STATE OF ILLINOIS
ILLINOIS COMMERCE COMMISSION

COMMONWEALTH EDISON COMPANY	)	
	)	
Investigation of Commonwealth Edison	)	**Docket No. 14-0384**
Company's Cost of Service for Low-Use	)	
Customers in Each Residential Class	)	

RATE DESIGN COMPARISONS

CUB EXHIBIT 1.02

DECEMBER 4, 2014

Study of the Correlation between Usage and Demand for Non-Space Heat Residential Consumers of ComEd

Executive Summary

1. Analysis of data provided by ComEd demonstrates that distribution cost is not driven by the number of ratepayer accounts and that all of the distribution costs should be allocated on the basis of energy usage.

2. Temporary low usage caused by unoccupied apartments and single family homes does not affect the conclusion that no amount load is correlated with usage and not the existence of a consumer account. The same is true for low use caused by home vacancy due to vacations.

3. Load factors tend to be better for low use consumers than for high use consumers in the residential class when correctly measured on a coincident peak or on a class basis.

4. ComEd's load research data contains many errors and biases that maybe increasing costs to residential consumers by a large margin.

5. There have been dramatic changes in the multi-family load factor over the years which have increased cost allocation and prices to the class.

6. The load factor for multi-family consumers is more than 39% inside the City of Chicago while it is about 32% outside of the City. The load factor of single family consumers is also higher inside the City than outside the City although there are biases in ComEd's measurement of City of Chicago single family loads which makes the comparisons less rigorous. The difference in load factors implies that outside city consumers should be allocated 23% more distribution cost than City consumers.

Objectives of the Analysis

Flaws in ComEd's Residential Usage Study (ComEd Exhibit 2.33) can be demonstrated by constructing an objective analysis derived from statistical evaluation of usage and load research data. The basic propositions that should be tested in an objective analysis of demand and usage include the following:

1. What is the correlation between maximum energy usage in an hour, day or month relative to the average energy usage over a year (i.e. how much of the variation in peak demand can be explained by variation in average use)? This correlation between alternative measures of demand and energy use can be tested by performing statistical analysis of demand and energy for individual consumers or alternatively by examining the correlation for groups of consumers. Because of differences in the relationship between usage and peak across individual consumers, the discussion below demonstrates that it is better to evaluate the correlation by creating consumer group increments according to usage level and then correlating demand and energy for the usage level groups.

2. If there is some randomness in the relationship between usage and demand across individual consumers, should this randomness be attributed to the existence of a ratepayer account? This issue involves whether some of the variation in demand is explained by usage and other variation is explained by the simple presence of a ratepayer account. The question of whether some variation in demand is completely independent of usage can be tested by arranging ratepayers into different usage groups as described above, and then evaluating the level of the intercept in a regression equation. The reason this analytical approach of aggregating consumers into groups and then running a regression equation on the usage groups addresses the issue of whether any randomness in demand comes from the existence of a customer account is demonstrated through modeling alternative relationships using Monte Carlo simulation.

3. What is the load factor for groups of consumers with different usage levels? If the proposition of ComEd is correct that usage is independent of load, then the load factor of large consumers should be much higher than the load factor for small consumers. On the other hand if the load factor is constant across usage or even decreases with usage, then it is highly inequitable to charge higher prices for distribution equipment for consumers with low usage levels. In computing load factor it is important to use the coincident peak or the class peak and not the peaks of individual consumers.

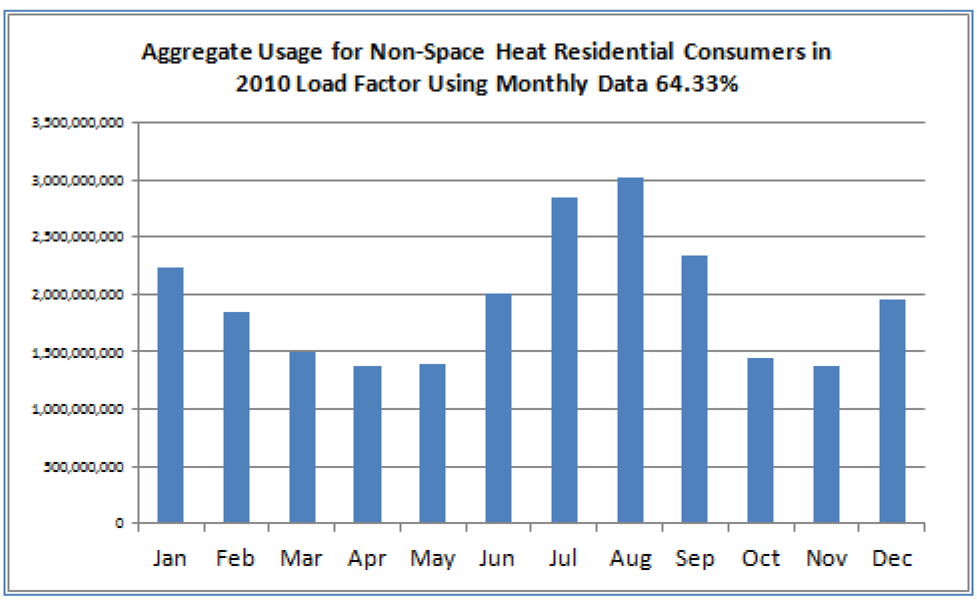
4. Can randomness in the relationship between peak load and usage that arises from home vacancy when moving or from vacations correctly be attributed to a customer acconunt rather than usage?

Part 1: Analysis using Comprehensive 2010 Data for Non-Space Accounts

ComEd has provided monthly usage data for virtually all of the individual ratepayers for the year 2010. This data can be used to test the relationship between usage during the peak month and usage over the course of a year. In analyzing the data only the non-space heat consumers have been selected as space heating involves a fundamental difference in the manner in which electricity is used. While the monthly data for non-space heat consumers provides a general indication of peak usage relative to average usage, the monthly usage data cannot be used to evaluate the relationship between energy usage in the peak hour and average energy use over the year. The analysis of average usage relative to peak hourly usage is presented below in part 3 below.

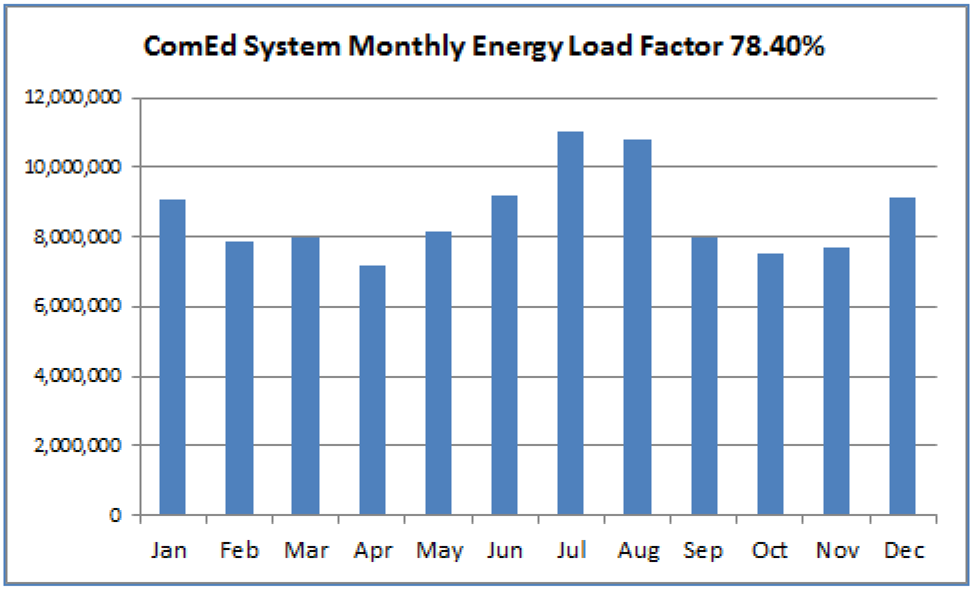
In 2010 the maximum aggregate usage across the non-space heat residential class occurred in the month of August as shown in the graph below. Relative to peak usage in August, average use across the year was 64% of the usage in the highest month. The 64% statistic can be termed a load factor. (Any load factor is defined as some level of average use over an extended period divided by some definition of maximum use during a shorter period.) In the context of the monthly usage data, the load factor can be defined as:

Load Factor from Monthly Usage Data = Average Monthly Use/August Monthly Use



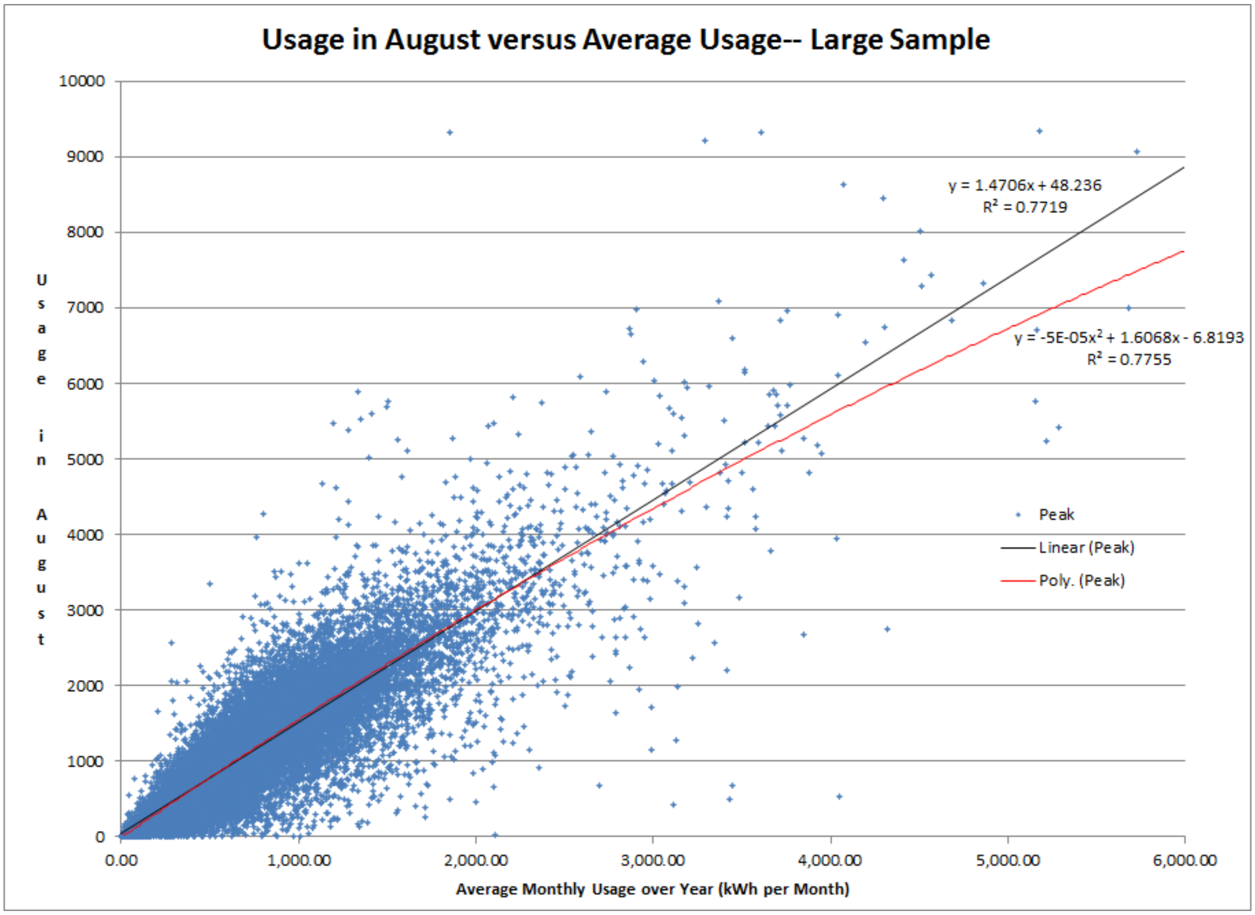
The aggregate residential usage data can be compared to energy usage on the entire ComEd system that includes all non-residential as well as residential consumers. For the entire ComEd system in 2010 the load factor was 78.4% and the month with the maximum peak occurred in

July (hourly data from the PJM website was used for this analysis). For the entire system, the peak load occurred on August 12 at 5:00 PM. Given that the peak usage month was August for residential consumers and because the peak hour occurred in August, the peak month for analysis below is defined to be August.

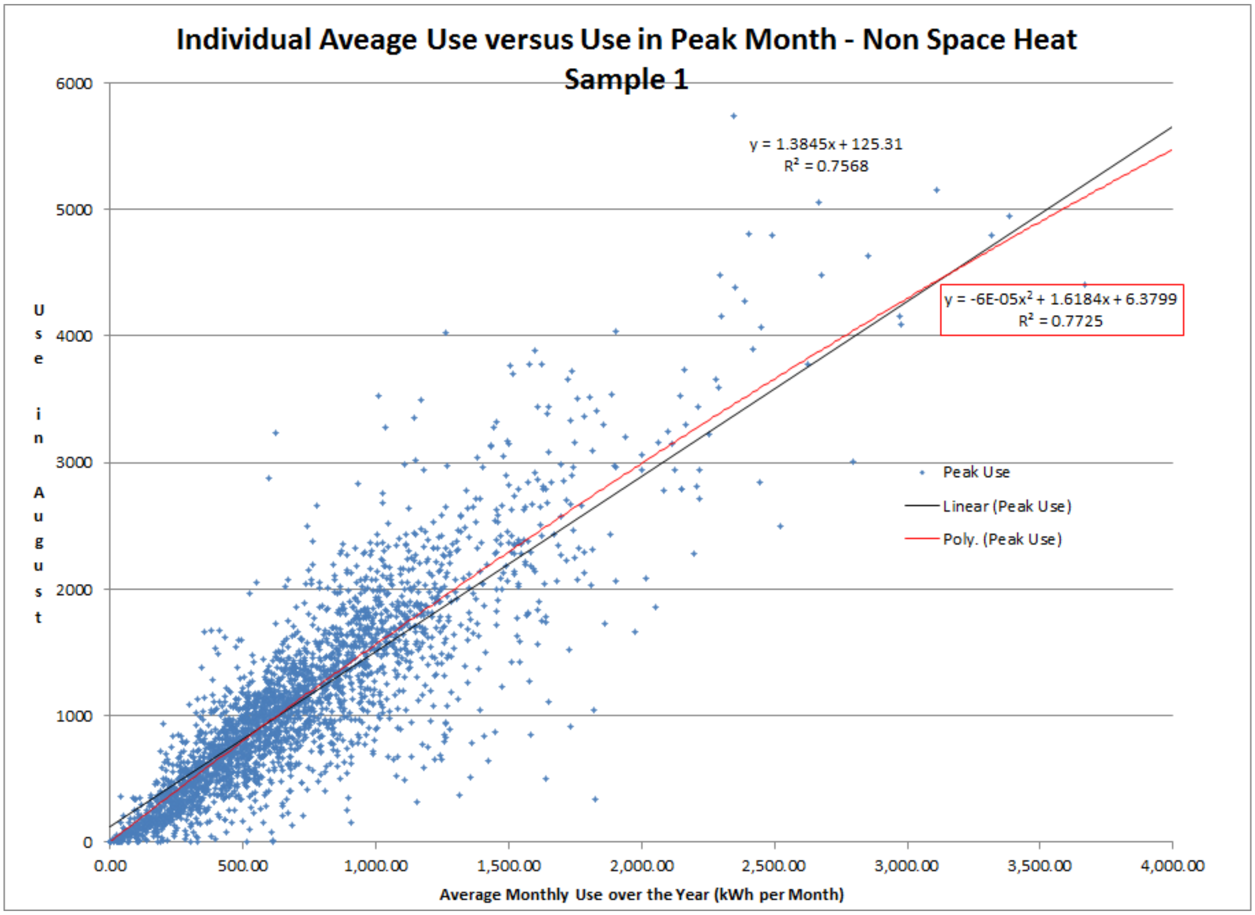


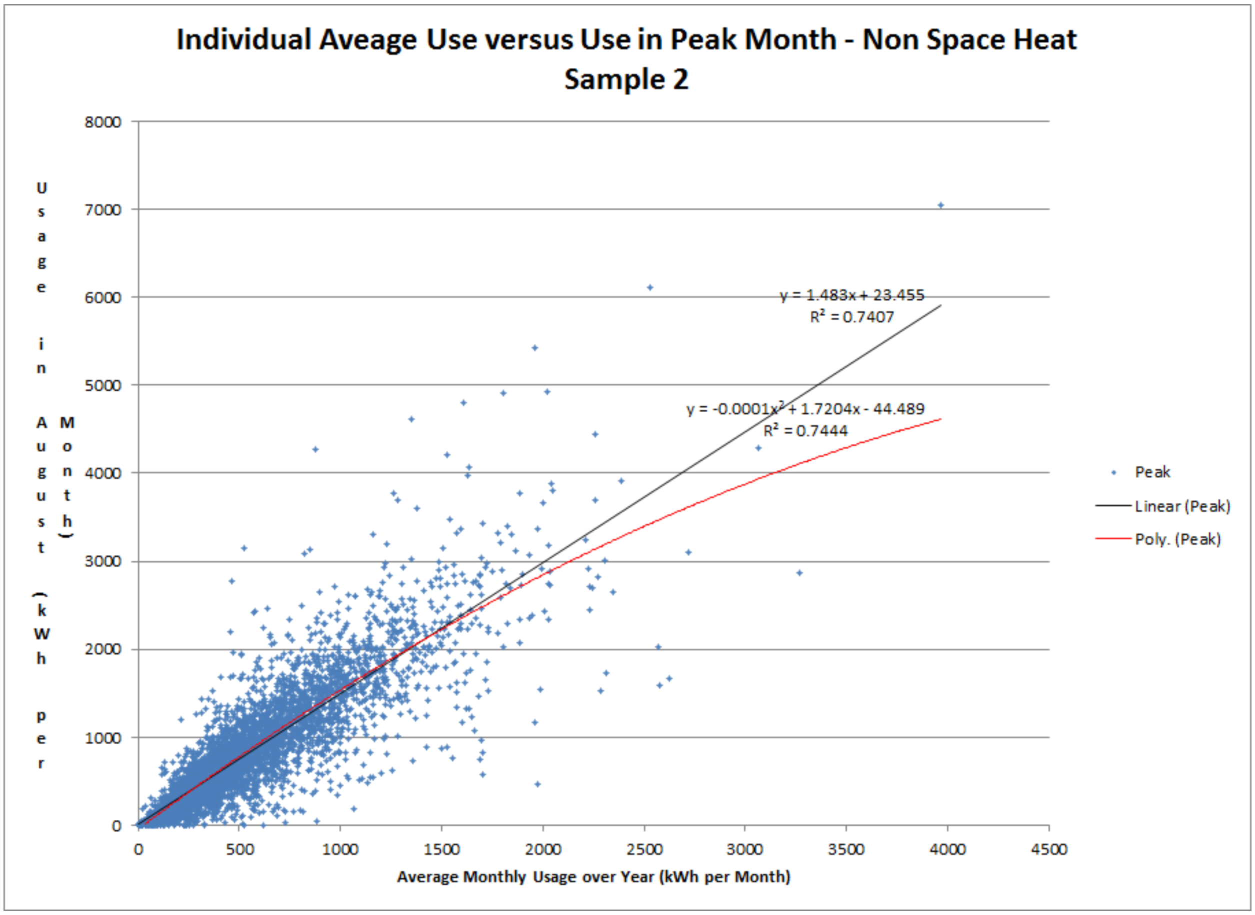
Average Annual Use and August Monthly Use for Individual Ratepayers

The relationship between average customer use over the whole year and use during the month of August provides a general picture of the correlation between usage and demand. To evaluate this relationship I have constructed scatter plots of the average use and the peak use for various samples of consumers using the 2010 usage data (I was unable to make a graph of every consumer because of the volume of the data and the graphs with very large density do not show overlapping points). In presenting the scatter plots, I have included lines that fit the data as well as the R-squared which measures the percent of variance in the peak load that can be explained by variance in usage.



Usage in August versus Average Usage-- Large Sample
Usage in August
Average Monthly Usage over Year (kWh per Month)
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
0.00
1,000.00
2,000.00
3,000.00
4,000.00
5,000.00
6,000.00
y = 1.4706x + 48.236
R² = 0.7719
y = -5E-05x² + 1.6068x - 6.8193
R² = 0.7755
Peak
Linear (Peak)
Poly. (Peak)


Individual Aveage Use versus Use in Peak Month - Non Space Heat
Sample 1
y = 1.3845x + 125.31
R² = 0.7568
y = -6E-05x² + 1.6184x + 6.3799
R² = 0.7725
Peak Use
Linear (Peak Use)
Poly. (Peak Use)
Use in August
Average Monthly Use over the Year (kWh per Month)
0
1000
2000
3000
4000
5000
6000
0.00
500.00
1,000.00
1,500.00
2,000.00
2,500.00
3,000.00
3,500.00
4,000.00



This scatter plot data demonstrates the following:

1. The correlation coefficient between average use and maximum use is about 87% (the square root of the 77% R-squared). This suggests that about 13% of the standard deviation in peak load for individual ratepayers cannot be explained directly by an equation where each consumer is assumed to have the same relationship between peak and average load.

   The correlation of less than 100% may come about because different ratepayers have different use/peak relationships. For example, consumer A may have an equation where Peak = 1.5 x Average Use while consumer B has a relationship that is characterized by Peak = 2.5 x Average Use. I term this the consumer peak/use relationship in the discussion below. To the extent that consumers have different peak/use relationships they appear as dots away from the fitted lines in the graph. Further, if an apartment is vacant during the summer months, then the average use may be high but the peak use would be almost nothing. Other factors that explain why the correlation is not 100% include situations such as when a family is away on vacation for much of August. In this case the peak load would seem to be low relative to the average use and the dot would be below the trend line. The important point in the context of the propositions tested is whether the correlation of below 100% implies that some of the peak demand is caused by the simple existence of a customer account. Randomness caused by different peak/use relationships or vacations does imply that variation on the graph is caused by consumer accounts.

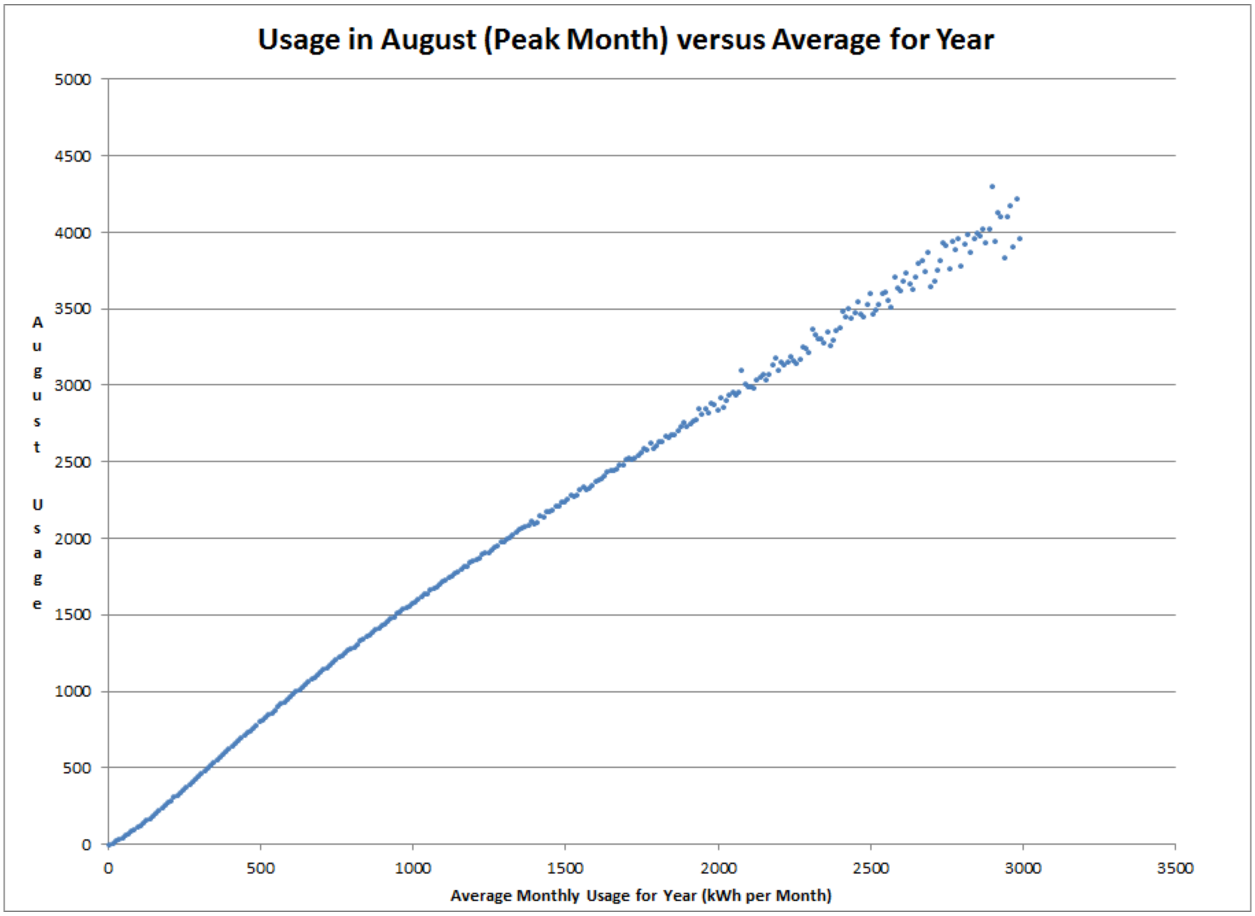
2. Given statements in the ComEd Residential Usage Study that consumers who have low use may suddenly have high use during peak periods (for example in the part of the report that discusses vacation homes), one would expect many dots to occur where average usage is low but peak usage is high. These dots would appear in the upper left hand quadrant of the graph. The graphs above demonstrate there are not many points in these areas and demonstrate that ComEd's antidotal stories have virtually no impact whatsoever in the context of the entire sample.

3. If part of the peak load was due to the presence of a customer account and not related to average use, then one would expect that intercept on the graphs (where the trend line crosses the y axis) to be clearly positive. In the extreme case where peak use is unrelated to average use, the intercept would be the average value of the peak use. To illustrate this point consider an example where a house is divided from one ratepayer into two ratepayers with the average use cut in half. Pretend that the house before being divided was a dot on the graph as well as the two accounts from the divided house. If peak usage was not related to use then the three points (the non divided house and the two parts of the divided house) would all have the same peak monthly usage. However the divided house would have half of the average use. This implies there would be three points along a straight line at the level of the peak load. The graphs above look nothing like what would occur if average usage were entirely unrelated to peak usage. As explained below, because load factors tend to be better for low use consumers, the relationship between peak use and average use is best represented by a polynomial equation. The graphs above show that when a polynomial equation is used, the intercept is approximately zero (it may not be exactly zero because of mathematical specifications of the relationship between peak and average use.) As explained in the simulation analysis below, an intercept of zero implies that none of the variation in peak load can be explained by the simple existence of a customer account.

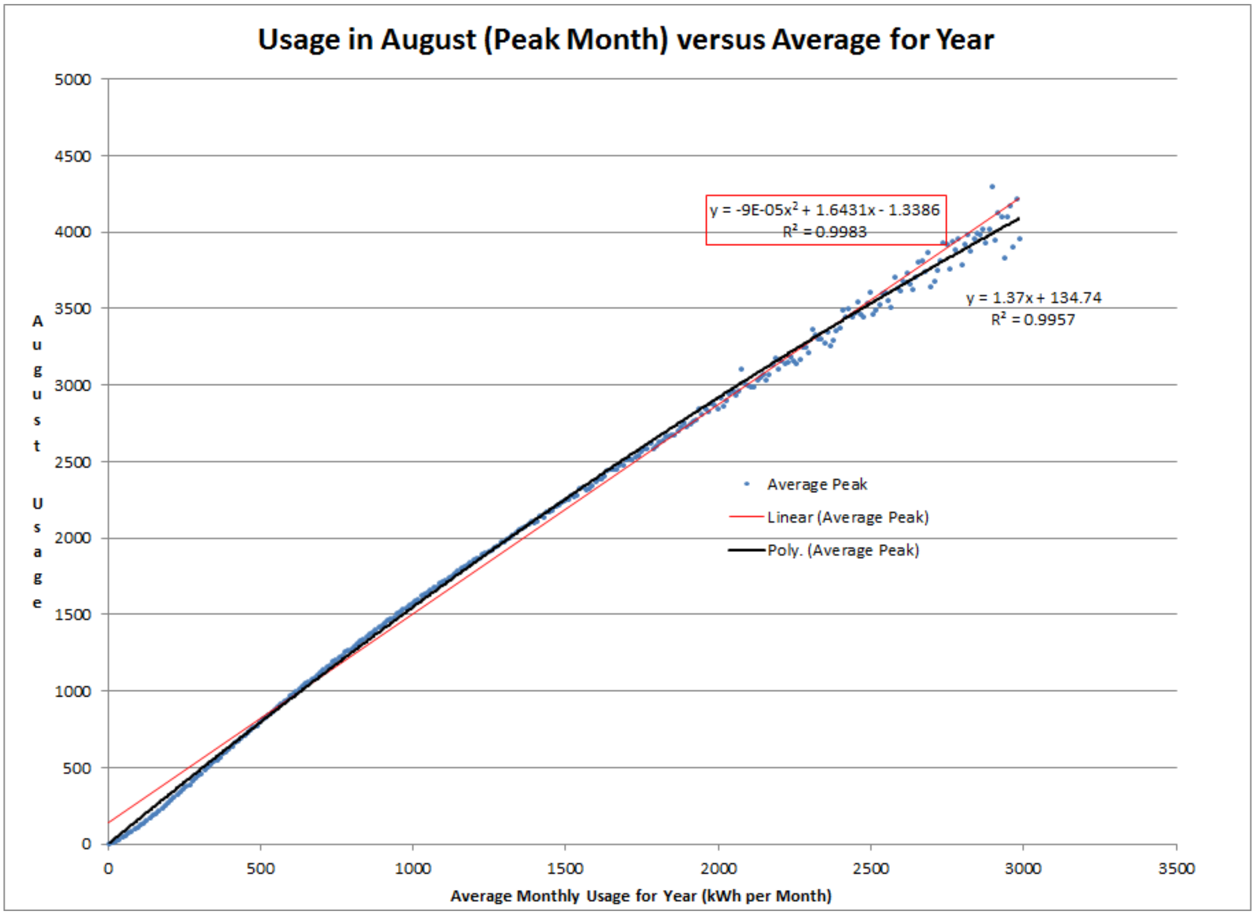
Average Annual Use and August Monthly Use for Ratepayers Aggregated by Usage Group

To further test the relationship between average usage and peak month usage, the average usage data for each consumer was categorized into 10 kWh per month increments. For each 10 kWh per month increment the sum of the average annual use and the sum of the August monthly use was computed. After summing the data, the average annual use and average use for the peak month were calculated. For example in the monthly usage increment between 490 and 500 kWh per month there were 27,000 accounts and the average usage was 485 kWh per month while the peak was 782 kWh per month. The process of averaging use by small increments corrects for situations such as the vacant apartment or the vacation scenarios discussed above. For example if one apartment is vacant in August and another in January and a third in March, then after the data is aggregated, the distortions in individual accounts are eliminated. One can then evaluate whether such vacancies cause low use categories to have a different use/peak relationship relative to high use categories. Similarly if different ratepayers have different peak/usage relationships but all of the consumers have a positive usage/peak relationship, the differences tend to be averaged out. The remaining variation in peak use that is not related to average use can be attributed to random variation associated with the simple existence of an account.

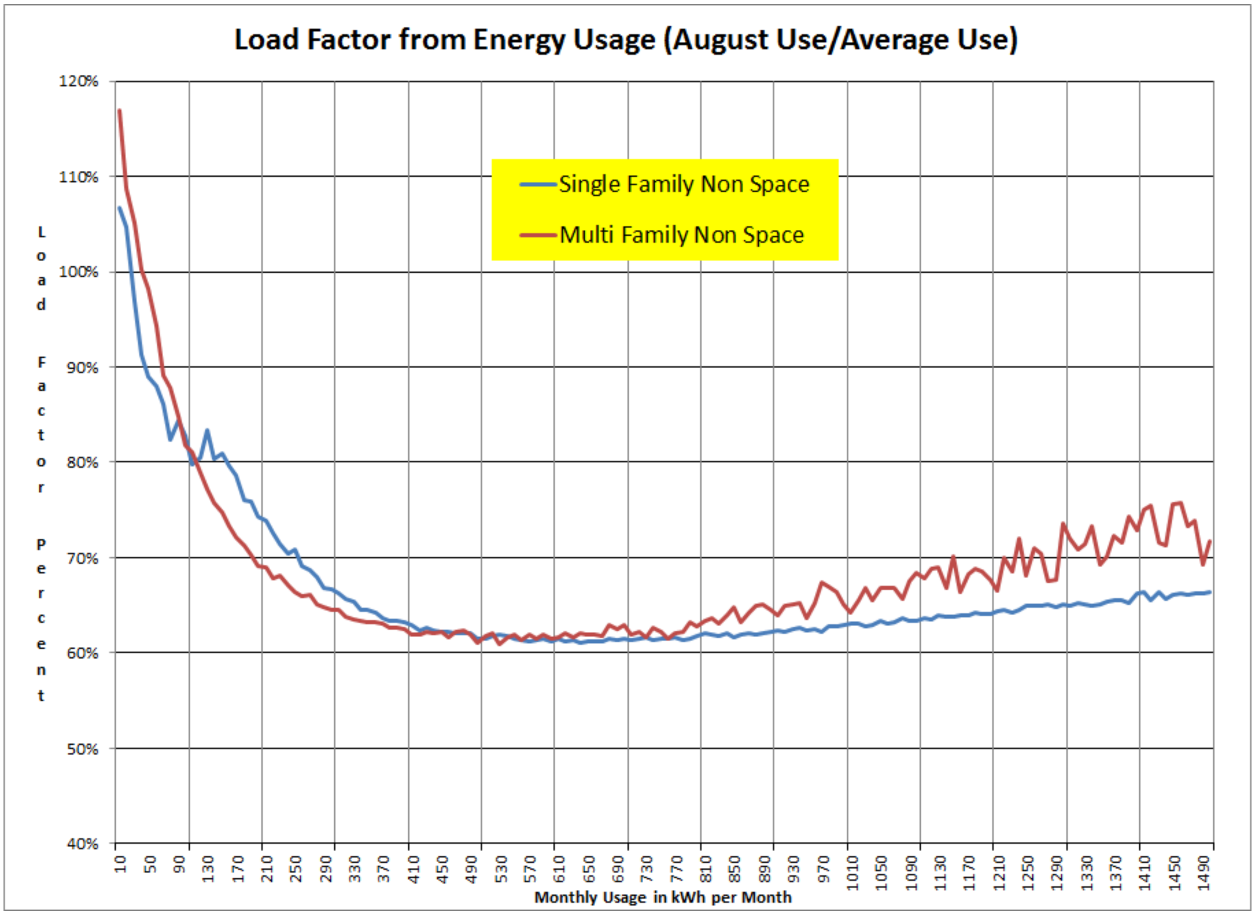
The first chart below shows a scatter plot of the average annual usage and the peak usage for the various 10 kWh usage increments. Note that after aggregating individual consumers into small usage groups virtually all of the randomness goes away. This implies that different consumers have somewhat different usage/peak relationships but over multiple consumers these differences average out. More importantly, the fact that the graph crosses they-axis at zero demonstrates that there is no variation in peak demand that can be explained by the simple presence of a customer account. Reasons for the differences in scatter plots for individual customer accounts relative to the scatter plots with aggregated accounts are explained in the simulation analysis below.



When fitting a line to the data above, it is apparent that the relationship between average usage and usage during the peak period does not follow a simple straight line. For the low usage increments (below 500 kWh per month) there is relatively less peak usage than for the higher usage increments. For extremely high use increments (above 2,500 kWh per month) this the peak to the usage declines. The graph below shows that when a polynomial equation is fitted to the data the correlation as measured by r-squared is 99.83%. The high correlation combined with the fact that the line crosses they-axis at zero demonstrates that there is effectively no peak use that can be explained by the existence of a ratepayer account.



The fact that the relationship between peak monthly use and average use does not follow a straight line can be translated into different load factors. The graph below shows that load factors are higher for low use categories until the monthly usage reaches 500kWh per month. Then, after falling and hitting a plateau, for usage levels of more than 900 kWh per month the load factor begins to increase. The pattern of load factor can be explained by low use consumers being very careful with electricity and using limited electricity for air conditioning. On the other hand the somewhat improved load factor for very high use consumers can be explained by very high use of appliances other than air conditioners over the course of the year. To the extent that load factors are higher for low use consumers, this graph directly contradicts the principle of setting the price of distribution on the basis of fixed customer charges. Because the customer charges impose higher prices on low use consumers, ComEd's pricing policy could only be justified if the load factor would increase as usage level increases.



Part 2: Monte Carlo Simulation that Demonstrates Why Use of Consumer Groupings is Appropriate in the Analysis

A central question in the above analysis is the statistical issue of whether load and usage analysis should be evaluated by the scatter plots for individual consumers or alternatively whether it is better to draw conclusions from the analysis where usage groups are tabulated. This and other questions can be answered through constructing a simulation model of consumer behavior. The simulation model of consumer behavior is particularly useful in this context because one can directly test the ComEd hypothesis that peak demand is not related to usage. Further one can test the more reasonable question as to whether some of the variation in peak demand is related to the mere existence of a ratepayer account and some is related to usage.

The simulation model of consumer behavior creates an equation that incorporates random variation due to alternative factors such as the presence of a ratepayer account or different demand/usage relationships. Then it presents the statistical results – scatter plots and regression equations that result creating structured random variables from the postulated equations. If the simulated results are completely different from the actual analysis shown above, then one can reject the hypothesis that the underlying equation that modeled the data is correct. On the other hand if the modeled data generates analogous individual and grouped graphs as shown in the above analysis, this confirms the underlying structure of the assumed model in the simulation.

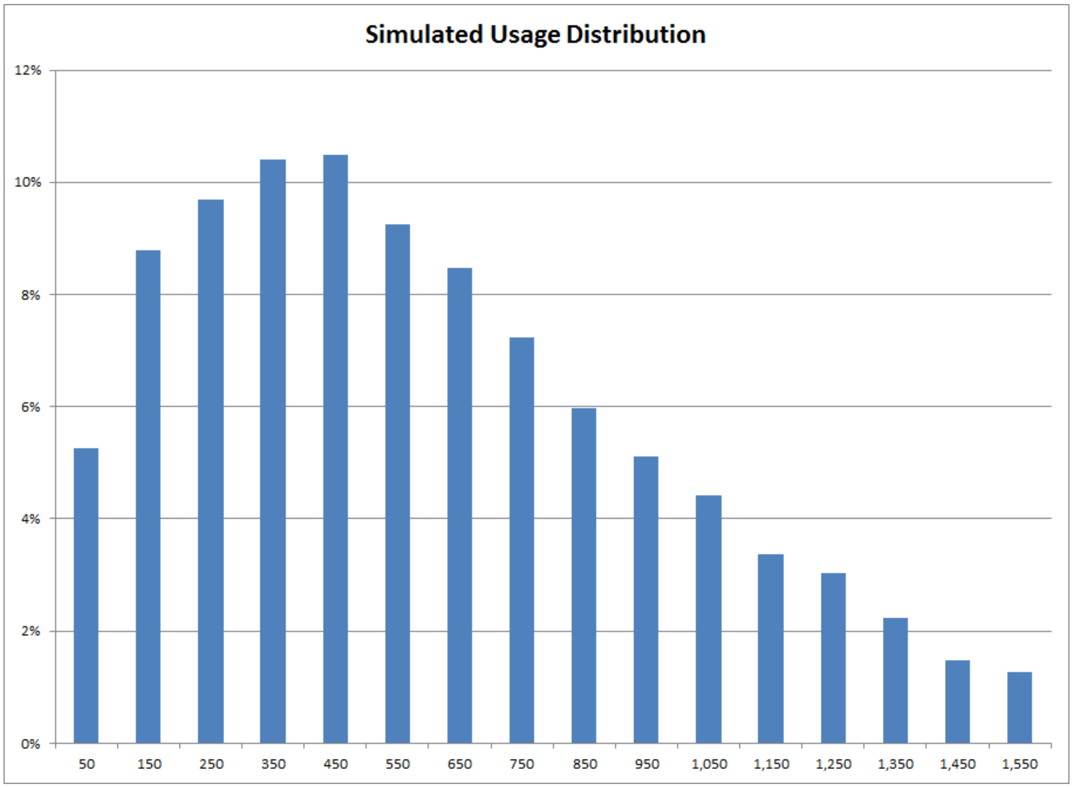
The simulation process can be described by the following three step process:

Step 1: Construct an equation of electricity demand that allows for random variation that can be created from deviations in demand/usage relationships, presence of ratepayer accounts, vacancies during on and off peak periods and so forth.

Step 2: Simulate randomness in the demand resulting from the equation through performing multiple random draws and filtering the random draws through a normal distribution or a Wiebull distribution.

Step 3: Perform statistical analysis on the simulated outcomes in the same manner as the analysis above and evaluate whether the output (scatter plots) are consistent with the actual data.

For each consumer behavior model, average usage is simulated using a Wiebull distribution. The Wiebull distribution yields a probably distribution that can be skewed to the right. This probability distribution can be used together with random draws to simulate thousands of different values for consumer use. An example of usage distribution resulting from this process is shown in the graph below.



In using simulation, the following alternative models of consumer behavior were tested that assume (1) usage is independent of demand (ComEd's hypothesis); (2) some of the variation in demand comes from ratepayer by ratepayer demand/usage variation but some demand variation is independent of usage and implicitly driven by the simple presence of a consumer account (a more reasonable version of ComEd's hypothesis); (3) demand is driven only by usage and includes random variation across individual consumers in the demand/usage relationship; and (4) variation in demand comes from both demand/usage variation as well as variation caused by consumers not using demand at the peak and consumers not using demand during off peak period periods. The last case is intended to represent vacation homes and vacancies in the occupancy of apartment buildings.

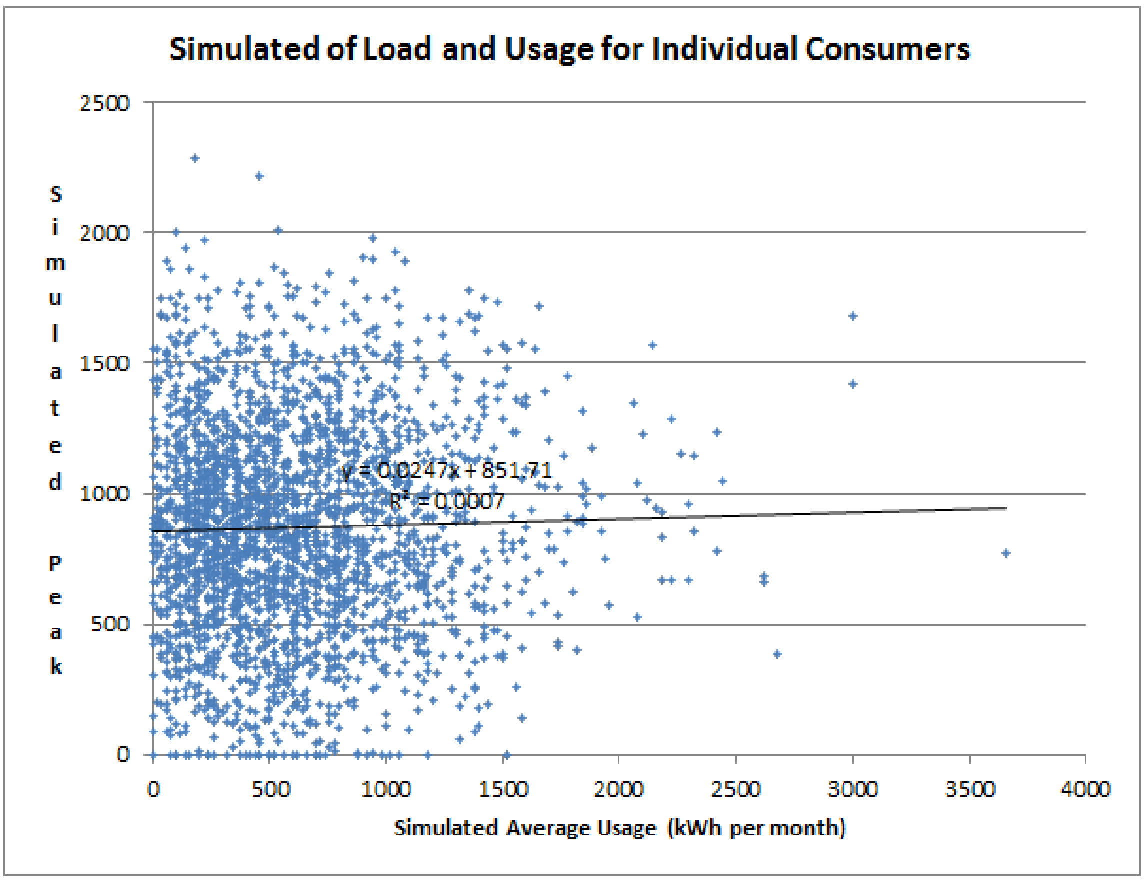
Case 1: No Relationship between Demand and Usage

To create a model where there is no relationship between demand and usage the following equation structure can be used:

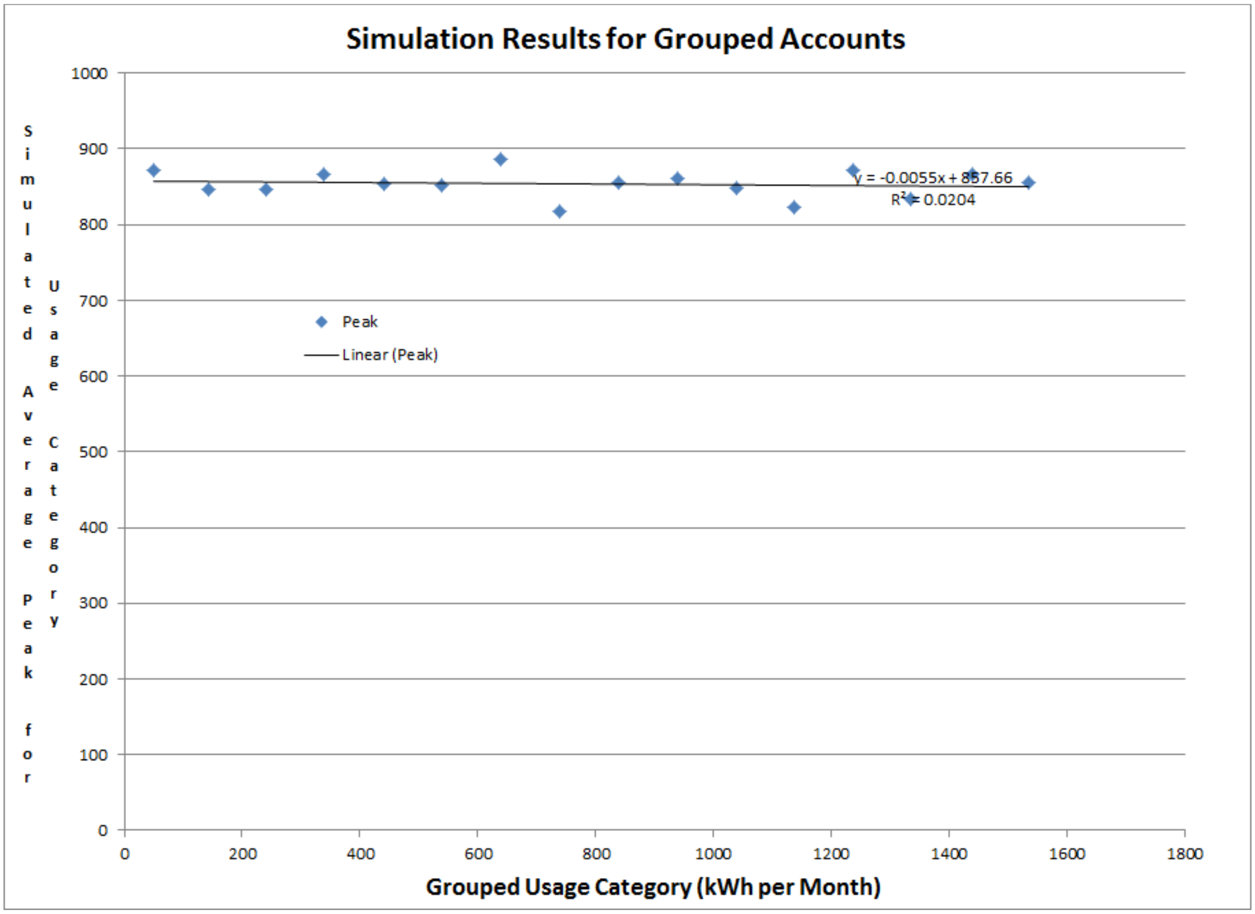
Demand = Constant Demand x Random Factor + Usage x 0

Constant demand is assumed to be 850 kWh per month with a standard deviation of 400 kWh per month.

In this case the usage has no influence on the demand because the usage coefficient is zero, but there is variation in demand that is just random. The random variation is modeled with a normal distribution that yields a demand level that can be above or below the assumed constant of 850 kWh per month. Usage over the course of the year has nothing at all to do with the level of the peak demand as in ComEd's repeated vacation home antidotes. The results of this case in terms of individual scatter plots (shown below) look nothing at all like the actual data that was presented in the previous part of the report. For the individual accounts, there is no correlation and nothing close to the notion of a zero intercept. Instead, the intercept is simply the average level of demand.



For the scatter plot of grouped usage, there is no positive correlation and the intercept is again the level of the average demand. The grouping of accounts does remove the random variation associated with different demands for individual consumers. As with the scatter plot for individual consumers, the graph looks nothing at all like the scatter plot generated from actual data. The simulation demonstrates that any supposition that average usage is not related to peak usage can be clearly rejected using the 2010 database. The simulation confirms that a suggestion that usage is not related to demand is simply absurd.

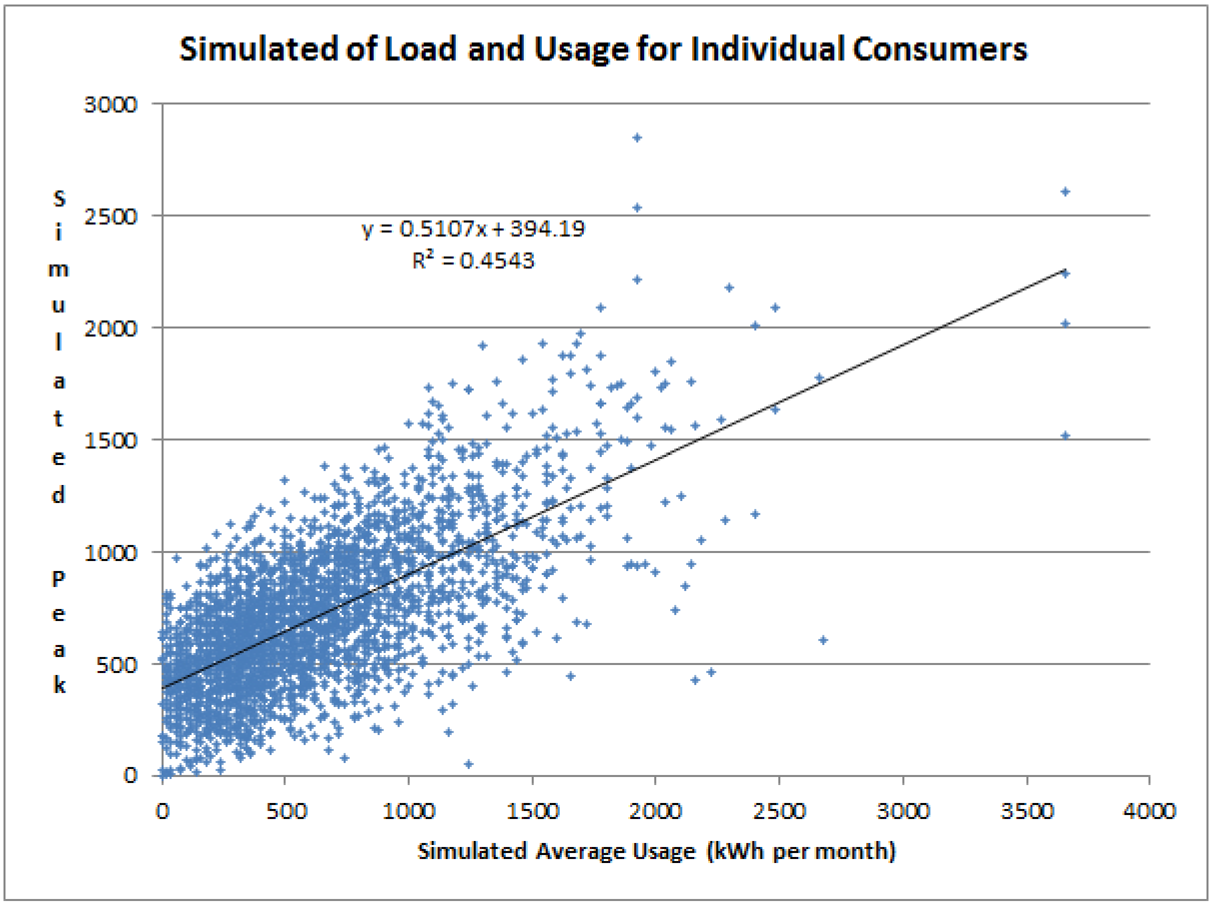


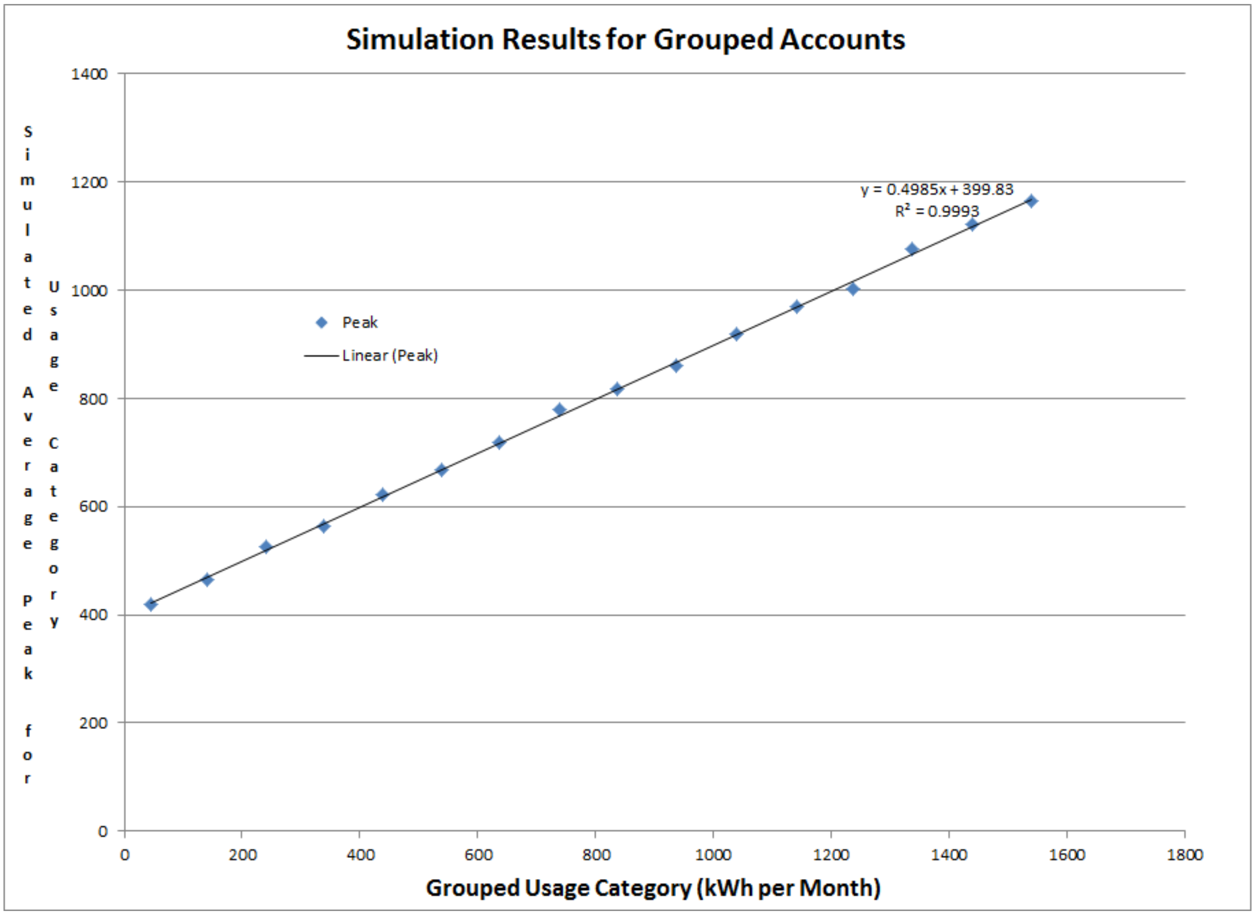
Case 2: Demand Variation from Both Ratepayer Accounts and Usage

In the second case some of the demand is a function of energy usage over the year and some is fixed and independent of usage as above. Both the fixed level and the relationship with demand are random variables meaning that each consumer can have a different equation. In the second case the relationship with demand can be written as:

Demand = Constant Demand x Random Factor + Usage x (Usage Coefficient x Random Factor)

In this equation the Constant demand is assumed to be 400 and the usage coefficient is assumed to be .5. Both the usage and the demand have random variation meaning that some consumers may have a coefficient of .3 and others of .7. The variation is driven by the assumed standard deviation of .2 in the usage coefficient and the standard deviation of 200 kWh per month around the constant. The graph below shows that the individual scatter plot looks more like the actual data except that the correlation is only 45% and the intercept is not close to the value of zero and the r-squared of 45% is less than the r-squared of about 77% for actual sample.



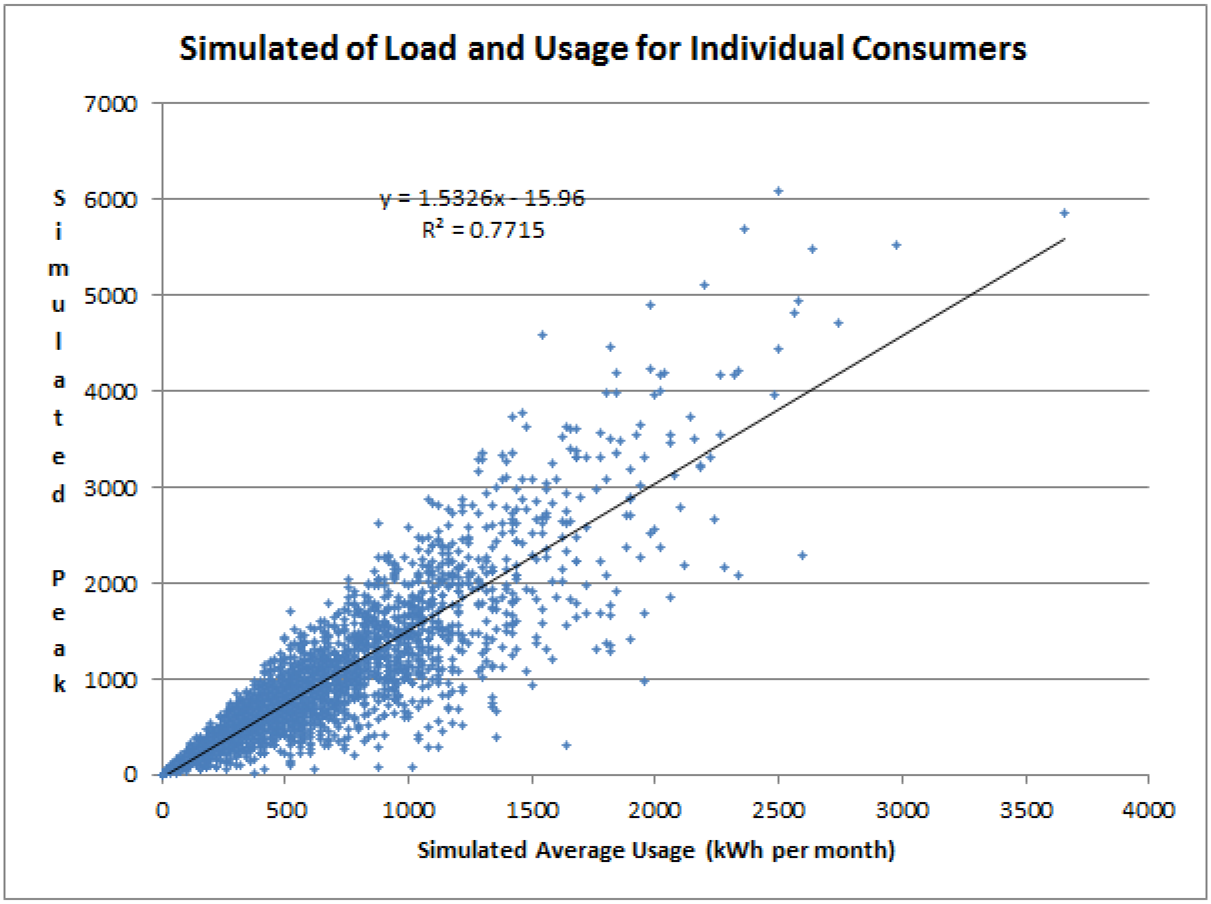


Case 3: Demand Variation from only Usage

In the third case there is no constant term and the only variation in individual consumer demand is derived from: (1) the usage level and (2) random variation in the demand/use relationship across individual consumers. The equation for case three is:

Demand = Usage x (Usage Coefficient x Random Factor)

When randomness only comes from differences in the peak/use relationship, the simulated consumer behavior closely resembles the actual scatter plots presented in part 1. The first graph below shows that for individual simulated consumers there is a lot of variation around the fitted line. This graph shows that even though all demand variation is driven by usage, the R-squared is well below 100% as because of randomness in the demand/use relationship across individuals. In this graph the intercept term is very close to the origin because when the usage level is zero, so is the demand.



For the simulation model where demand is not driven by random behavior unrelated to usage, the scatter plots with grouped usage and demand also resemble the actual data. The graph below shows that after the data is grouped, the intercept term is close to zero and the R-squared is close to 100% as was the case for the case for the actual data. The simulation of this case with no variation derived from consumer accounts confirms that none of the demand comes from the existence of a ratepayer account and ComEd's theory that distribution costs should be priced on the basis of the number of accounts is not valid.