will continue into the future. A model that does not fit the data well introduces additional uncertainty into the forecast.

The traditional measure of how well the model fits the data is the R^2 statistic. The better the model fits the data, the higher will be the R^2 score. For the summer peak models, this statistic ranges from 0.9097 to 0.9549, with a median of 0.9429. For the winter peak models, R^2 ranges between 0.8537 and 0.9202, with a median value of 0.9001.

The F-statistic is a joint test that all of the slope coefficients are zero. All of the models reject this hypothesis.

A.1.5. Stationarity.

A time series is said to be (weakly or covariance) stationary if the mean and autocovariances of the series do not depend on time. If a dependent variable is trending, and there must be at least one non-stationary explanatory variable to explain this trend in a correctly specified equation[13]. In the case of a nonstationary dependent variable with a nonstationary independent variable, the residual series must be stationary;[14] otherwise, it has one or more unit roots. This means the variance is not constant over time, and the equation is considered to be "spurious." The residuals were tested for stationarity (tests for unit roots) using the Augmented Dickey-Fuller (ADF) test, where possible[15]. In each of the models, the ADF test rejected the null hypothesis of a unit root in the residuals at the 5% critical level or better.

A.1.6. Consistency.

It is reasonable to expect the models to be consistent from year to year. It also would be unreasonable for this year's outlook to differ substantially from last year's, unless underlying economic drivers change significantly. One additional year of weather data should not change the relationship between peaks

---

[11] See the EVIEWs documentation for more information: Quantitative Micro Software, LLC, *EViews 7 User's Guide II*. The criteria for identifying influential observations are: DFFITS: cutoff ≈ 2 (similar to Cook's D); RStudent: cutoff ≈ 2 (about 2 s.d.); |CovRatio-1|≥ 3p/N (p=parameter; N=number of observations).

[12] Alternative methods for accounting for the effects of influential observations include robust regression and quantile regression. In EVIEWs, neither of these techniques allows corrections for AR error processes, and the resulting models did not fit the data well.

[13] When an equation contains a non-stationary dependent variable and non-stationary explanatory variables that are stationary at the same order of differencing, the equation is said to be co-integrated.

[14] See *Review of the Econometric Models Used by ISO-NE in April 2007 to Determine the Long-Run Forecasts of Energy and Peak Load*, prepared by Tim Mount from Cornell University.

[15] Where it was not possible to use the ADF, alternative tests were evaluated.

and weather, and variations in modeling methodology should not change the overall relationships among variables substantially.

A.2. Energy Model Evaluation

The statistical model evaluation process is somewhat similar to the Peak Load model, but with several exceptions and limitations.

(1) Serial Correlation. The Durbin-Watson statistic tests for first-order autocorrelation in the error terms, and it cannot be used in the presence of lagged dependent variables. The Breusch-Godfrey Serial Correlation LM test checks for higher order serial correlation in the error terms and was evaluated at lag=2. For all of the models, the results of this test, combined with examination of the correlogram, suggest that the errors are not serially correlated.

(2) Heteroscedasticity. Given the small sample size, visual inspection of the residuals suggests that the error terms are homoscedastic in each of the models. Nevertheless, formal tests for heteroscedasticity were conducted, with no evidence found.

(3) Influential outliers. Influence statistics could not be calculated because the sample size is too small. Instead, large outliers were visually detected by examining the standardized residuals. Dummy variables were included to account for the outliers.

(4) Unit Root Test. The Augmented Dickey-Fuller (ADF) Test shows that the residuals from all seven models are stationary, with a significance level of 5% or better.

(5) Standard Statistical Tests. The R^2 goodness-of-fit measure ranges from 0.9781 to 0.9963. These high values are not unexpected, given the aggregate time-series nature of the data. The t-statistics on the coefficients of most regressors are over 2. Exceptions include some constant terms and an occasional variable considered *a priori* to be important for model fit.