Technical Paper

Short- and Long-Term Monthly Sales Forecasting Using Statistically Adjusted End-Use Models

The traditional approach to forecasting monthly sales for a customer class is to develop an econometric model that relates monthly sales to weather, seasonal variables, and economic conditions. From a forecasting perspective, the strength of econometric models is that they are well suited to identifying historical trends and to projecting these trends into the future. In contrast, the strength of the end-use modeling approach is the ability to identify the end-use factors that are driving energy use. By incorporating end-use structure into an econometric model, the statistically adjusted end-use modeling framework exploits the strengths of both approaches.

There are several advantages to this approach.

- The equipment efficiency trends and saturation changes embodied in the long-run end-use forecasts are introduced explicitly into the short-term monthly sales forecast. This provides a strong bridge between the two forecasts.
- By explicitly introducing trends in equipment saturations and equipment efficiency levels, it is easier to explain changes in usage levels and changes in weather-sensitivity over time.
- Data for short-term models are often not sufficiently robust to support estimation of a full set of price, economic and demographic effects. By bundling these factors with equipment-oriented drivers, a rich set of elasticities can be built into the final model.

This paper describes this approach using an example for residential class sales.

Statistically Adjusted End-Use Modeling Framework

The statistically adjusted end-use modeling framework begins by defining energy use (*USE*) for an average customer in year (y) and month (m) as the sum of energy used by heating equipment ($Heat_{y,m}$), cooling equipment ($Cool_{y,m}$) and other equipment ($Other_{y,m}$). Formally,

$$USE_{y,m} = Heat_{y,m} + Cool_{y,m} + Other_{y,m}$$

Although monthly sales (*USE) ARE* measured for individual customers, the end-use components (*Heat*, *Cool* and *Other*) are not. This implies that the above relationship can only be estimated and not measured. Substituting estimates for the end-use elements gives the following econometric equation.

$$\mathrm{USE}_{y,m} = b_1 \times \mathrm{XHeat}_{y,m} + b_2 \times \mathrm{XCool}_{y,m} + b_3 \times \mathrm{XOther}_{y,m} + \varepsilon_{y,m}$$

Here, *XHeat*, *XCool*, and *XOther* are estimates of the true heating, cooling and other equipment usage values. These estimates are constructed from end-use information, weather data, and market data. As will be shown below, the equations used to construct these estimates are simplified end-use models. The *XHeat, XCool,* and *XOther* are the estimated usage levels for each of the major end uses based on these models. The above econometric equation can then be thought of as a statistically adjusted end-use model, where the estimated slopes (b_1, b_2 and b_3) are the adjustment factors and ε is a random error.

Constructing XHeat

As represented in end-use models, energy use by space heating systems depends on the following types of variables.

- Heating degree days (or more generally, cold weather),
- Heating equipment saturation levels,
- Heating equipment operating efficiencies,
- Thermal integrity of homes, and
- Average household size, household income, and energy price.

Based on variables available from the REEPS database, the heating variable is represented as the product of an annual equipment index and a monthly usage multiplier. That is,

$$\mathrm{XHeat}_{y,m} = \mathrm{HeatIndex}_{y} \times \mathrm{HeatUse}_{y,m}$$

where $XHeat_{y,m}$ is estimated heating energy use in year (y) and month (m), $HeatIndex_y$ is the annual index of heating equipment, and $HeatUse_{y,m}$ is the monthly usage multiplier.

Heating Equipment Index. The heating equipment index is defined as a weighted average across equipment types (Type) of equipment saturation levels (*HeatShare*) normalized by operating efficiency levels (*Eff*). Given a set of fixed weights (*Weight*), the index will change over time with changes in equipment saturations and operating efficiencies. Formally, the equipment index is defined as:

$$\text{HeatIndex}_y = \sum_{\text{Type}} \text{Weight}^{\text{Type}} \times \frac{\left(\text{HeatShare}_y^{\text{Type}} \big/ \text{Eff}_y^{\text{Type}} \right)}{\left(\text{HeatShare}_{\text{base}}^{\text{Type}} \big/ \text{Eff}_{\text{base}}^{\text{Type}} \right)}$$

In this expression, *base* corresponds to a base year for normalizing the index. The ratio on the right is equal to 1.0 in the base year. In other years, it will be greater than one if equipment saturation levels are above their base year level. This will be counteracted by higher efficiency levels, which will drive the index downward.

The weights are defined by the estimated heating energy use per household for each equipment type in the base year.

$$\text{Weight}^{\text{Type}} = \frac{\text{HeatingEnergyUse}_{\text{base}}^{\text{Type}}}{\text{HouseHolds}_{\text{base}}}$$

With these weights, the *HeatIndex* value in the base year will be equal to the estimated annual heating energy use per household in that year. Variations from this value in other years will be proportional to saturation and efficiency variations around their base year values.

For electric heating equipment, the REEPS database contains four equipment types: electric resistance furnaces, baseboard and other resistance room units, heat pumps, and furnace fans. Examples of weights for these four equipment types are as follows:

Equipment Type	Weight (kWh)
Electric Resistance Furnace	421
Electric Resistance – Room Unit	71
Electric Heat Pump	516
Furnace Fans	532

Monthly Heating Usage Multiplier. Heating system usage levels are impacted on a monthly basis by several factors, including weather, household size, income levels, and prices. Using the REEPS default elasticity parameters, the estimates for space heating equipment usage levels are computed as follows:

$$\text{HeatUse}_{y,m} = \left(\frac{\text{HDD}_{y,m}}{\text{NormHDD}} \right) \times \left(\frac{\text{HHSize}_y}{\text{HHSize}_{\text{base}}} \right)^{.25} \times \left(\frac{\text{Income}_y}{\text{Income}_{\text{base}}} \right)^{.20} \times \left(\frac{\text{Price}_{y,m}}{\text{Price}_{\text{base}}} \right)^{-.30}$$

where *HDD* is the number of heating degree days in year (y) and month (m),
NormHDD is the normal value for annual heating degree days,
HHSize is average household size in a year (y),
Income is average real income per household in a year (y),
Price is the average real price of electricity in a year (y) and a month (m), and
base indexes the base year.

Figure 1: Heating Equipment Shares

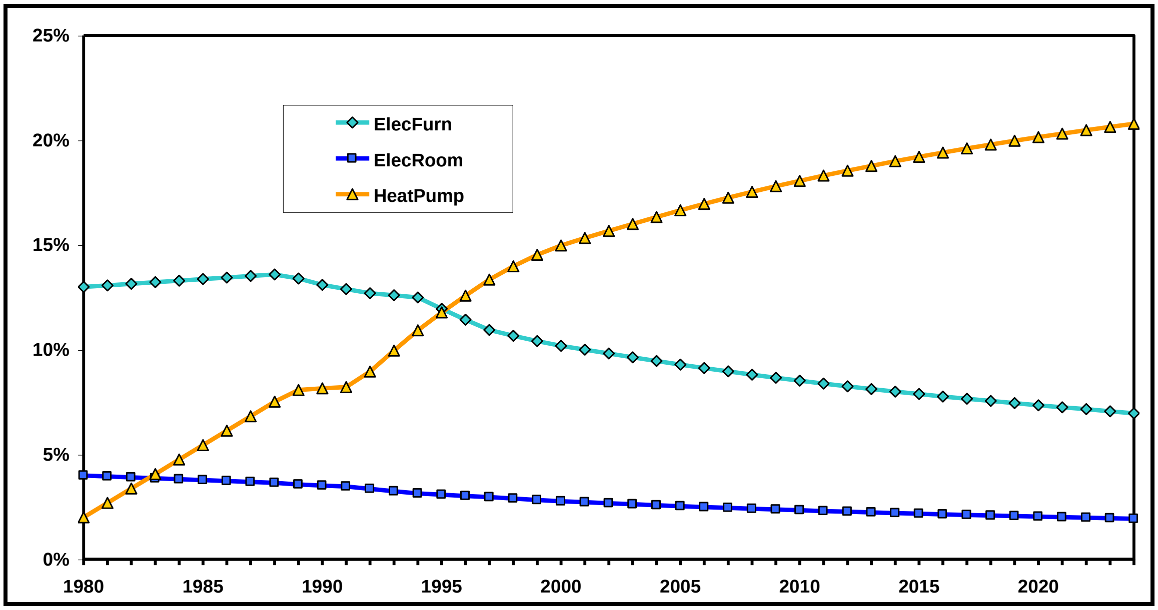


Figure 2: Heating Equipment Efficiencies

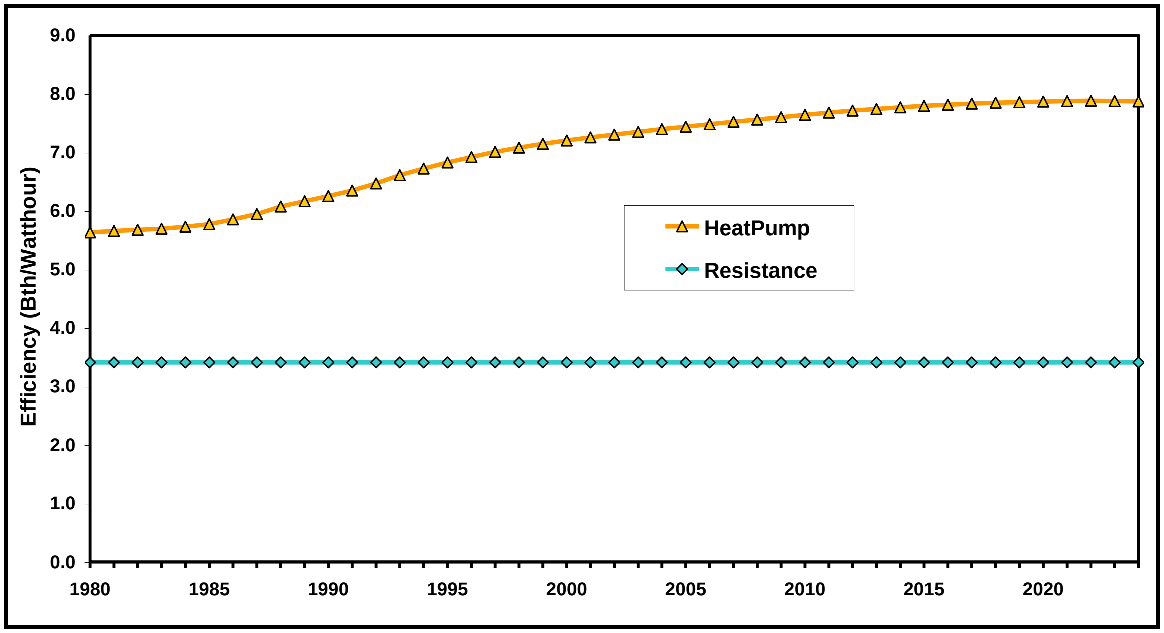
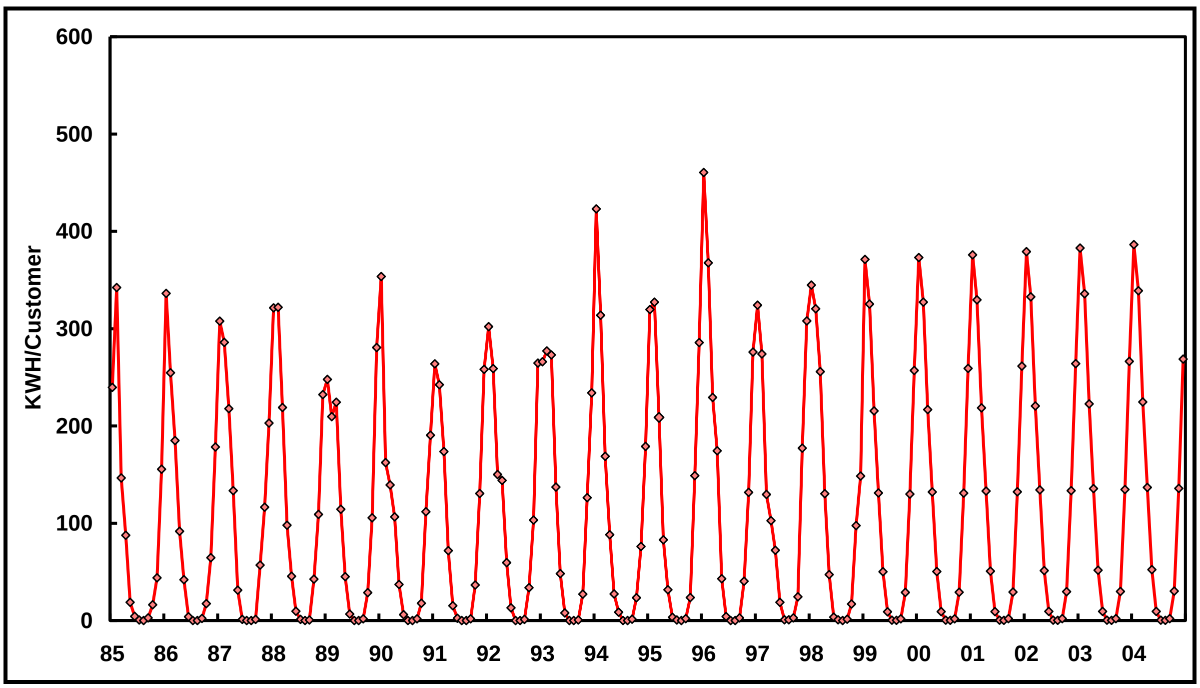


Figure 3: Heating Variable (XHeat)



By construction, the *HeatUse* variable has an annual sum that is close to one in the base year. The first term, which involves heating degree days, serves to allocate annual values to months of the year. The remaining terms average to one in the base year. In other years, the values will reflect changes in the economic driver changes, as transformed through the end-use elasticity parameters. For example, if the real price of electricity goes up 10% relative to the base year value, the price term will contribute a multiplier of about .97 (computed as 1.10 to the -.30 power).

Data for heating equipment shares, average equipment efficiency levels, and the final explanatory variable for heating (*XHeat*) are shown in Figure 1 through Figure 3.

Constructing XCool

The explanatory variable for cooling loads is constructed in a similar manner. The amount of energy used by cooling systems depends on the following types of variables.

- Cooling degree days (or more generally, warm weather) and humidity levels,
- Cooling equipment saturations,
- Cooling equipment operating efficiencies,
- Average household size, household income, and energy price.

Based on variables available from the REEPS database, the cooling variable is represented as the product of an equipment-based index and monthly usage multiplier. That is,

$$\text{XCool}_{y,m} = \text{CoolIndex}_{y} \times \text{CoolUse}_{y,m}$$

where $XCool_{y,m}$ is estimated cooling energy use in year (y) and month (m), $CoolIndex_y$ is an index of cooling equipment, and $CoolUse_{y,m}$ is the monthly usage multiplier.

Cooling Equipment Index. As with heating, the cooling equipment index is defined as a weighted average across equipment types of equipment saturation levels normalized by operating efficiency levels. Formally, the cooling equipment index is defined as:

$$\text{CoolIndex}_{y} = \sum_{\text{Type}} \text{Weight}^{\text{Type}} \times \frac{\left(\text{CoolShare}_{y}^{\text{Type}} \Big/ \text{Eff}_{y}^{\text{Type}} \right)}{\left(\text{CoolShare}_{\text{base}}^{\text{Type}} \Big/ \text{Eff}_{\text{base}}^{\text{Type}} \right)}$$

Data values for a selected base year (base) are used for normalizing the index. By construction, the ratio on the right is equal to 1.0 in the base year. In other years, it will be greater than one if equipment saturation levels are above their base-year level. This will be counteracted by higher efficiency levels, which will drive the index downward. The weights are defined as the estimated base-year cooling energy use per household for each equipment type.

$$Weight^{Type} = \frac{CoolingEnergyUse_{base}^{Type}}{HouseHolds_{base}}$$

With these weights the *CoolIndex* value in the base year will be equal to estimated annual cooling energy use per household in that year. Variations from this value in other years will be proportional to saturation and efficiency variations around their base values.

For cooling equipment, the REEPS database contains three equipment types: Central air conditioning, Heat pump, and Room air conditioning. Examples of weights for these three equipment types are as follows:

Equipment Type	Weight (kWh)
Central Air Conditioning	2513
Heat Pump	522
Room Air Conditioning	405

Monthly Cooling Usage Multiplier. Cooling system usage levels are impacted on a monthly basis by several factors, including weather, household size, income levels, and prices. Using the REEPS default parameters, the estimates of cooling equipment usage levels are computed as follows:

$$CoolUse_{y,m} = \left(\frac{CDD_{y,m}}{NormCDD}\right) \times \left(\frac{HHSize_{y}}{HHSize_{base}}\right)^{.25} \times \left(\frac{Income_{y}}{Income_{base}}\right)^{.20} \times \left(\frac{Price_{y,m}}{Price_{base}}\right)^{-.30}$$

where, *CDD* is the number of cooling degree days in year (y) and month (m),
NormCDD is the normal value for annual cooling degree days,
HHSize is average household size in a year (y),
Income is average real income per household in a year (y), and
Price is the average real price of electricity in year (y) and month (m)

By construction, the *CoolUse* variable has an annual sum that is close to one in the base year. The first term, which involves cooling degree days, serves to allocate annual values to months of the year. The remaining terms average to one in the base year. In other years, the values will change to reflect changes in the economic driver changes.

Data for cooling equipment shares, average equipment efficiency levels, and the final explanatory variable for cooling (*XCool*) are shown in Figure 4 to Figure 6.

Figure 4: Cooling Equipment Shares

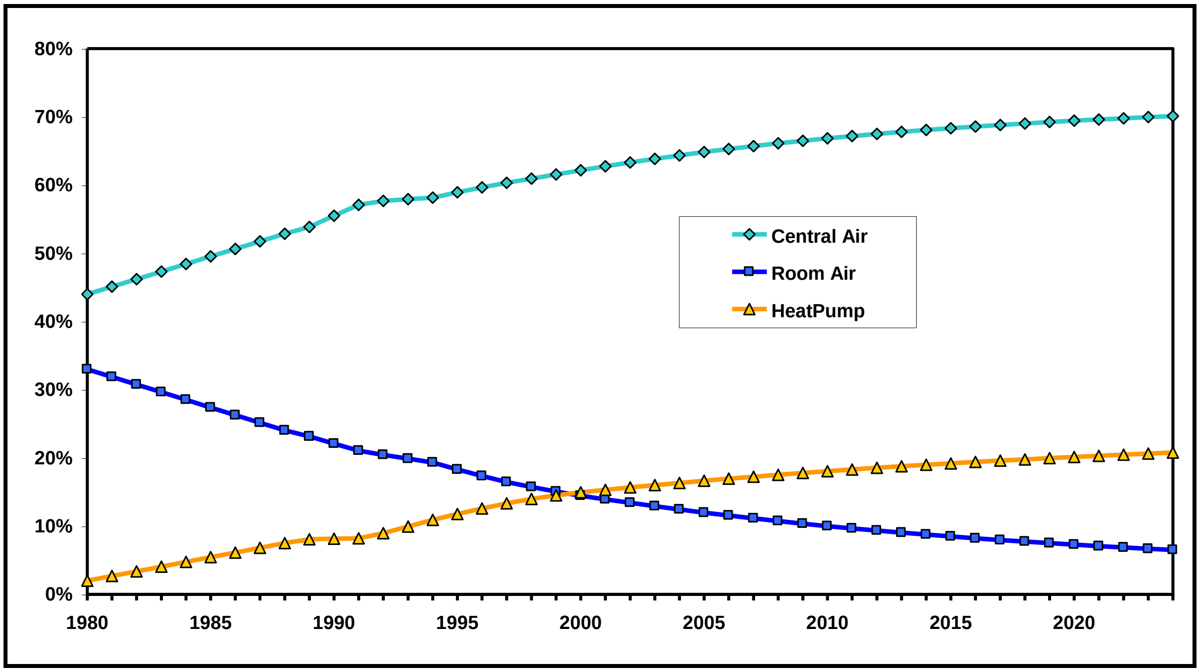


Figure 5: Cooling Equipment Efficiencies

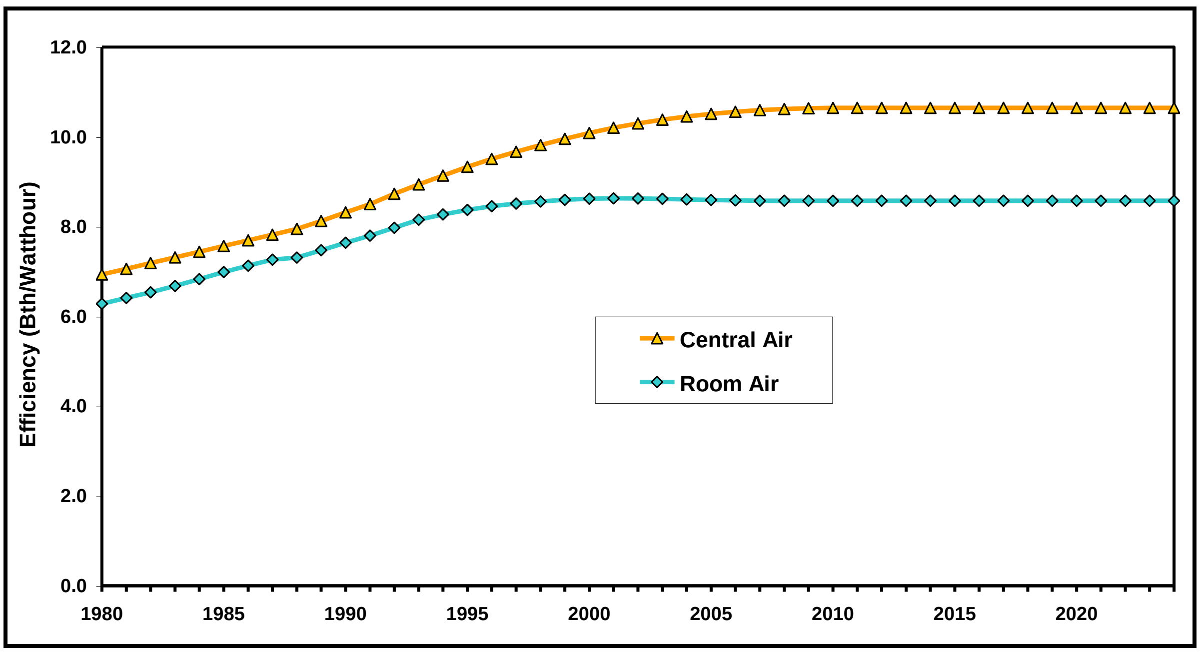
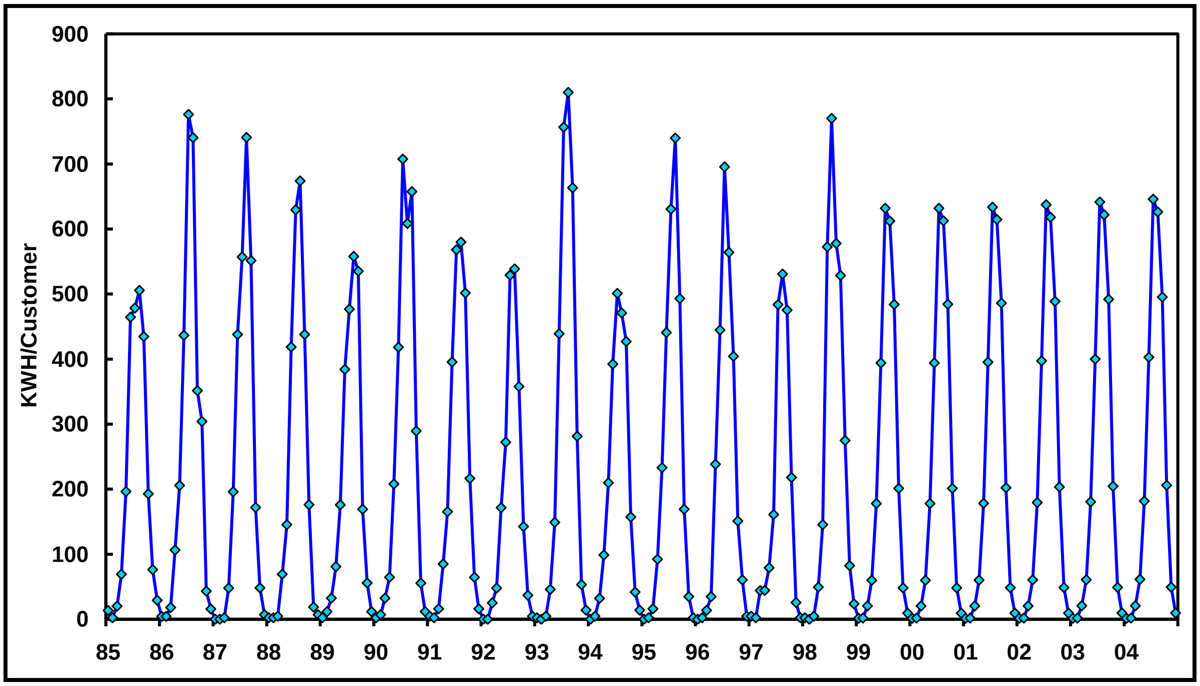


Figure 6: Cooling Variable (XCool)



Constructing XOther

Monthly estimates of non-weather sensitive sales can be derived in a similar fashion to space heating and cooling. Based on end-use concepts, other sales are driven by:

- Appliance and equipment saturation levels and densities,
- Appliance efficiency levels,
- Average number of days in the billing cycle for each month,
- Hours of light to reflect higher lighting loads in the winter,
- Water temperatures implying higher water heating loads in the winter,
- Indoor temperatures implying higher refrigerator loads in the summer,
- Average household size and real income, and
- Real prices.

The variable for other sales can be developed using the equipment stock data contained in the REEPS database, augmented by other factors relating to monthly variation. The explanatory variable for other uses is defined as follows:

$$XOther_{y,m} = OtherIndex_{y,m} \times OtherUse_{y,m}$$

Other Equipment Index. The *OtherIndex* term in this expression embodies information about appliance saturation levels and efficiency levels. To reflect the special influence of seasonal factors on lighting, water heating, and refrigeration, the equipment index for other uses is defined as follows:

$$\text{OtherIndex}_{y,m} = \left[\sum_{\text{Use}} \text{Weight}^{\text{Use}} \times \left(\frac{\frac{\text{Sat}_y^{\text{Use}}}{\text{Eff}_y^{\text{Use}}}}{\frac{\text{Sat}_{\text{base}}^{\text{Use}}}{\text{Eff}_{\text{base}}^{\text{Use}}}} \right) \times \text{Mult}_m^{\text{use}} \right] + \left[\text{UEC}_y^{\text{Light}} \times \text{Mult}_m^{\text{Light}} \right] + \left[\text{UEC}_y^{\text{Misc}} \times \text{Mult}_m^{\text{Misc}} \right]$$

where, Sat^{use} represents the fraction of households who have an appliance type,
$Mult^{use}$ is a monthly multiplier for the Use in month (m),
Weight is the weight for each use, and
UEC is the unit energy consumption for lighting and miscellaneous uses in year (y).

This index combines information about trends in saturation levels and efficiency levels for the main appliance categories with monthly multipliers for lighting, water heating, and refrigeration, and with general assumptions about trends in unit energy consumption values (UEC) for lighting and miscellaneous uses. As with heating and cooling, the weights are defined as the base year values of energy use per household for each end use.

Monthly Other Usage Multiplier. Further monthly variation is introduced by multiplying by usage factors that cut across all end uses, constructed as follows:

$$\text{OtherUse}_{y,m} = \left(\frac{\text{BillingDays}_{y,m}}{365} \right) \times \left(\frac{\text{HHSize}_y}{\text{HHSize}_{\text{base}}} \right)^{.46} \times \left(\frac{\text{Income}_y}{\text{Income}_{\text{base}}} \right)^{.10} \times \left(\frac{\text{Price}_{y,m}}{\text{Price}_{\text{base}}} \right)^{-.15}$$

In this expression, the end-use elasticities on income, household size, and real price are taken from the REEPS default database. The annual data for the equipment stock and UEC drivers for the main uses are presented below in Figure 7 and Figure 8. In Figure 8, the appliance category includes data for cooking, dishwashers, clothes washers, clothes dryers, and televisions. Figure 9 shows the final monthly variable for this category on a per customer basis. The main source of variation comes from the billing-days variable.

Figure 7: Other Uses – Refrigerators, Lighting, DHW

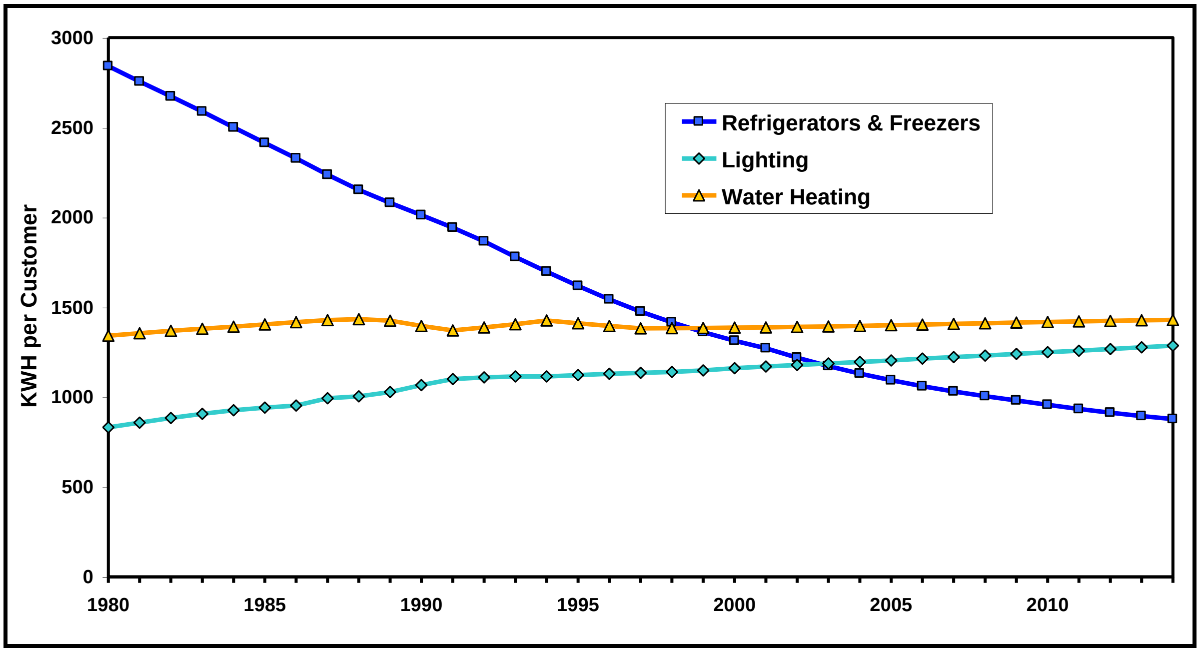


Figure 8: Other Uses – Appliances & Miscellaneous

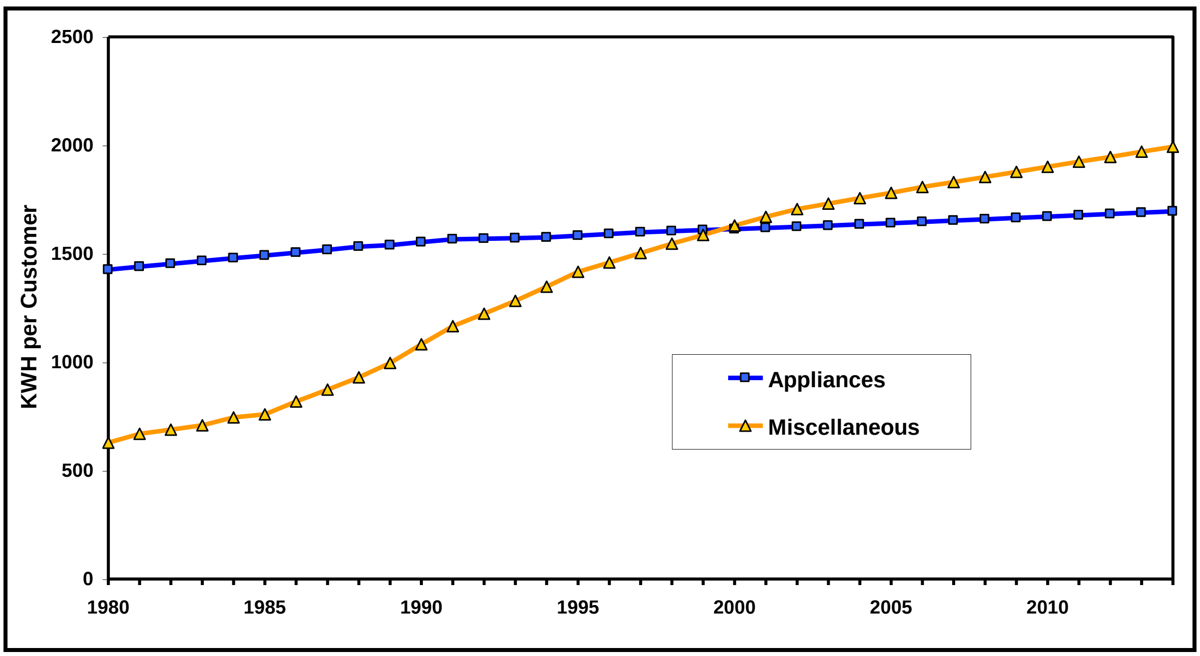
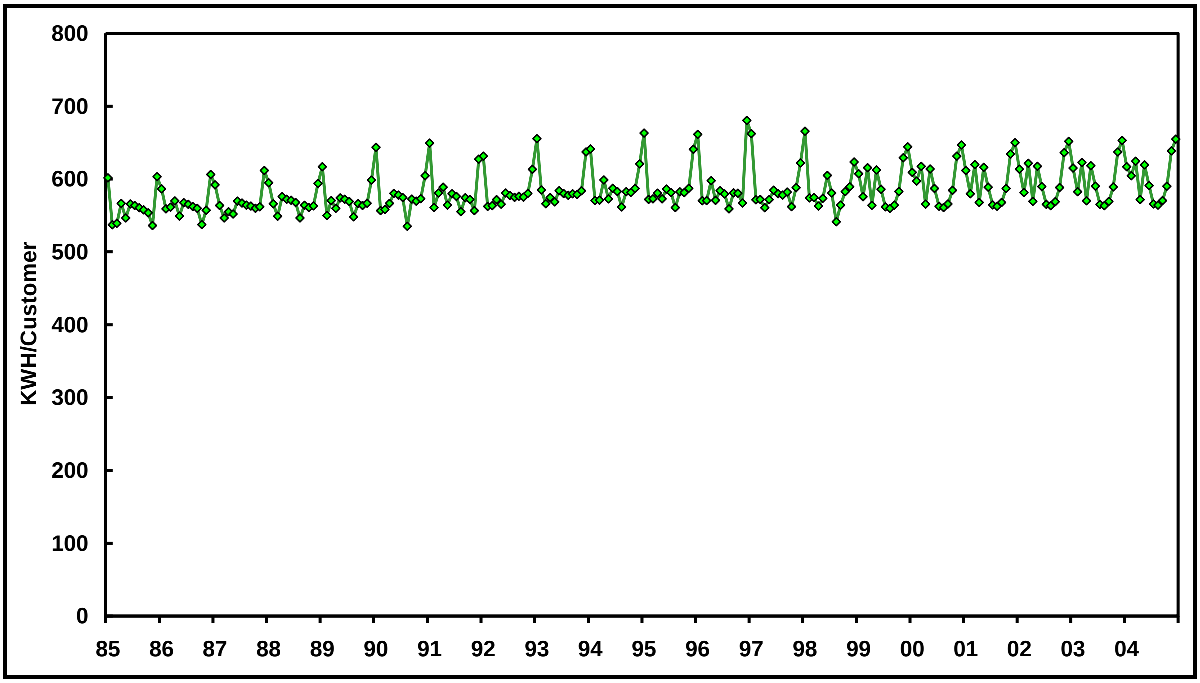


Figure 9: Other Uses Variable (XOther)



Example of Approach

A sequence of three models was estimated starting with the simplest specification and then moving to the final model specification. The results for the base model are presented in Table 1. As expected, the parameter values on the cooling, heating, and other variables are positive, strongly significant and close to one. The parameters on the two weather sensitive end uses are significantly greater than one, and the parameter on the Other end use is significantly less than one. The DW statistic indicates significant first order serial correlation.

The overall fit of the model is illustrated in Figure 10, which shows a scatter plot between the actual and predicted values. The base model performs well for those months were sales are under 1,500 GWh, but the residual dispersion is visibly larger for the high-sales months. The mean absolute deviation in the sample period is about 59 GWh, which corresponds to an in-sample MAPE of 4.5%. These values will be used as a point of reference for the two other model specifications.

In the second model, the specification is extended to include a set of monthly binary variables. The results from this model are presented in Table 2. The t-Statistics on the monthly variables are significant for many months, indicating seasonal variation that is not directly correlated with weather variables. Inclusion of these monthly factors reduces the slopes on the heating variable significantly, and the slopes on the Cooling and Other variables converge closer to one.

Inclusion of the monthly variables reduces the in-sample MAD to 48 GWh, and the MAPE improves to 3.6%. The overall model fit is illustrated in Figure 11. Although this specification fits better than the base case model, there are still large errors in the months when sales levels are highest.

Estimates for the final model specification are presented in Table 3. In this specification the following four variables were added:

1 *XHeatTrend*, is computed by interacting *XHeat* with a linear time trend. The coefficient on this variable indicates that there is a slight, but insignificant, positive trend in electric space heating sales that is not captured by the end-use trends already embedded in the *XHeat* variable.

2. *XCoolHum*, is computed by interacting with *XCool* with a monthly humidity variable. The coefficient on this variable indicates that sales are higher when humidity levels are up.

3. *XCoolTrend*, is computed by interacting *XCool* with a linear time trend. The coefficient on this variable indicates that there is a strong and significant positive trend in cooling sales that is not captured by the end-use trends embedded in the *XCool* variable.

4. *XOtherTrend*, is computed by interacting *XOther* with a linear time trend. The coefficient on this variable indicates that there is no additional underlying growth in miscellaneous equipment usage beyond what is captured in the *XOther* variable.

Table 1: Base Model Results

Variable	Coefficient	StdErr	T-Stat
CustXHeat	1.333	0.063	21.028
CustXCool	1.227	0.030	40.385
CustXOther	0.859	0.024	35.924
Summary Statistics			
R-Squared			0.937
Adjusted R-Squared			0.936
Durbin-Watson Statistic			0.52
AIC			8.897
BIC			8.953
Std. Error of Regression			84.75
Mean Abs. Dev. (MAD)			58.56
Mean Abs. % Err. (MAPE)			4.52%

Figure 10: Actual Vs. Predicted Scatter Plot for Base Model

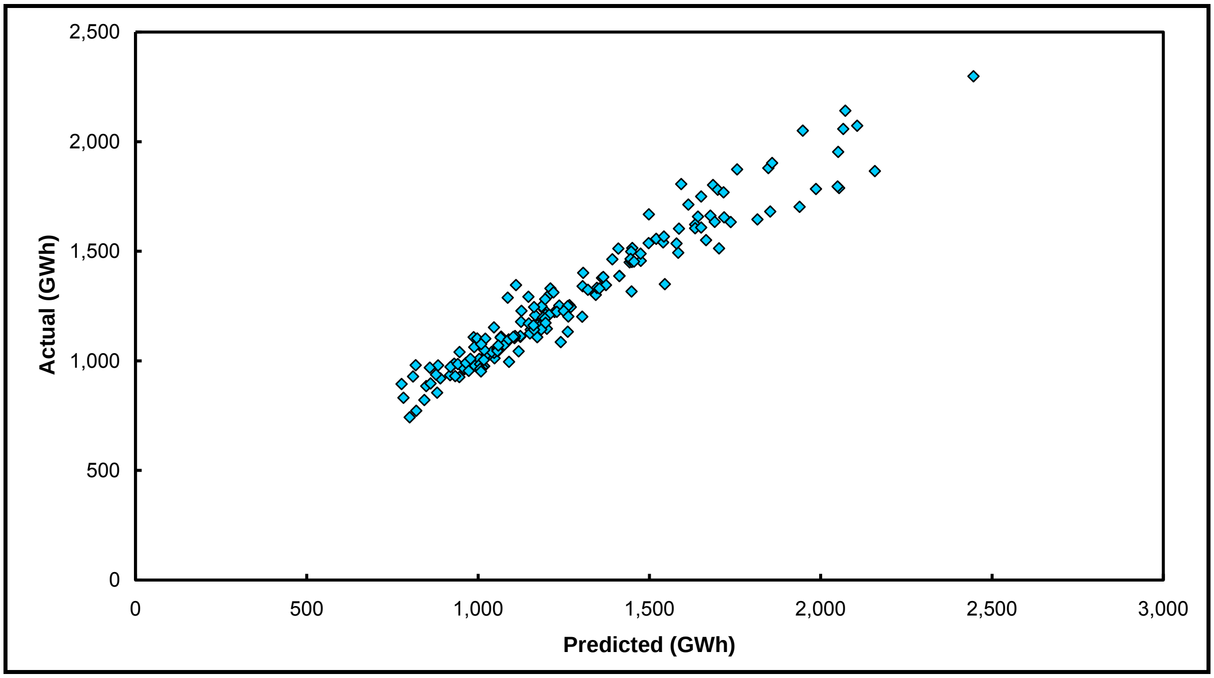


Table 2: Model Two - Base Plus Monthly Binaries

Variable	Coefficient	StdErr	T-Stat
CustXHeat	0.751	0.111	6.735
CustXCool	1.074	0.059	18.289
CustXOther	1.133	0.047	24.232
Jan	38.242	29.028	1.317
Feb	18.752	26.758	0.701
Mar	-56.297	24.573	-2.291
Apr	-156.560	29.280	-5.347
May	-233.805	36.791	-6.355
June	-197.214	47.414	-4.159
July	-116.515	58.315	-1.998
Aug	-41.367	58.008	-0.713
Sep	-51.869	51.084	-1.015
Oct	-112.029	37.766	-2.966
Nov	-104.520	28.762	-3.634
Summary Statistics			
R-Squared			0.961
Adjusted R-Squared			0.958
Durbin-Watson Statistic			0.559
AIC			8.543
BIC			8.803
Std. Error of Regression			68.83
Mean Abs. Dev. (MAD)			48.12
Mean Abs. % Err. (MAPE)			3.58%

Figure 11: Actual Vs. Predicted Plot for Model Two

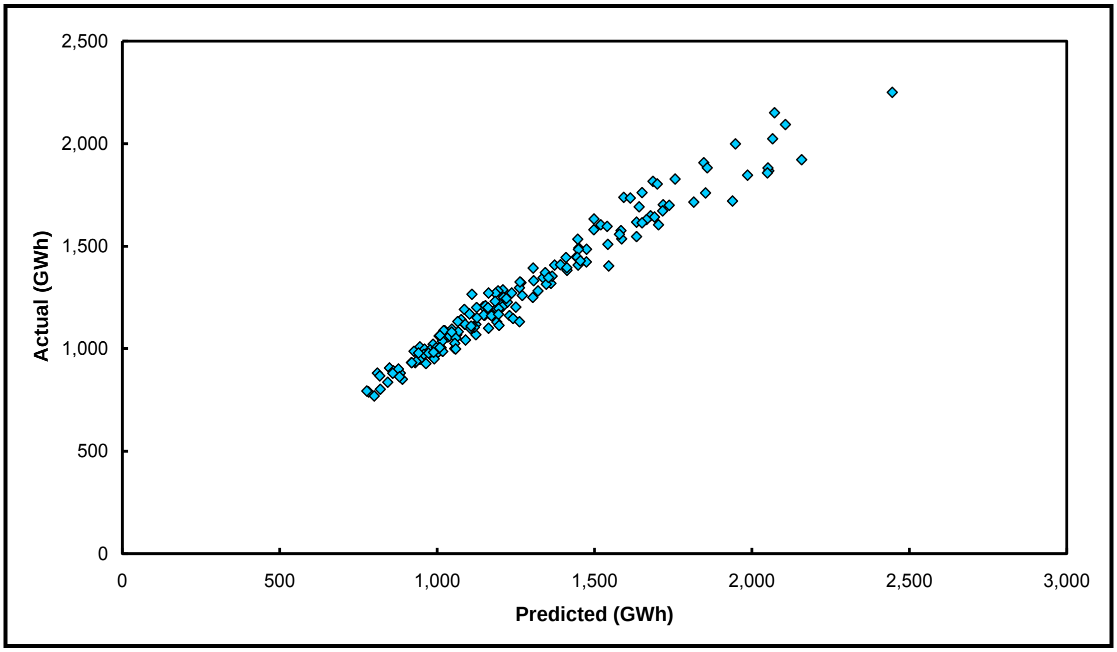


Table 3: Final Model Specification

Variable	Coefficient	StdErr	T-Stat
CustXHeat	1.136	0.104	10.878
XheatTrend	0.001	0.006	0.194
CustXCool	0.644	0.056	11.495
XcoolHum	0.005	0.001	7.325
XcoolTrend	0.028	0.003	10.005
CustXOther	0.946	0.038	24.642
XotherTrend	0.000	0.002	0.136
Jan	25.127	13.515	1.859
Feb	13.371	12.353	1.082
Mar	-12.179	11.066	-1.101
Apr	-55.885	13.644	-4.096
May	-126.129	19.934	-6.327
June	-106.424	33.137	-3.212
July	-19.543	41.983	-0.465
Aug	44.853	43.193	1.038
Sep	15.502	38.168	0.406
Oct	-82.698	27.303	-3.029
Nov	-71.600	15.570	-4.599
Summary Statistics			
R-Squared			0.993
Adjusted R-Squared			0.992
Durbin-Watson Statistic			1.042
AIC			6.91
BIC			7.245
Std. Error of Regression			30.1
Mean Abs. Dev. (MAD)			22.41
Mean Abs. % Err. (MAPE)			1.74%

The coefficients on *XHeat* (1.14) and *XOther* (.95) are not significantly different from one in this final specification. However, the coefficient on *XCool* (.64) is significantly less than one, indicating lower sensitivity to hot weather at the beginning of the sample period. To understand how this sensitivity changes, the time trend value is zero in 1980 and increases by 1 each year in equal monthly increments. As a result, through the interaction term, each decade adds .28 to the slope on *XCool*. This implies that the combined slope on this variable is 1.20 by the year 2000. Of course, this slope growth will continue in the forecast period, and this growth will more than counteract the forecasted improvement in cooling equipment efficiency levels, resulting in a continuing increase in weather sensitivity on the cooling side.

The overall model fit improves significantly over the base model with an in-sample MAD of 22.4 and a MAPE of 1.7%. The overall fit is illustrated in Figure 12 and Figure 13. As illustrated, this model does a good job in predicting sales for all months. In particular, the model shows a significant improvement in tracking the months with sales above 1,500 GWh.

Figure 12: Final Model Specification Actual Vs. Predicted

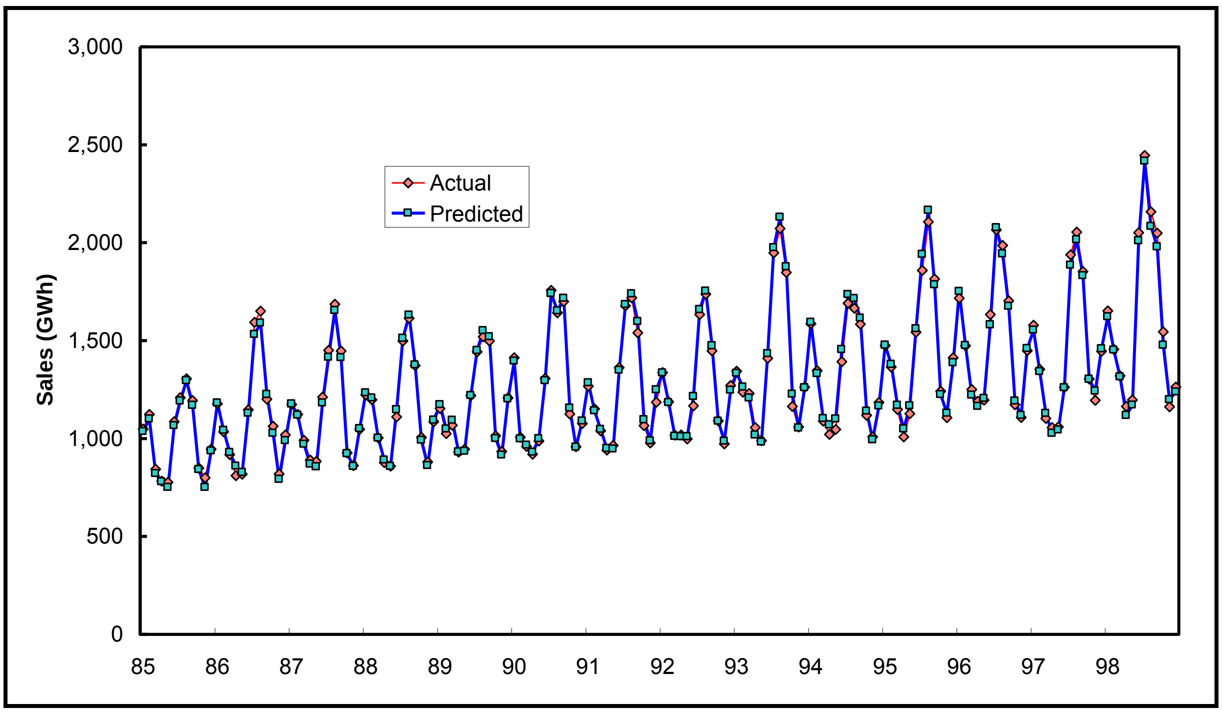


Figure 13: Actual Vs. Predicted Plot for the Final Model Specification

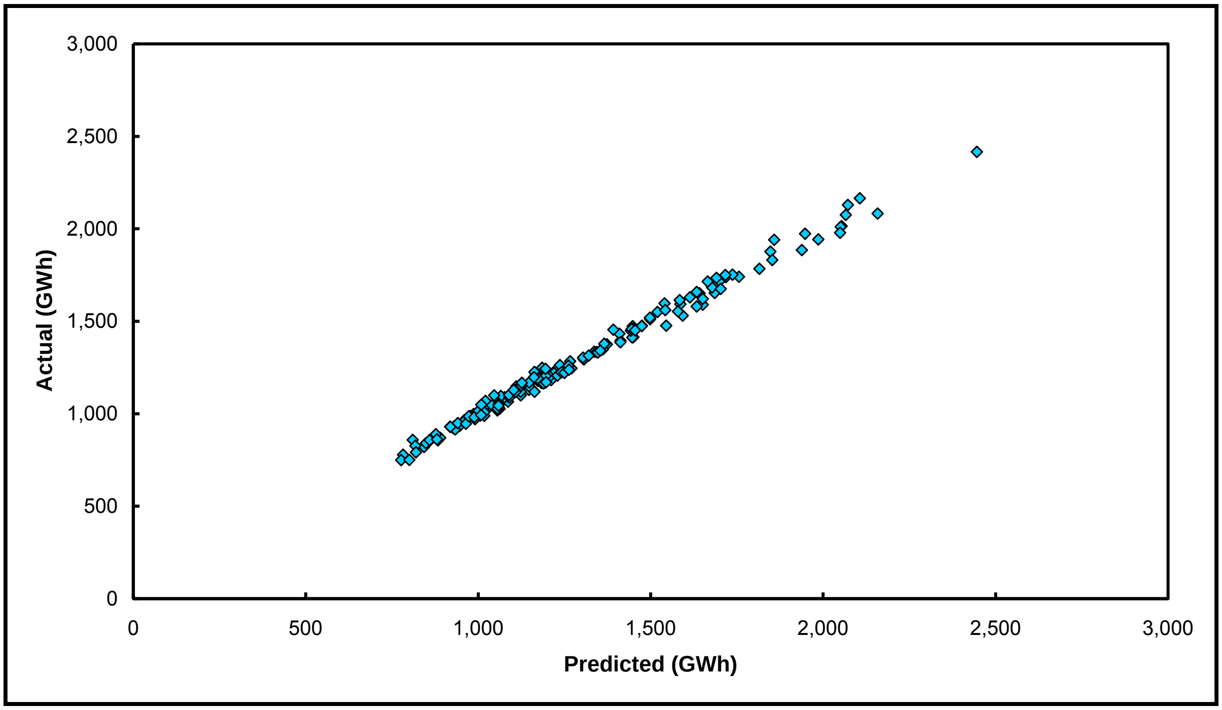
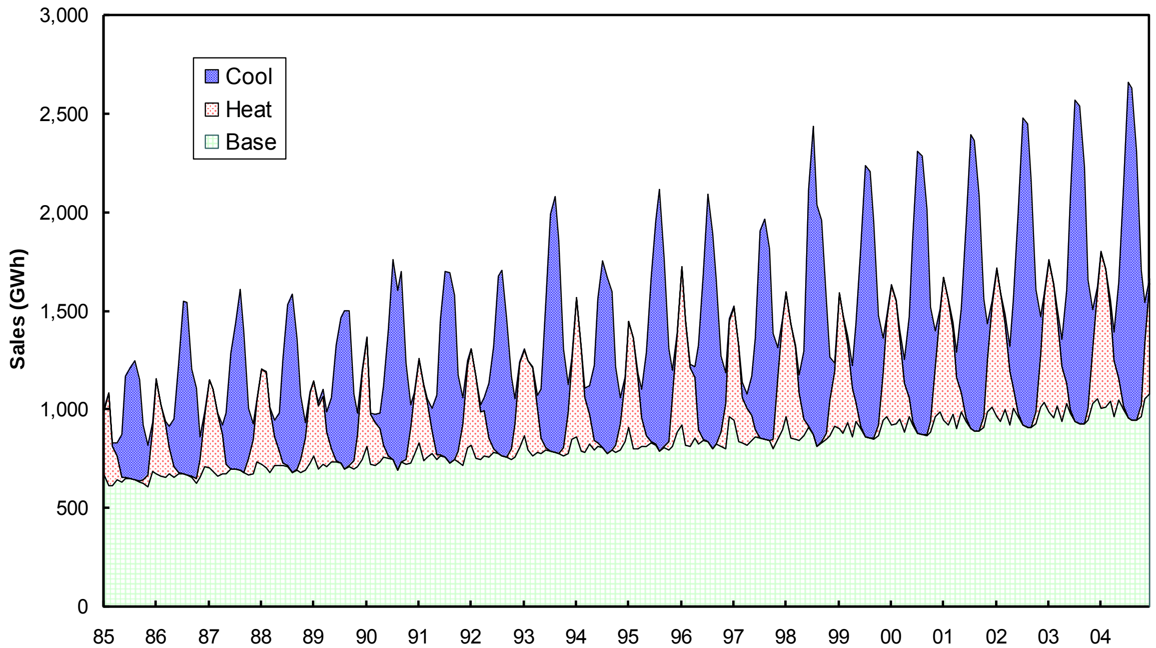


Figure 14 provides a plot of the contribution of each end-use component to the overall predicted value. As can be seen from the figure, other uses contribute significantly to overall monthly sales. There are also significant monthly variations in these base sales that are driven by variations in billing days and seasonal cycles associated with lighting, water heating, and refrigeration.

Figure 14: End-Use Contribution to Total Predicted Sales (GWh)



Conclusion

The SAE approach provides a powerful framework for development of short-run and long-run energy forecasts. Historical data and forecast assumptions about end-use equipment stocks and efficiency levels are inputs at the annual level. These explanatory variables are then used in to construct variables that are used in a monthly regression model to estimate multipliers and trend adjustments that provide the best historical fit. The estimated coefficients are highly significant and within expected ranges.

By construction, this approach gives estimates of weather sensitivity that vary over time, reflecting changes in equipment shares and efficiency levels as well as estimated trend adjustments. And these sensitivities will continue to change in the forecast period to reflect further changes in equipment stocks. Like time-varying parameter approaches, this has the advantage that weather sensitivity values will be different at the end of the historical period, which is important for short-term forecasting and weather normalization in the near-by years.

In addition to providing short-term forecasting and analysis capabilities, the model is sufficiently structured for development of long-run forecasts and the forecast results can be decomposed into end-use components. Like detailed end-use models, this provides a rich story behind the forecast.