AEIC Load Research Committee

Estimation Errors In Demand Response with Large Customers

Applications for Load Research

November 2009

Acknowledgements

The AEIC Load Research Committee would like to acknowledge the contributions of the author of this paper: Joseph Lynch, Ph.D., LRC member and manager of Resource Planning in South Carolina Electric & Gas Company.

Table of Contents

White Paper Focus

The primary focus of this paper is to estimate how well a customer's profile can be estimated and how confident we can be in estimating what a single customer's load profile would have been on a demand response (DR) event day in the absence of any demand reduction. In this paper, several heuristic methods and several regression models were used to generate distributions of errors resulting from the estimation of 365 customer load profiles.

The results of this work were first presented at the Marcus Evans Resource Planning Conference in Atlanta, Georgia in September 2008. Results were again presented at the AEIC Load Research Committee meeting in Charleston, SC in October 2008 under the less descriptive title "DSM Measurement & Verification from a Load Research Perspective."

1 Introduction

1.1 Introduction to DSM, DR, and BLP

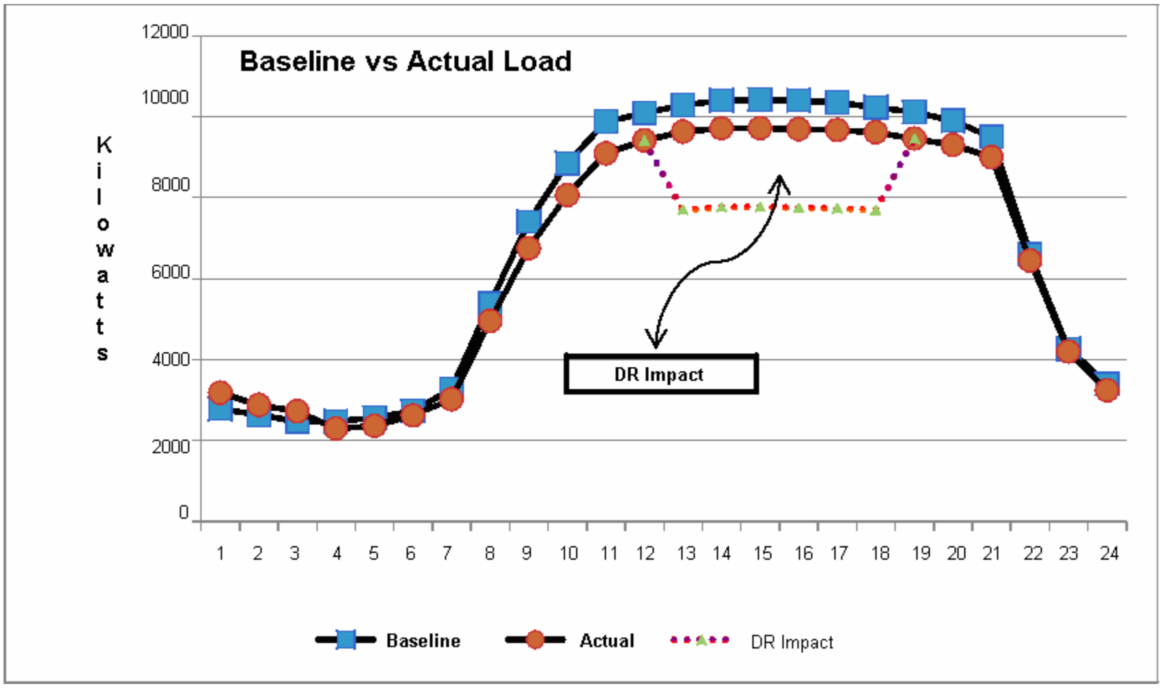
It is generally recognized that the country's power markets cannot be fully efficient without integrating demand-side management (DSM) into market functioning and that demand response (DR) is a key component of DSM (FERC, 2007)[1]. Demand Response refers to programs that result in a lower customer demand for a few hours per event -- usually to help mitigate the impact of high market prices or high system demands. In this paper, estimation errors for DR for large commercial and industrial customers will be discussed. "Large" refers to customers with a peak demand of 1,000 kW or higher.

One of the most significant challenges to DR implementation is estimation of the DR impact which is needed for reasons of system reliability as well as for financial equity related to customer compensation. It is easy to use an interval data recorder to measure what a customer's actual hourly loads were on a particular day. In order to measure the DR impact, it is necessary to estimate what the customer's hourly loads would have been absent the call for energy curtailment. Base line profile (BLP) is the term given to what the customer's load profile would have been absent the DR event. This paper concerns errors in estimating the BLP and what should be done about them. The terms "forecasting" and "estimation" are used interchangeably in this paper. With estimation, all relevant data are typically used while forecasting is limited to data occurring prior to the event. That said, both tasks are related in the sense that if a good estimation technique can be identified, then it is likely to suggest a good forecasting technique.

Before continuing, it is important to make the distinction between error in estimating the BLP and error in estimating the DR impact. Versions of the graphic on the following page have been seen in many papers on this topic. In the Demand Response Impact Example (Fig. 1.1), the actual load profile of a particular customer was assumed to produce a 20% DR impact from 12pm-6pm. For simplicity in this example, we will estimate the BLP to be 1,000 kW, with an estimation error of 3%, or +/-30 kW. The DR impact of 20% equates to 200 kW. Therefore the error associated with the estimated DR impact will be 30 kW divided by 200 kW or 15%.

The BLP was the projection developed using one of the forecasting methods discussed later in the paper. If there is +/-3% error in the BLP, then there is +/-15% error in the estimate of the DR impact, i.e., 3% divided by 0.20. In this case there is a five-fold increase in the percentage of error as the necessary result of using the smaller number in the denominator. The point to keep in mind is that the relevant error is the 15% error – not the 3% error. Clearly the goal of the estimation process is to estimate the DR impact – not the BLP. This paper will discuss the estimation error of the BLP. The reader should keep in mind that – the relevant error around the DR impact estimate will be significantly higher, perhaps 5 times higher, as in this example.

Figure 1.1 – Demand Response Impact Example



1.2 Data and the AEIC Load Research Committee

The Association of Edison Illuminating Companies (AEIC) is a utility organization that was formed by Thomas Edison in 1885. The AEIC Load Research Committee (LRC) is one of six AEIC committees dedicated to various aspects of the electric utility profession. The purpose of the LRC is to advance the practice of load research which involves the collection, analysis and interpretation of the load profiles of customers. It is only natural that the members of the LRC would be intimately involved in the measurement and verification (M&V) of DR.

Seven members of the committee provided the load profiles and associated weather profiles of a number of commercial and industrial customers for the summer of 2007. The customers were kept anonymous except for their classification as commercial or industrial. Hourly load data from May 1 through August 31 of 2007 was provided for 365 customers evenly split between commercial and industrial. The data covered 14 states with 20 different weather patterns from ERCOT and the Eastern Interconnect.

2 Methodology

2.1 Forecasting Methods for the BLP

Today DR is an important component of system resources so there are currently methods being used to estimate it. The most well studied and documented are the methods used by various Independent System Operators (ISOs) and Regional Transmission Organizations (RTOs) around the country. The table in Fig. 2.1 summarizes four methods for the ISOs of New England (NE), New York (NY), California (CA) and the PJM RTO. (PJM is the RTO coordinating wholesale electricity in all or parts of 13 states and the District of Columbia.)

Figure 2.1 – Four Methods for the ISOs of New England (NE), New York (NY), California (CA) and the PJM RTO

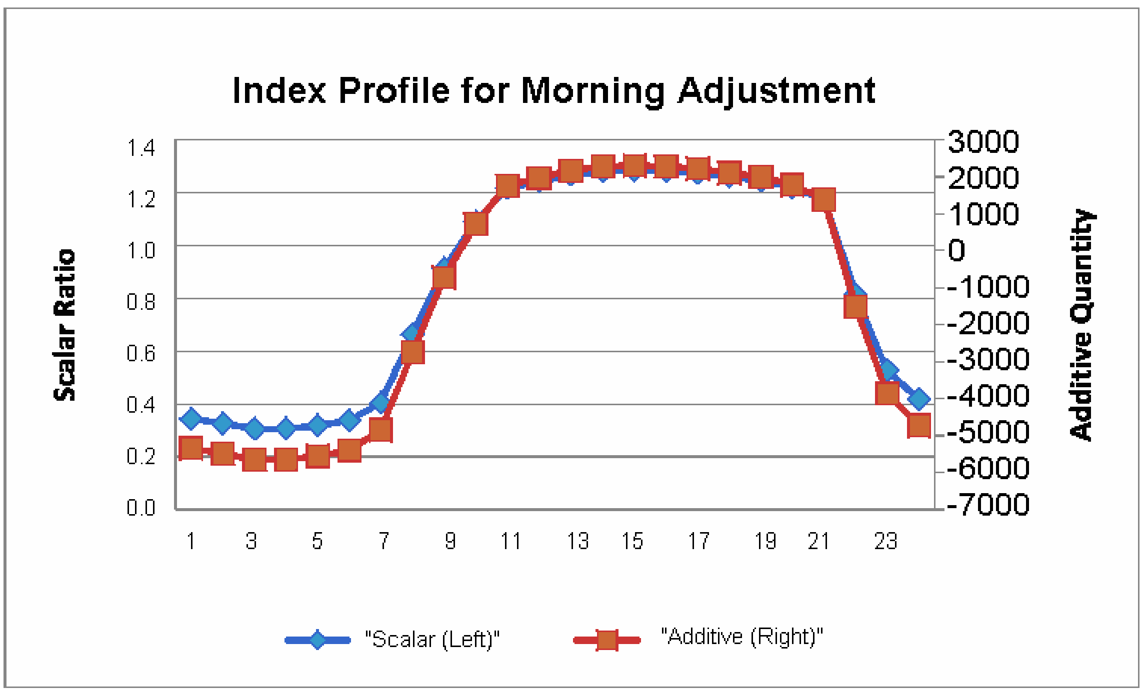
- **ISO NE:** uses exponential smoothing
  - BLP=0.1*previous day + 0.9 * previous BLP
  - Starting BLP is average of first 5 eligible days.
  - Uses morning additive adjustment
- **ISO NY:**
  - Profile from 5 highest days of last 10 eligible days
  - Eliminate days where KWH is less than 25% of prior day.
  - Morning scalar adjustment using 3^{rd} and 4^{th} hour prior to event
- **ISO California:**
  - Profile from 3 highest days of last 10 eligible days
  - No morning adjustment
- **PJM RTO:**
  - Profile from 5 highest days of last 10 eligible days
  - Use days where KWH greater than 75% of average profile.
  - Morning scalar adjustment using hours 9 and 10 am

A good exposition of these methods is contained in “Evaluating Baselines for Demand Response Programs (Clifford Grimm, 2008)[2]. All four methods are based on the load profiles from “eligible days” which exclude weekends, holidays and DR event days. The methodologies of ISO NY, CA and PJM RTO are similar in that they all use the simple average profile of 3 or 5 days out of the previous 10 eligible days to form the BLP. ISO NE uses an exponential smoothing algorithm which takes a weighted average of all eligible days prior to the DR event day with the weights declining geometrically in time. Although being reconsidered, ISO CA is the only one of the four that does not use a “morning adjustment”. Morning adjustment refers to the practice of using a customer’s actual load from the day of, but prior to, the DR event. These methods used by the ISOs are referred to as heuristic to distinguish them from statistical methods such as regression analysis used later in the paper.

2.2 Morning Adjustment

This use of data from the DR event day or "morning adjustment", is problematic because of possible bias resulting from the DR event itself. The morning adjustment cannot be ignored however because, as will be seen, it does significantly increase the explanatory power of most forecasting models. This result agrees with KEMA (2003)[3] and Lawrence Berkeley National Lab (LBNL) (2008)[4]. This morning adjustment bias can arise legitimately when customers pre-cool their premises or change a production schedule in anticipation of the DR event. The bias can also be the result of gamesmanship when customers try to unnaturally increase their morning load anticipating that the morning adjustment would increase their BLP resulting in higher payments for the resulting increased estimate of their DR impact. More will be said about detecting gaming later in the paper. A scalar morning adjustment or an additive adjustment can be made. When using a morning adjustment, the BLP no longer relies on the average kW profile of previous eligible days but uses instead an average "indexed" profile. In this paper the average load between 8am and 10am, i.e., the average of the 9th and 10th hour of the day, will be used to adjust each day's kW profile. For the scalar adjustment each kW value of an eligible day is divided by the average of the 9th and 10th kW value to create an "indexed" profile. For the additive adjustment the average of the 9th and 10th kW value is subtracted from each kW value of the eligible day.

Figure 2.2 – Index Profile for Morning Adjustment



The Index Profile for Morning Adjustment (Fig. 2.2) shows the average "indexed" profile for a particular customer where the average is taken over five eligible days. The shapes of the two curves are similar but differences will become more significant, when expanded, because the scalar index will be multiplied by a day's morning kW reading while the additive index will be added to it.

3 Results

3.1 Preliminary Results

In the following analysis, a DR event is assumed to occur from 12pm to 6pm. The quantity to be estimated is the average load for these six hours. Of course the load in each of the six hours is important as would be the maximum load over the time period. This paper focuses on one quantity, i.e. the average of the six afternoon loads, so that conclusions can more easily be presented and understood.

It is also believed that the results regarding error with the average would at least be indicative of the magnitude of error that would result if each hour were analyzed separately.

Hottest Day Results

Figure 3.1 – Average Percent Error for ISO NE (Hottest Day)

Average Percent Error					
Exponential Smoothing (ISO NE)					
	Hottest Day Results				
Percentiles	**10**	**25**	**50**	**75**	**90**
No Adjustment	**-16.9**	**-10.8**	**-5.7**	**-1.1**	**8.7**
Additive Adj.	**-9.1**	**-2.3**	**0.4**	**2.8**	**10.0**
Scalar Adj.	**-9.2**	**-3.0**	**0.4**	**2.6**	**10.4**

Fig 3.1 summarizes the forecast average percent error associated with the BLP methodology used by ISO NE for the hottest day. There are 365 customer profiles being estimated and the table presents certain percentiles from the distribution of errors. At the 50th percentile in the distribution of errors, half of the accounts are above or below the average percent error shown. The "No Adjustment" row contains the results when the morning adjustment is not employed. At the 50th percentile, the average percent error of -5.7%, suggests that the methodology is somewhat biased and tends to understate the actual load profile. When a morning adjustment is applied, whether a scalar or additive one, the bias seems to disappear and the error tends to be smaller at just 0.4%. This is precisely why ISO NE uses a morning adjustment in their process.

Figure 3.2 – Average Percent Error for ISO NY (Hottest Day)

Average Profile (ISO NY)					
	Hottest Day Results				
Percentiles	**10**	**25**	**50**	**75**	**90**
No Adjustment	**-8.9**	**-5.4**	**-1.3**	**2.5**	**16.1**
Additive Adj.	**-8.5**	**-2.6**	**0.2**	**3.2**	**9.6**
Scalar Adj.	**-8.8**	**-2.6**	**0.2**	**2.9**	**10.4**

The Average Percent Error for the ISO NY is shown in Fig 3.2. There is less bias in the distribution and the forecast error seems smaller. Since the NY approach uses only the last 5 highest load days while the NE method of exponential smoothing uses all prior eligible days, it should be expected that the NY approach would work better for the hottest day. The morning adjustment clearly improves the forecast accuracy with no apparent significant difference between the additive and the scalar approach. Both methods produce average percent errors that are centered closer to zero (0.25%) at the 50th percentile.

5th Hottest Day Results

Figure 3.3 illustrates the results for the fifth hottest day for both ISO NE and ISO NY, for reference in the event of any unusual circumstances for the hottest day.

Figure 3.3 – Average Percent Error for ISO NE and ISO NY (5th Hottest Day)

Average Percent Error					
Exponential Smoothing (ISO NE)					
	5th Hottest Day Results				
Percentiles	**10**	**25**	**50**	**75**	**90**
No Adjustment	**-14.4**	**-8.1**	**-3.5**	**0.4**	**12.7**
Additive Adj.	**-9.3**	**-2.6**	**0.5**	**3.9**	**12.0**
Scalar Adj.	**-9.2**	**-2.7**	**0.4**	**3.7**	**13.9**
Average Profile (ISO NY)					
	5th Hottest Day Results				
Percentiles	**10**	**25**	**50**	**75**	**90**
No Adjustment	**-8.1**	**-3.7**	**0.6**	**5.5**	**22.5**
Additive Adj.	**-9.0**	**-2.4**	**0.5**	**4.5**	**12.8**
Scalar Adj.	**-9.5**	**-2.7**	**0.5**	**4.1**	**12.7**

These results from the fifth hottest day seem similar to the results from the hottest day.

10 Hottest Days Results

The metric that will be used to judge the error of each methodology is the mean (or average) absolute percent error (MAPE) over the 10 hottest days of the summer. By using an average of 10 days, any idiosyncrasies related to a particular day are somewhat mitigated. In addition, use of the absolute error avoids the cancelling out of large positive and negative errors. Here we are more interested in estimating the magnitude of the error while taking into account the different scales among the customers analyzed, not whether the error is too high or too low. There are other metrics that can be justified, but this 10 day MAPE is certainly reasonable.

Figure 3.4 – MAPE for ISO NE and ISO NY (10 Hottest Days)

Average Absolute Percent Error					
Exponential Smoothing (ISO NE)					
	10 Hottest Days Results				
Percentiles	**10**	**25**	**50**	**75**	**90**
No Adjustment	**3.0**	**4.5**	**7.0**	**12.0**	**38.4**
Additive Adj.	**1.4**	**2.3**	**4.0**	**8.0**	**24.3**
Scalar Adj.	**1.3**	**2.2**	**4.0**	**8.1**	**26.7**
Average Profile (ISO NY)					
	10 Hottest Days Results				
Percentiles	**10**	**25**	**50**	**75**	**90**
No Adjustment	**2.3**	**3.5**	**5.6**	**11.4**	**44.3**
Additive Adj.	**1.4**	**2.3**	**4.2**	**8.4**	**24.8**
Scalar Adj.	**1.4**	**2.2**	**4.2**	**8.7**	**32.3**

Fig. 3.4 shows the results using the MAPE for the two methods discussed thus far.

When using this metric, there does not seem to be much difference in the forecasting ability of the two methods. The ISO-NY methodology might perform slightly better when the morning adjustment is not used; however, with a morning adjustment, the distribution of MAPE errors seems very similar.

3.2 Considerations

Before proceeding, it is worthwhile to highlight two considerations:

1. Using MAPE we can rank order the forecasting errors from lowest to highest. The low end of the distribution of MAPEs is not of great interest. Rather, the median or 50th percentile and above are the critical points to examine in the distribution. In the previous example, the large values at the 90th percentile indicate that forecasting error will be unacceptable for at least 10% of the customers in our analysis.

2. These tables contain the distribution of errors in estimating the BLP. As discussed earlier, if a 20% DR impact is expected, then the error in estimating the DR impact will be some multiple of the entry, perhaps 5 times the entry in the table. For example, in the ISO NE methodology with an additive adjustment, the 4.0% error in the BLP could become a 20.0% error in the DR impact estimate.

4 Analysis

4.1 Regression Analysis

The focus of this paper is to estimate how well a customer's profile can be estimated or, put another way, how confident we can be in estimating what a single customer's load profile would have been absent a call for Demand Response. This is basically a statistical problem, so we turn now to statistical methods. Ordinary least squares (OLS) regression analysis is one of the best tools available to explain variations in customer loads. While the heuristic methods of ISO-NE and ISO-NY used only the prior 10 eligible days, the regression techniques will use all eligible days.

There were three different regression equations specified which are summarized in Fig. 4.1. All three equations use a "zero-one" indicator variable "I" to capture day-of-week effects. The "Weather Model" relies on daily cooling degree days (CDD) to explain changes in demand. The "Same Day Model" uses CDD and the average kW for two hours in the morning of the DR event. This is a morning adjustment type model. Lastly there is the "Previous Day Model" which uses the dependent variable from the prior day and the change in CDD to explain the dependent variables. The motivation for this model is to see if there is a way to avoid the morning adjustment and still maintain accuracy. Because of the potential for bias in the morning adjustment, it would be very useful to have this model perform well. The dependent variable in all the regression equations "Avgkw" is the average load from 12pm to 6pm which is the same quantity being forecasted by the heuristic techniques.

Figure 4.1 – Three Different Regression Equations

- **Weather Model**
  - **Avgkw**=a + b***CDD** + c***I*m*** + d***I*t*** + e***I*w*** + f***I*h***
- **Same Day Model**
  - **Avgkw**=a + b***AMKW** + c***CDD** + d***I*m*** + e***I*t*** + f***I*w*** + g***I*h***
- **Previous Day Model**
  - **Avgkw**=a + b***lag(avgkw)** + c***delta(CDD)** + d***I*m*** + e***I*t*** + f***I*w*** + g***I*h***

WHERE:

Avgkw = average of load from Noon through 6pm (KW13-KW18)
CDD = cooling degree days
AMKW = average of KW at 9 and 10 am of event day
I*m*,I*t*,I*w*,I*h* = indicator variables (0,1) for days of the week M,T,W,H

Figure 4.2 provides the estimation results for the hottest day for the three different regression equations. None of the models seemed significantly biased and, as expected, the morning adjustment as measured in the same day model ("Same Day kW") seems to increase the forecast accuracy.

Figure 4.2 – Results for the Hottest Day by Three Different Regression Methods

OLS Regression					
	Hottest Day Results				
Percentiles	**10**	**25**	**50**	**75**	**90**
Weather	**-7.3**	**-2.8**	**0.3**	**3.4**	**9.5**
Previous Day KW	**-8.8**	**-4.1**	**-1.4**	**2.0**	**11.2**
Same Day KW	**-5.9**	**-2.0**	**0.2**	**2.0**	**6.8**

The MAPE for the 10 hottest days is presented in Fig. 4.3. The morning adjustment clearly helps reduce the forecast error at each point in the distribution but especially at the high end. At the 90th percentile the error decreases from 34.3% and 29.6% to 17.8%. The results of the attempt to avoid a morning adjustment with the "Previous Day kW" model are disappointing. Its accuracy is only slightly better than the weather model. Use of the morning adjustment, i.e. actual loads from the morning of the DR event, clearly increases the explanatory power of the regression equation.

Figure 4.3 – Results for the 10 Hottest Days by Three Different Regression Methods

OLS Regression					
	10 Hottest Days Results				
Percentiles	**10**	**25**	**50**	**75**	**90**
Weather	**1.7**	**2.4**	**4.3**	**8.9**	**34.3**
Previous Day KW	**1.7**	**2.5**	**4.1**	**8.5**	**29.6**
Same Day KW	**1.0**	**1.6**	**2.6**	**5.6**	**17.8**

Each row in Fig 4.3 represents the results from 365 regression equations. The statistical estimation results from each equation will not be analyzed. However, since the volume of data is so great and OLS is so sensitive to outliers, it is valuable to validate the results, at least to some extent, by turning to robust estimation techniques.

4.2 Robust Regression

Robust regression refers to a class of estimation techniques that can identify outliers and produce parameter estimates that are resistant to outliers. Additionally when there are no outliers, the parameter estimates are still unbiased but do not have minimum variance. Thus, there is a small efficiency penalty in using robust estimation techniques when there are no outliers in the data and simple OLS should be used. The estimation algorithm uses iteratively re-weighted least squares to estimate parameters, which is more complicated than OLS where parameter estimates are solutions to a closed form equation. This complication may explain why good robust techniques have not been available in some statistical software packages. SAS® statistical analysis software has added an easy- to-use robust regression technique in Version 9.1. PROC ROBUSTREG[5] is the function that is now available. The MAPE forecast results, derived using robust regression, are shown in Fig. 4.4. The distributions of errors are remarkably similar to the OLS results. This provides a degree of confidence in both sets of results. Since the robust regression will identify outliers, it may be interesting to see the errors when the outliers are eliminated.

Figure 4.4 – MAPE Errors of Robust Regression for 10 Hottest Days

Robust Regression					
	10 Hottest Days Results				
Percentiles	10	25	50	75	90
Weather	1.6	2.2	3.9	8.2	34.5
Previous Day KW	1.5	2.3	3.8	8.0	27.4
Same Day KW	0.9	1.5	2.5	5.4	13.2

Fig 4.5 illustrates the distribution of MAPE errors for the 10 hottest non-outlier days. With outliers taken out of the results, the errors are much smaller, especially at the upper end of the distribution. In fact, at the 90th percentile in the "Same Day kW" distribution, the error of 6.1% is at least approaching a tolerable level. Outliers can occur for a number of reasons, such as an industrial customer shutting down for maintenance or a commercial customer closing temporarily for renovations. This type of information may or may not be available to a utility or ISO for their DR program.

The results suggest that it would be advantageous to include such information if possible. However, incorporating such detailed customer-specific information in this paper is beyond the scope of this research.

Figure 4.5 – MAPE Errors of Robust Regression for 10 Hottest Days without Outliers

Robust Regression					
	10 Hottest Days Results Without Outliers				
Percentiles	10	25	50	75	90
Weather	1.3	1.9	3.2	6.8	20.3
Previous Day KW	1.3	2.0	3.2	6.1	17.6
Same Day KW	0.6	1.0	1.7	3.2	6.1

4.3 Summary

The table in Fig. 4.6 summarizes the distribution of errors for the best of the approaches studied. For ISO-NE and ISO-NY, the results are based on the additive adjustment approach; for the regression technique, the results are from the "Same Day kW" model. Additionally, the OLS and robust regression approaches were re-estimated using only the 30 hottest eligible days of the summer. Since our metric is based on the 10 hottest days, it was thought that a regression equation based only on hot days might perform better than one based on all days.

Figure 4.6 – Summary Table for Different Approaches

Average Absolute Percent Error					
Summary of Best Results					
	10 Hottest Days Results				
Percentiles	**10**	**25**	**50**	**75**	**90**
Exponential	**1.4**	**2.3**	**4.0**	**8.0**	**24.3**
Average Profile	**1.4**	**2.3**	**4.2**	**8.4**	**24.8**
OLS Regression	**1.0**	**1.6**	**2.6**	**5.6**	**17.8**
Robust Regression	**0.9**	**1.5**	**2.5**	**5.4**	**13.2**
OLS Regression (30 Hottest)	**0.9**	**1.4**	**2.5**	**5.4**	**13.6**
Robust Regression (30 Hottest)	**0.8**	**1.3**	**2.4**	**5.2**	**12.6**

Focusing on the 50th and higher percentiles in each distribution, it is clear that the regression techniques outperform the heuristic techniques. It also appears that all the regression results are similar. Even the regression results based on models estimated with only the 30 hottest eligible days ("30 Hottest") perform as well but no better than the models estimated with all the eligible days.

Before proceeding with what to do with these results, Fig 4.7 illustrates a separate analysis for commercial and industrial classes.

Figure 4.7 – Summary Table for Different Approaches on Commercial and Industrial Classes

Average Absolute Percent Error						
Summary of Best Results						
	10 Hottest Days Results					
	Commercial			Industrial		
Percentiles	**10**	**50**	**90**	**10**	**50**	**90**
Exponential	**1.2**	**3.5**	**16.4**	**1.7**	**4.9**	**32.9**
Average Profile	**1.2**	**3.7**	**19.3**	**1.7**	**5.1**	**34.7**
OLS Regression	**0.9**	**2.3**	**10.4**	**1.1**	**3.2**	**26.7**
Robust Regression	**0.8**	**2.2**	**9.3**	**1.0**	**3.3**	**23.3**
OLS Regression (30 Hottest)	**0.8**	**2.1**	**10.0**	**1.0**	**3.3**	**23.1**
Robust Regression (30 Hottest)	**0.8**	**2.0**	**9.2**	**0.9**	**3.0**	**23.6**

It appears that the industrial loads are more volatile. While beyond the scope of this paper to pursue, a likely explanation may be that industrial customers are more prone to have maintenance days or vacation outages than the commercial class. This emphasizes the need for the utility or ISO to use specific customer information as part of the DR program where industrial customers are concerned.

5 Other Analysis

5.1 Cut-Off Error Rate: Saying No to Some Customers

For the sake of system reliability and for the sake of financial equity, it is critical that the estimates of DR impact be reliable. This suggests that a cut-off error rate be adopted to eliminate from the DR program those customers with load profiles that are too volatile and unpredictable. Referring to the summary results table (Fig. 4.6), a reasonable expectation for the average error in the BLP would be approximately 2.5%. The corresponding error for the DR impact might be 5 times as great -- i.e., 12.5%. This may be an acceptable error rate when considering the value of the DR impact to the system and the fact that, when loads are estimated for a single customer, the estimation errors will usually be much higher than when doing so for a group of customers, or for the system as a whole. However, referring again to the results in Fig. 4.6, a DR program with unrestricted access could have 25% of its participants with error rates higher than 5.4% (the 75th percentile). A 5.4% error in the BLP could mean a 27.0% error in the DR impact. This may or may not be acceptable, but clearly a point can be reached at which no confidence can be placed in the estimates for some customers.

Assuming a 3% cut-off rate for the BLP estimation error, we find that 232 customers out of 365, or 64%, meet the criterion and would be allowed in the DR program, with 133 excluded. Prior research in this area suggests there is a need for flexibility in choosing a forecast methodology, and the data in Fig. 5.1 support this advice. A variable “icount” was created to count in how many of the 6 estimation methods each customer would meet the 3% cut-off criterion. The table shows the output of a PROC FREQ analysis created in SAS.

Figure 5.1 – SAS Output for 3% Cut-off Rate for BLP Estimation Error

The FREQ Procedure

icount	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	12	5.17	12	5.17
2	45	19.40	57	24.57
3	12	5.17	69	29.74
4	62	26.72	131	56.47
5	18	7.76	149	64.22
6	83	35.78	232	100.00

For example, there are 12 customers (shown under “Frequency”) who meet the criterion under only 1 estimation method (shown under “icount”). There are 45 customers meeting the criterion with 2 methods, and finally, 83 meeting the criterion in all 6 methods.

When the criterion becomes more restrictive and is set at 1%, only 61 of the 365 customers would meet the criterion. Alternatively, with a less restrictive criterion of 5%, 290 customers would meet the criterion. It is also interesting to note how many customers meet the criterion by each method. This would provide a glimpse of the result of a utility or ISO restricting the estimation method to only one technique. The next table (Fig. 5.2) presents these results using the 3% criterion.

Figure 5.2 – Summary Table for Number of Customers Meeting 3% Criterion by Methods

Number Meeting 3% Criterion By Methods		
Method	Count	%
Exponential	108	30%
Average Profile	103	28%
OLS Regression	181	50%
Robust Regression	190	52%
OLS Regression (30 Hottest)	194	53%
Robust Regression (30 Hottest)	198	54%

It was shown in the summary results table (Fig. 4.6) that the exponential method of ISO-NE and the average profile method of ISO-NY did not perform as well as the regression techniques. Here we see a similar result: Fewer customers meet the 3% cutoff. The regression techniques would allow more than two-thirds as many customers into the DR program as the heuristic methods if there were a 3% cut-off error criterion.

Next, we examine a possible option for a customer who is excluded from a DR program because of a volatile load profile.

5.2 Aggregation: A Possible Solution

A customer whose load profile contains too much random variation could join with other customers and offer their combined profile for entry into the DR program. Ideally, the random variations in each would somewhat offset the others, producing a sufficiently predictable profile that would add value to the DR program.

To test this approach, each of the 365 customers in our database was grouped with four randomly chosen customers from the same state. The sum of the 5 resulting load profiles was then analyzed. Figure 5.3 illustrates the results for the ISO-NY methodology. The non-aggregated results shown previously are repeated here for reference.

Figure 5.3 – Average Profile for Aggregated 5 Random Accounts by State

Average Profile (ISO NY) - *Aggregated* 5 Random Accounts By State					
	10 Hottest Day Results				
Percentiles	**10**	**25**	**50**	**75**	**90**
No Adjustment	**2.3**	**3.2**	**4.7**	**8.0**	**13.6**
Additive Adj.	**1.5**	**2.0**	**3.0**	**5.9**	**10.2**
Scalar Adj.	**1.5**	**2.0**	**3.0**	**5.8**	**10.4**
Non-Aggregated Results For Reference					
No Adjustment	**2.3**	**3.5**	**5.6**	**11.4**	**44.3**
Additive Adj.	**1.4**	**2.3**	**4.2**	**8.4**	**24.8**
Scalar Adj.	**1.4**	**2.2**	**4.2**	**8.7**	**32.3**

Clearly the distribution of errors improves. This is particularly apparent when comparing the MAPEs for the 90th percentile of the aggregated results with the original non-aggregated results. Under the scalar adjustment, for example, a 32.3% error at the 90th percentile (non-aggregated results) drops significantly to a 10.4% error (aggregated results.) It is projected that results would be even more dramatic with aggregation of more than 5 customers, or if there were a wider pool of customers from which to choose than was available in this database.

5.3 Detecting Gaming: A Way to Instill Confidence

It has been seen in this paper that a morning adjustment can make a significant improvement in the BLP estimate. However, when the morning adjustment is used, the issue of gaming may cast a shadow over the results. Consequently, confidence in the estimates of DR impacts and trust in the equity of the process may suffer. One solution to this dilemma could be a method of detecting gaming when it occurs. One approach is to produce two estimates of the BLP – one with the morning adjustment and one without. If estimates are sufficiently close, it is reasonable to conclude that the customer is not engaged in gaming. The table in Fig. 5.4 shows the mean absolute percent difference for the 10 hottest days between estimates made with the morning adjustment and estimates made without it.

Figure 5.4 – The Mean Absolute Percent Difference between Estimates for 10 Hottest Days: With and Without Morning Adjustment

Mean Absolute Percent Difference					
Between Forecasts With and Without Morning KW					
	10 Hottest Days Results				
Percentiles	**10**	**25**	**50**	**75**	**90**
Exponential	3.2	4.9	7.3	13.9	31.2
Average Profile	2.8	4.3	6.5	12.4	27.1
OLS Regression	0.9	2.0	3.7	7.9	19.0
Robust Regression	0.9	2.0	3.2	6.6	19.2
OLS Regression (30 Hottest)	0.8	1.5	3.1	7.6	17.3
Robust Regression (30 Hottest)	0.8	1.4	3.0	7.0	18.0

Based on the results at the 75th percentile level in the distribution of mean absolute differences for the regression techniques, a reasonable gaming detection protocol might require that the two estimates be within 10% of each other. If they fall beyond this limit, then the possibility of gaming is present and further analysis or research would be needed. With the presence of a gaming detection protocol, the few unscrupulous customers who might otherwise try to beat the system would be discouraged from doing so. The vast majority of customers who would not employ gaming would have more confidence that the system is fair and that everyone is “playing by the same rules.”

6 Summary and Conclusions

The primary purpose of this research was to answer the question: How well can we estimate the load profile for a single large customer on a DR event day in the absence of any demand reduction? Several heuristic methods and several regression equations were used to generate distributions of errors resulting from the estimation of 365 customer load profiles. A cut-off error level was examined to eliminate some profiles with very large errors. An aggregation option was studied to suggest that it might offer hope for these customers to participate in a DR program. Finally a solution to the problem of gaming was addressed.

Here are the conclusions:

1. Practicality dictates the need for selection criteria such as a cutoff error rate.
2. Approximately two-thirds of customers in the study met a 3% error cutoff. Similar results might hold in a particular market or utility service area.
3. The percent of error in estimating the DR impact can be significantly larger than the percent of error in estimating a customer's load profile. This is critical since the ultimate purpose of the estimation is to ascertain the DR impact.
4. There is a need for flexibility in choosing an estimation technique. There is certainly merit in the qualities of simplicity and understandability of the technique, but it can be safely said that accuracy and dependability trump these qualities.
5. The use of morning loads from the event day significantly increases accuracy and is therefore recommended. The use of morning loads, however, also creates concern for "gaming the system."
6. Gaming may occur with reliance on morning kW; therefore, gaming detection methods should be developed. This paper compared an estimate with and without the morning adjustment to detect gaming, the results of which suggest this protocol would work in practice.
7. Customers rejected by the cut-off selection criterion may be able to meet the standard and enter the DR program through aggregation with others.

7 References

1 "Demand Response & Advanced Metering 2007", FERC Staff Report, September 2007, http://www.ferc.gov/legal/staff-reports/09-07-demand-response.pdf

2 "Evaluating Baselines for Demand Response Programs", Clifford Grimm, 2008 AEIC Load Research Workshop, paper and presentation:
http://www.aeic.org/load_research/docs/12_Evaluating_Baselines_for_Demand_Response_Programs_PAPER.pdf
http://www.aeic.org/load_research/docs/12_Evaluating_Baselines_for_Demand_Response_Programs_PRESENTATION.pdf

3 "Protocol Development for Demand Response Calculation – Findings and Recommendations", KEMA-Xenergy Report to California Energy Commission, February 2003, http://www.energy.ca.gov/reports/2003-03-10_400-02-017F.PDF

4 "Estimating Demand Response Load Impacts: Evaluation of Baseline Load Models for Non-Residential Buildings in California", Lawrence Berkeley National Lab (LBNL), January 2008, http://repositories.cdlib.org/lbnl/LBNL-63728

5 SAS/STAT® 9.2 User's Guide, Second Edition, Proc ROBUSTREG Statement : http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/statug_rreg_sect007.htm