Modelling Marginal System Costs for the Commonwealth Edison Distribution Network

Midwest Big Energy Data

08/17/2018

Kim Sha - ksha@princeton.edu

Ben (Jianda) Liu - jl3726@cornell.edu

Table of Contents

Abstract

The installation of Advanced Metering Infrastructure (AMI) across the Commonwealth Edison distribution system provides renewed potential for big data analytics to characterize the interaction between consumer loads and the utility's system-wide costs. In this paper, we investigate different consumers' marginal system cost impacts as a result of their load profiles.

In the first portion of the study, we outline the process of customer profiling and segmentation, leveraging the k-means clustering algorithm to group Commonwealth Edison customers into clusters across different billing periods, seasons and customer delivery classes. Two pricing schemes were developed and applied to clustered load profiles. A marginal rate pricing model was built to exclusively reflect the impact of system costs on customers. For comparison, a flat rate was determined using historical price-to-compare data.

Analysis of the resulting cost profiles reveals not only the role of load magnitudes, but also the significant impact of load shapes on marginal system costs. Regression tests that were conducted indicate that the time deviation between a cluster's peak load and the system-wide peak load plays an important role in dictating the cost to consumers. In particular, time deviation was shown to be the dominant driver of system costs during summer months when the system costs are the highest.

Introduction

When constructing pricing schemes or policy recommendations in the context of utility billing, one must have a firm grasp of the marginal impact that consumers have on underlying system grid costs. Constructing an idealized model from this baseline would price electricity according to core system costs associated with serving a customer on a real-time basis. However, while traditional Residential Real Time Pricing (RRTP) schemes can serve as close proxies, they fail to isolate these core system costs from other overhead administrative costs baked in by the utility provider. Thus, it's worth constructing an idealized pricing model based on the perfect knowledge of historical system costs on a real-time basis for a particular provider: Commonwealth Edison.

Developing this pricing scheme is insightful, but to truly investigate the marginal impact of different consumers, one must apply it to customer load data to yield cost profiles reflecting their relative system impacts. However, conducting such an analysis on the entire customer pool is impractical. With the widespread adoption of smart meters among ComEd customers in Illinois, the wealth of data available for analysis is quickly reaching unwieldy proportions.

This introduces the challenge of drawing broad conclusions from customer behaviors that are inherently distinct. It's not hard to understand that consumers have different preferences and consume energy in different ways, but this makes it extremely challenging to formulate comprehensive recommendations that are relevant to every customer. Indeed, different types of electricity consumers are more responsive to recommendations that are catered to their particular consumption behavior. The ability to draw distinctions between different types of consumers is what makes customer segmentation and profiling valuable. With literally millions of customers' half-hourly consumption data, big data techniques are required to analyze electricity usage on a macroscopic level.

Customer Segmentation and Profiling

For the purposes of this big energy data analysis, the raw anonymized customer data is pulled from Commonwealth Edison's Advanced Metering Infrastructure (AMI) readings. The data was first grouped and cleaned before segmented using the k-means clustering algorithm.

Data Grouping and Cleaning

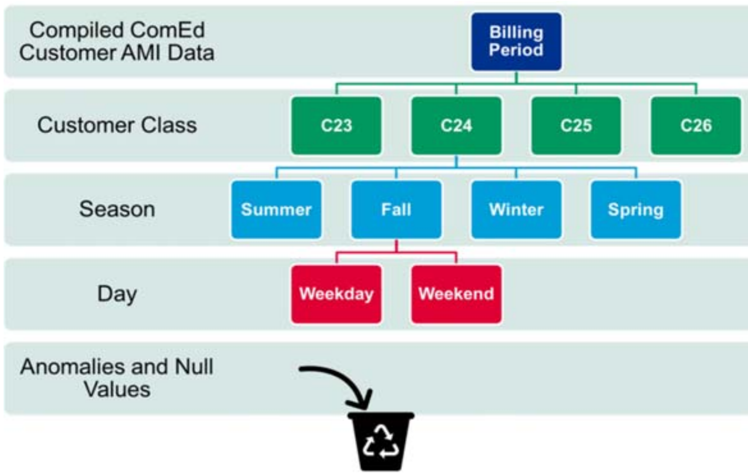
The raw customer load data (in kWh) contains indices that mark the date associated with the measured data, as well as the customer delivery class. The four residential classes are as follows:

- C23 – Residential single family without electric space heat
- C24 – Residential multi-family without electric space heat (three or more customers served through separate meters from a single service connection e.g. an apartment building)
- C25 – Residential single family with electric space heat
- C26 – Residential multi-family with electric space heat

Because these categories were readily available attached to the raw data, it was taken advantage of in order to pre-segment the data by billing period (June 1st of any given year to May 31st of the following year), delivery class, season, and weekday or weekend. Note that the data for each individual consumer was averaged and scaled by their average total energy usage. This meant the algorithm would segment customers into clusters based exclusively on similarities in the *shape* of their load profile, as opposed to accounting for the impacts of magnitude.

Before implementing k-means clustering, anomalies and null values that would otherwise interfere with the algorithm were dropped from the data. Specifically, customers consuming on average less than 1 kWh a day were considered outliers and discarded in this analysis. In addition, any rows containing N/A values were dropped. Any other extreme outliers were removed on a case-by-case basis.

Figure 1 – Visual representation of the data segmentation process prior to implementing K-Means Clustering.



K-Means Clustering

Now that our customer data has been averaged, scaled and grouped, we can implement k-means clustering to profile customers. Since no explicit labels to describe the data are provided for the algorithm to learn from, this is a form of unsupervised machine learning. The similarities between customers are measured based on 48 attributes (load usage at every 30min. interval over 24 hours) at the same time, using a Euclidean distance metric. An outline of our implementation of the algorithm is as follows:

1. Determine the number of clusters *k* to output using the elbow method and silhouette scores. Since the elbow method can be subjective, silhouette scores were designated as deciding factors in close cases.
2. Randomly select *k* customers to form the starting centers for the clusters (i.e. the centroids).
3. Data assignment – sort each observation (customer) into clusters based on its Euclidean distance to the nearest centroid:

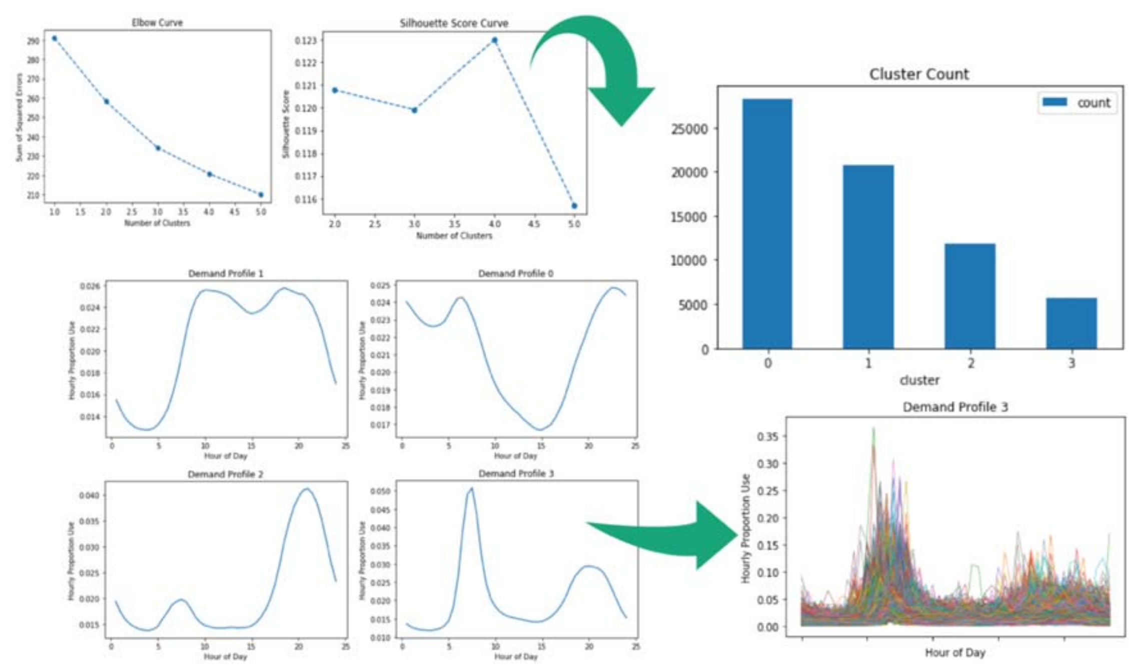
Where

4. Centroid adjustment – adjust the positions of centroids by averaging the observations assigned to that cluster.
5. Iterative refinement – repeat steps 3-4 until cluster assignments remain unchanged and the intra-cluster sum of squares is minimized:

Where and

The resulting clusters and their respective observations are exported as .csv files and the centroids are plotted for reference.

Figure 2 – A representation of the K-Means clustering process (for C26 weekdays in spring 2016). The input number of centroids (K = 4) is first determined via examination of the elbow curve or silhouette score. The algorithm then segments the data into (K = 4) clusters. The right-hand-side plots show the distribution of customers across the clusters and a visual compilation of one such cluster.



Modelling Price Schemes

In this study, we examine two different pricing schemes: a marginal real-time cost model and a flat-rate model. The flat-rate scheme reflects actual prices charged to consumers on a flat-rate plan. The marginal cost scheme is a pricing framework designed to reflect real-time system grid costs.

Different price components introduced hereafter may vary based on:

Marginal Cost Pricing Model

There are four components that make up the design of the marginal real-time cost pricing rate (hereafter referred to as "Marginal Rate"):

1. *Delivery charges* – common charges indicating the price of delivering electric supply through the local distribution grid. They are calculated from ComEd's Embedded Cost of Service Study (ECOSS)[1].
2. *Capacity Charge* – rates that account for the procurement of resources required by the regional transmission grid (i.e. PJM), in particular for grid emergencies. In practice, electricity suppliers must pledge reserve resources in addition to meeting its customers' demand. These are determined through PJM Residual Auction data for ComEd[2] and divided by delivery class based on capacity peak load contributions (PLC)[3].
3. *Transmission charges* – charges used by utility providers (i.e. ComEd) to offset costs accrued through electricity transmission services from generators to local distribution centers via high voltage power lines. These are determined from

[1] Docket No. 17-0196, ComEd Ex. 7.01, 2017 FRU ECOSS

[2] http://www.pjm.com/markets-and-operations/rpm.aspx

[3] https://www.comed.com/SiteCollectionDocuments/PLCs_by_Delivery_Class.pdf

ComEd's Annual Transmission Revenue Requirement[4] and divided by delivery class based on network PLCs[2].

4. Energy charges – market rates retrieved from Locational Marginal Pricing data (LMPs)[5]. As a reflection of electric prices at every location in the distribution grid, LMPs represent the system cost components of electric supply charges.

Three of these components (delivery charges, capacity charges and transmission charges) vary between different yearly billing periods, seasons and customer classes, but remain constant on an hourly basis. Energy charges vary by billing period, season and hour, but not by class.

The half-hourly marginal rate in ¢/kWh,, is designed as:

Because D, C and T are "fixed" throughout the 24 hours of the day, we group them in a defined fixed rate:

More specifically, the daily marginal rate in ¢/kWh, , is a 48 by 1 matrix:

More detailed derivations of the individual rate components can be found in Appendix I. Reflecting these components, the finalized marginal cost pricing model yields a unique price for each billing period, season, customer class and hourly interval.

Flat-Rate Pricing Model

There are three components that make up the flat-rate pricing model:

1. *Historical Price-to-compare* – a summation of ComEd's electric supply charge (or energy charge) and transmission services charge (which encompasses capacity and transmission charges)[6].

[4] https://www.pjm.com/markets-and-operations/billing-settlements-and-credit/formula-rates.aspx

[5] https://www.pjm.com/markets-and-operations/energy/real-time/lmp.aspx

[6] https://www.pluginillinois.org/FixedRateBreakdownComEd.aspx

2. *Purchased electricity adjustment (PEA)* – a price adjustment accounting for the difference between ComEd price-to-compare revenues and the actual cost of supplying electricity to customers.
3. *Delivery Charges* – see marginal cost pricing model.

Both price-to-compare and PEA varies by year and season, but not by customer class or an hourly basis. The design of delivery charges has been adjusted to also vary exclusively by year and season (i.e. a "flat" delivery charge):

The flat rate in ¢/kWh, , is therefore defined as:

Rate Variation Summary

Table 1 – A representation of the variability of each pricing scheme. A ✓ indicates that the rate of the corresponding row index varies with the corresponding column index.

	Year	*Season*	*Class*	*Hour*
Flat Rate	✓	✓		
Marginal Rate	✓	✓	✓	✓

Note that in the case of marginal rates, different customer classes' energy usage patterns and contributions to system costs result in each class being charged very differently. Such different consumption behaviors may result in very different rate assignments, which will be explained in later sections.

Hourly Indifference Ratio and Critical Point

On an hourly basis, is defined as

Preference Implications

If, consumers are indifferent between the two pricing schemes.

If, flat rate will be higher, and consumers who are on flat rates are “over-paying”, effectively subsidizing system costs for others.

If, marginal rates are higher, and consumers on flat rates are “underpaying”, effectively shifting their impact on system costs to others.

The HIR theoretically shows the optimal option for the consumer at every hour. However, consumers can’t switch between different pricing schemes on an hourly basis, and we need to construct our consumer preference implications from a cost perspective.

Modelling Costs

As in modelling price schemes, different cost components introduced hereafter may vary based on:

Mathematically, daily total costs can be defined as:

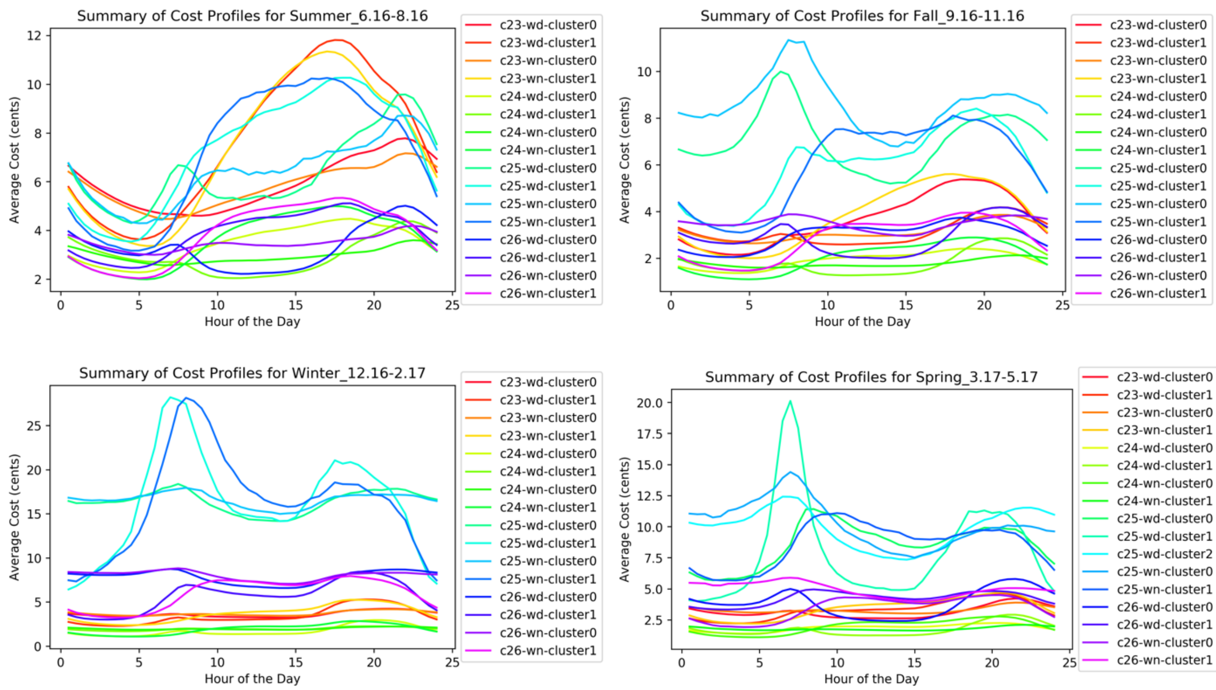
Where daily loads are given in 48 30-minute intervals. These can be expressed mathematically as 1 by 48 matrices:

Flat-Rate Cost

Following the general equation outlined above, we can calculate the total flat-rate cost for a load profile on a daily basis:

Note that the total flat-rate cost can be broken down into hourly costs to yield a flat-rate cost profile for each customer in a cluster.

Figure 3 – Hourly-averaged flat-rate cost profiles for each cluster in the 6/1/16 – 5/31/17 billing period.



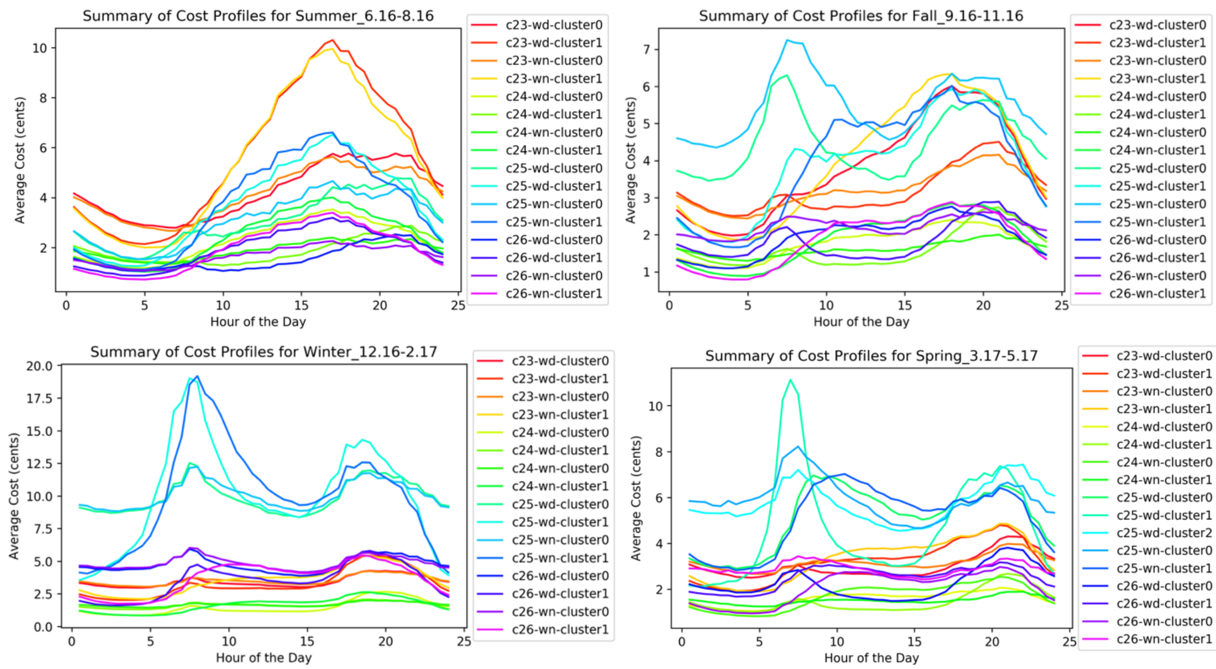
Marginal Rate Cost

We define an analogous relationship for the total marginal rate cost of a load profile on a daily basis:

$$=$$

As with the previous case, the total marginal rate cost can be broken down into hourly costs to yield a marginal rate cost profile for each member of a cluster.

Figure 4 - Hourly-averaged marginal-rate cost profiles for each cluster in the 6/1/16 – 5/31/17 billing period.


Summary of Cost Profiles for Summer_6.16-8.16
Average Cost (cents)
Hour of the Day
c23-wd-cluster0
c23-wd-cluster1
c23-wn-cluster0
c23-wn-cluster1
c24-wd-cluster0
c24-wd-cluster1
c24-wn-cluster0
c24-wn-cluster1
c25-wd-cluster0
c25-wd-cluster1
c25-wn-cluster0
c25-wn-cluster1
c26-wd-cluster0
c26-wd-cluster1
c26-wn-cluster0
c26-wn-cluster1
Summary of Cost Profiles for Fall_9.16-11.16
Average Cost (cents)
Hour of the Day
c23-wd-cluster0
c23-wd-cluster1
c23-wn-cluster0
c23-wn-cluster1
c24-wd-cluster0
c24-wd-cluster1
c24-wn-cluster0
c24-wn-cluster1
c25-wd-cluster0
c25-wd-cluster1
c25-wn-cluster0
c25-wn-cluster1
c26-wd-cluster0
c26-wd-cluster1
c26-wn-cluster0
c26-wn-cluster1
Summary of Cost Profiles for Winter_12.16-2.17
Average Cost (cents)
Hour of the Day
c23-wd-cluster0
c23-wd-cluster1
c23-wn-cluster0
c23-wn-cluster1
c24-wd-cluster0
c24-wd-cluster1
c24-wn-cluster0
c24-wn-cluster1
c25-wd-cluster0
c25-wd-cluster1
c25-wn-cluster0
c25-wn-cluster1
c26-wd-cluster0
c26-wd-cluster1
c26-wn-cluster0
c26-wn-cluster1
Summary of Cost Profiles for Spring_3.17-5.17
Average Cost (cents)
Hour of the Day
c23-wd-cluster0
c23-wd-cluster1
c23-wn-cluster0
c23-wn-cluster1
c24-wd-cluster0
c24-wd-cluster1
c24-wn-cluster0
c24-wn-cluster1
c25-wd-cluster0
c25-wd-cluster1
c25-wd-cluster2
c25-wn-cluster0
c25-wn-cluster1
c26-wd-cluster0
c26-wd-cluster1
c26-wn-cluster0
c26-wn-cluster1

Cost Analysis

Now that cost calculations have been conducted under flat-rate and marginal rate schemes for each cluster, we wish to compare them against one another. While subtracting one cost profile from another would provide a simple measure of the cost deviation between the two pricing schemes, such a metric scales with load magnitude and would result in unfair comparisons between clusters that have differing total loads. To provide a valid metric for comparison purposes, we define a non-dimensional number: the daily indifference ratio.

Daily Indifference Ratio and Critical Point

We define the as

More specifically,

Preference Implications

If, consumers are indifferent between the two pricing schemes.

If, flat rates will be costlier, and consumers who are on flat rates are 'over-paying', effectively subsidizing system costs for others.

If, marginal rates are costlier, and consumers on flat rates are "underpaying", effectively shifting their impact on system costs to others.

Intra-Season Analysis

Within the same season, different classes have different values of DIR. To compare, we can see the averaged DIR for different classes across the two billing periods:

Table 2 – DIR for the 2017/2018 billing period.

		C23	C24	C25	C26
Average Value	Total	1.1184	1.259557143	1.824083333	1.832654545
	Summer	1.24805	1.421575	1.964125	1.9893

	Fall	1.024925	1.175066667	1.7527	1.7349
	Winter	1.082225	1.224275	1.755425	1.749325

Table 3 -DIR for the 2016/2017 billing period:

		C23	C24	C25	C26
Average Value	Total	1.0384375	1.17218125	1.619250588	1.6282625
	Summer	1.224225	1.378375	1.805665	1.830025
	Fall	0.96615	1.0944	1.55005	1.54685
	Winter	1.041975	1.17195	1.63815	1.63505
	Spring	0.9214	1.044	1.51036	1.501125

Price Scheme Differences

The differences between classes within the same season mostly result from different pricing schemes. C23 and C24 are assigned consistently higher marginal rates while C25 and C26 are assigned consistently lower marginal rates.

Table 4 – for the 2017/2018 billing period:

Classes	C23	C24	C25	C26
Marginal Rate	8.89696678	7.89723107	5.62501405	5.64070336

Table 5 – for the 2016/2017 billing period:

Classes	C23	C24	C25	C26
Marginal Rate	8.86940493	7.8133454	5.37756395	5.40219219

Intra-Class Analysis

Looking closer within the same class for the same season, different clusters also have different daily indifference ratios (DIRs). Because these clusters have *scaled* loads, we can attribute these differences to the shape of their load profile, which determines portions of their consumption that are charged with a higher price and portions that are charged with a lower price. To characterize the effects of their shape, we examine two related phenomenon, peak time and peak level.

Peak Level

In theory, the flatter the load profile, the more it should benefit from a real-time pricing scheme relative to a flat-rate scheme. This corresponds to a higher DIR. Energy charges (LMPs) can be seen as a congestion fee on the grid[7], which is higher with higher electricity usage. The peak of LMPs is a proxy to the system peak. Therefore, as system-wide peak loads are achieved, the marginal rate will, in theory, also hit its peak price. A running example to illustrate this theory can be found in Appendix II.

Metrics and Visualization

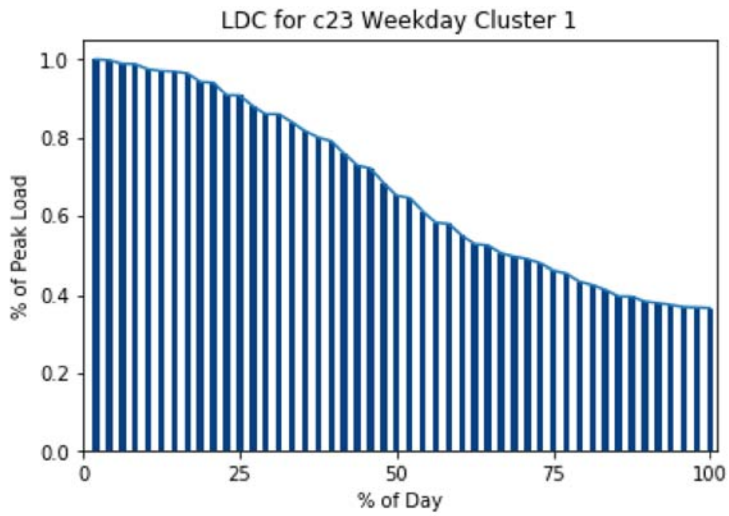
To evaluate the theory on peak level, we look at a couple metrics. For each cluster, the peak fraction (PF) can be defined:

This is a rough measure of how "peaky" the load profile is, such that a smaller PF is indicative of a higher peak level.

Since peak fraction is not a comprehensive metric, it is supplemented with a load duration curve (LDC) for a more qualitative interpretation of peak level. The LDC is a plot that scales the hourly energy load by the peak usage for that load profile and displays it in a descending order.

Figure 5 – Exemplary load duration curve (LDC).

[7] https://learn.pjm.com/three-priorities/buying-and-selling-energy/lmp.aspx



Given the LDC for each cluster, the scaled minimum load (SML) can be defined:

This effectively reflects the range of the LDC, making it a comparable metric between clusters. A smaller SML value would suggest a higher peak level.

Peak Time

In theory, the further away the cluster peak is from the system-wide peak, the more it should benefit from a real-time pricing scheme relative to a flat-rate scheme. This corresponds to a higher DIR. A running example to illustrate this theory can be found in Appendix II.

In this study, the effect of different peak times is largely measured by defining the time deviation (TD) of a cluster peak from system-wide peak load:

Here, a higher TD corresponds to larger peak time deviation and theoretically, a larger DIR.

Comparison of Factors

As historical system costs tend to peak in summers, shifting consumers' behavior during those seasons would significantly reduce system costs. As can be seen through the intra-class analysis, peak fraction (PF) and time deviation (TD) both play important roles in driving customer costs. As policy makers strive to make comprehensive recommendations, it's valuable to explore models to characterize the relative impact of the two factors.

Preliminary DIR Model

Consider a regression model to investigate how peak fraction (PF) and time deviation (TD) impact the daily indifference ratio (DIR). Through the model, we compare the relative influence of PF on DIR with the influence of TD on DIR.

We construct a multi-linear regression model such that:

By examining the sign and relative magnitude of and , we are able to first investigate how and impact . We then compare with to see which factor is a more important driver. According to the theoretical impacts of PF and TD (explored previously), and should both be positive.

However, as TD is a measurement of time and PF is a ratio designed to measure peak level, making direct unscaled comparisons would be challenging. Therefore, the independent and dependent variables are unity normalized, such that they're standardized to a mean of 0 and a standard deviation of 1.

Note that would be 0 as a result of unity normalization. The standardized variables are derived as follows:

Under standardized coefficients, direct comparisons between PF and TD are possible. Note that TD, in theory, is positively correlated to DIR, as shown in previous sections.

Table 6 – sample statistics for the two summer seasons. For each cluster, the PF coefficient is colored blue if larger than the TD coefficient. The TD coefficient is colored yellow if larger than the PF coefficient. The larger DIR of the two clusters in each season is colored gold.

c23 wd	Year	2016 Summer		2017 Summer	
	Cluster	0	1	0	1
	PF Coefficient	-0.1117	0.5985	0.6166	-0.0191
	TD Coefficient	0.2897	0.1669	0.1843	0.3164
	DIR	1.2501	1.2002	1.228	1.2722

c23 wn	Year	2016 Summer		2017 Summer	
	Cluster	0	1	0	1
	PF Coefficient	-0.0517	0.5927	0.0553	0.5667
	TD Coefficient	0.3342	0.2296	0.2797	0.2243
	DIR	1.2484	1.1982	1.269	1.223

c24 wd	Year	2016 Summer		2017 Summer	
	Cluster	0	1	0	1
	PF Coefficient	0.5722	-0.0496	-0.131	0.5279
	TD Coefficient	0.2955	0.3518	0.3455	0.3437
	DIR	1.3617	1.4103	1.4496	1.4026

c24 wn	Year	2016 Summer		2017 Summer	
	Cluster	0	1	0	1
	PF Coefficient.	-0.1965	0.4398	-0.1405	0.4209
	TD Coefficient	0.3281	0.2928	0.3124	0.2778
	DIR	1.398	1.3435	1.4427	1.3914

c25 wd	Year	2016 Summer		2017 Summer	
	Cluster	0	1	0	1
	PF Coefficient	-0.1326	0.349	0.3747	-0.0435
	TD Coefficient	0.3023	0.421	0.4458	0.3972
	DIR	1.8578	1.771	1.9324	2.015

c25 wn	Year	2016 Summer		2017 Summer	

	Cluster	0	1	0	1
	PF Coefficient.	-0.1091	0.279	0.2749	-0.0673
	TD Coefficient	0.3596	0.3882	0.3463	0.3107
	DIR	1.8389	1.75496	1.9076	2.0015

c26 wd	Year	2016 Summer		2017 Summer	
	Cluster	0	1	0	1
	PF Coefficient	0.0111	0.4737	-0.0528	0.4505
	TD Coefficient	0.3733	0.3812	0.3609	0.425
	DIR	1.8905	1.7974	2.0459	1.9541

c26 wn	Year	2016 Summer		2017 Summer	
	Cluster	0	1	0	1
	PF Coefficient	-0.1582	0.3775	-0.1008	0.3486
	TD Coefficient	0.2993	0.286	0.2955	0.3197
	DIR	1.8645	1.7677	2.028	1.9292

In general, the model agrees with our expectations. From the regression statistics, we can observe that the higher DIR often coincides with clusters that have higher TD coefficients, and the PF coefficient is more dominant in clusters that have a lower DIR.

After we are able to compare the coefficients directly, we need to unstandardize the coefficients in order to tell the "full story" behind the coefficients with the units that we wish to use. Therefore, we obtain unstandardized coefficients:

That is, one unit of change in TD will lead to a unit of change in DIR, and one unit of change in PF will lead to a unit of change in PF. Valid comparisons are based on and only

Refined MUR Model

Apart from regression on DIR, we also want to investigate how PF and TD directly impacts the total cost () consumers face under the marginal rate, as well as their relative impacts.

However, directly performing regression on is non-ideal because the higher total cost may be driven more by inherently larger magnitudes of consumption, as opposed to load profile shapes. Therefore, we run regression on the "average price per unit of electricity" consumers pay when charges are calculated based on marginal rates. In this way, consumers' total costs are scaled to a unit level, thereby controlling for the magnitude of the total load.

Define as:

Notice that DIR and MUR are negatively correlated:

Therefore, we construct a multi-linear regression model that is a refined version of the previous DIR model:

=

Note here we introduce an interaction term to investigate how the influence of one independent variable on the dependent variable varies as the other independent variable is assigned a different value. More specifically, we are investigating:

1. How the relationship between MUR and TD changes given different values of PF. Mathematically, the MUR regression could be rearranged as:

2. how the relationship between MUR and PF changes given different values of TD. Mathematically, the MUR regression could be rearranged as:

The same issue with direct comparison is resolved using the same standardization and unstandardization techniques as in previous sections. Mathematically speaking, we run regression on:

Where and are calculated as before and:

Under standardized coefficients, direct comparison is made possible. Note that DIR is theoretically positively correlated to TD and arithmetically negatively correlated to MUR. By extension, TD is expected to be negatively correlated with MUR.

Table 7 – sample statistics for the two summer seasons. For each cluster, the PF coefficient is colored blue if larger than the TD coefficient. The TD coefficient is colored yellow if larger than the PF coefficient. The larger DIR of the two clusters in each season is colored gold.

c23 wd	Year	2016 Summer		2017 Summer	
	Cluster	0	1	0	1
	PF Coefficient	0.0205	-0.2312	-0.1997	-0.0206
	TD Coefficient	-0.451	-0.19	-0.1592	-0.1506
	Interact Coefficient	0.2994	0.1457	0.1207	0.0871
	DIR	1.2501	1.2002	1.228	1.2722

c23 wn	Year	2016 Summer		2017 Summer	
	Cluster	0	1	0	1
	PF Coefficient	-0.023	-0.2815	-0.0266	-0.2462
	TD Coefficient	-0.4529	-0.399	-0.0947	-0.3162
	Interact Coefficient	0.2938	0.3308	0.0438	0.2564
	DIR	1.2484	1.1982	1.269	1.223

c24 wd	Year	2016 Summer		2017 Summer	
	Cluster	0	1	0	1
	PF Coefficient	-0.2243	-0.0485	-0.0218	-0.5395
	TD Coefficient	-0.3494	-0.7027	-0.4541	-0.7929

	Interact Coefficient	0.2907	0.4703	0.3549	0.5741
	DIR	1.3617	1.4103	1.4496	1.4026

c24 wn	Year	2016 Summer		2017 Summer	
	Cluster	0	1	0	1
	PF Coefficient	-0.0357	-0.3824	-0.0205	-0.3074
	TD Coefficient	-0.2322	-0.5899	-0.2487	-0.4404
	Interact Coefficient	0.1953	0.4395	0.1737	0.3125
	DIR	1.398	1.3435	1.4427	1.3914

c25 wd	Year	2016 Summer		2017 Summer	
	Cluster	0	1	0	1
	PF Coefficient	-0.02	-0.5101	-0.5848	-0.2922
	TD Coefficient	-0.7139	-1.1006	-1.246	-0.9809
	Interact Coefficient	0.4268	0.6974	0.8289	0.7198
	DIR	1.8578	1.771	1.9324	2.015

c25 wn	Year	2016 Summer		2017 Summer	
	Cluster	0	1	0	1
	PF Coefficient	0.0843	-0.5613	-0.5526	-0.2527
	TD Coefficient	-0.4076	-1.5390	-1.4176	-0.8201
	Interact Coefficient	0.0636	1.1955	1.1056	0.6679
	DIR	1.8389	1.75496	1.9076	2.0015

c26 wd	Year	2016 Summer		2017 Summer	
	Cluster	0	1	0	1
	PF Coefficient	-0.0993	-0.6049	-0.016	-0.6751
	TD Coefficient	-0.4942	-0.9397	-0.154	-1.1412
	Interact Coefficient	0.279	0.6819	0.0926	0.7741
	DIR	1.8905	1.7974	2.0459	1.9541

c26 wn	Year	2016 Summer		2017 Summer	
	Cluster	0	1	0	1
	PF Coefficient	0.0069	-0.1256	-0.0058	-0.1332
	TD Coefficient	-0.0745	-0.234	-0.1371	-0.236
	Interact Coefficient	0.0373	0.1895	0.0936	0.1717
	DIR	1.8645	1.7677	2.028	1.9292

The results of this regression generally corroborate the DIR regression model. Coefficients of TD are mostly higher than coefficients of PF, regardless of DIR. This is an "improvement" from the DIR regression where coefficients of TD are only mostly higher for clusters with the higher DIR. This suggests most costs facing clusters in summer months are driven by time deviation.

The process of unstandardization is the same as before:

Plugging in the unstandardized coefficients into the two rearrangements of the MUR regression to see how TD and PF are inter-dependent and simultaneously affect MUR, we will get:

1. A relationship between MUR and TD given certain values of PF:

Therefore, given a certain PF, one unit of change in TD will lead to a unit of change in MUR.

2. A relationship between MUR and PF given certain values of TD

Therefore, given a certain TD, one unit of change in PF will lead to a unit of change in MUR.

It should be noted that, in both DIR and MUR regression, there has been evidence of heteroscedasticity in the data distribution, especially as it pertains to the relationship between DIR and PF and the relationship between MUR and PF. However, the p-values obtained are sufficiently small to be statistically significant under both hetero- and homoscedasticity cases. Because the expected relation between these variables is not necessarily linear, standardized coefficients are solely for comparative purposes while unstandardized coefficients are solely for illustrative purposes. In this sense, the linear model only serves as an approximation of the positive or negative relationship between different variables, as the distribution of the data is varied and scattered.

Results and Discussion

Interpreting Summer DIR Regressions

Higher DIR is an indicator of higher savings potential when moving to a real-time pricing scheme. As the TD coefficient is higher than PF coefficients for summer clusters with a higher DIR, we can see that higher savings potentials are more driven by changes in TD. This suggests that for summer months, clusters that are overpaying relative to others are largely driven by time deviation. Therefore, policies should be more focused on shifting the peak time during summer months than in other seasons, in order to decrease system costs.

Interpreting Summer MUR Regression

We mostly observe from the summer MUR regression statistics. Such an inequality has important policy implications:

1. indicates that TD and PF both have a positive relationship with MUR. As MUR is negatively correlated with DIR, we can tell that TD and PF are negatively correlated with DIR, matching our theory. Therefore, policies that encourage peak shifting and peak flattening can both contribute to benefiting the entire system.
2. Moreover, indicates that TD has a stronger negative influence on MUR during summer months. Note that a lower MUR is an indicator of higher savings potential. Corroborating the DIR regression, we see that savings potentials are driven by changes in TD, and therefore policies should be more focused on shifting the peak time during the summer as opposed to other seasons.

As we rearrange the MUR regression to obtain the derived coefficients of for TD and for PF, it becomes clearer that TD and PF should not be considered in an isolated vacuum: the influence of PF on MUR is also determined by TD, and the influence of TD on MUR is also determined by PF.

More importantly, Consider the derived coefficients of for TD and for PF, together with the inequality and:

1. On the one hand, if we hold the peak level constant at a high value, then policies that encourage peak time shifting will benefit the system greatly. However, if the

consumers' load is very flat, then policies that encourage peak time shifting will not contribute much to reducing the system cost.

2. On the other hand, if we hold time deviation constant at a small value, then policies that encourage peak flattening would mitigate system costs. However, if consumers' load peak is further away from the system peak, then policies that encourage peak flattening will not contribute much to reducing the system cost.

References

Yang Yu, Guangyi Liu, Wendong Zhu, Fei Wang, Bin Shu, Kai Zhang, … Ram Rajagopal. (2018). Good Consumer or Bad Consumer: Economic Information Revealed From Demand Profiles. *IEEE*, Vol. *9* (No. 3).

Appendix I – Marginal Cost Pricing: Component Derivations

Different components introduced hereafter may vary based on:

Conceptually, capacity, delivery and transmission charges are "fixed" rates for each billing period, season and customer class. A broad definition for these rates is:

TAL is divided by billing period, season and class and is computed in the same manner for capacity, delivery and transmission charges.

TAC is divided into year, season and class as well, but is calculated differently for capacity, delivery and transmission charges:

Delivery Charge

The Delivery charge (in $/kwh) is:

Capacity Charge

The Capacity charge (in $/kwh) is

Transmission Charge

The Transmission charge (in $/kwh) is

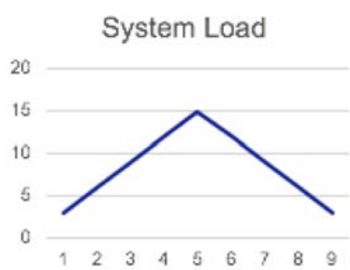
Energy Charge

Energy rates are dynamic as they change hourly. The energy charge used in the marginal rate scheme is the averaged seasonal hourly locational marginal prices (LMP). It can be expressed (in $/kwh) is

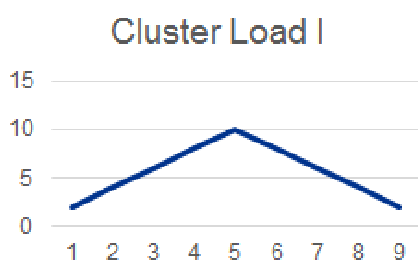
Appendix II – Illustrating Effects of Peak Level and Peak Time

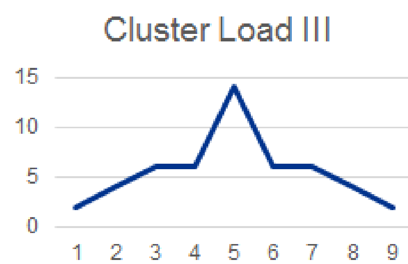
Peak Level

Examining an example case in which the system load has a certain distribution, we expect the unit price distribution to roughly coincide with system loads. In an ideal scenario:



Now, consider two clusters that have the same total daily load and both have their hourly peak at the same time as the system peak. The only variable is their peak level:

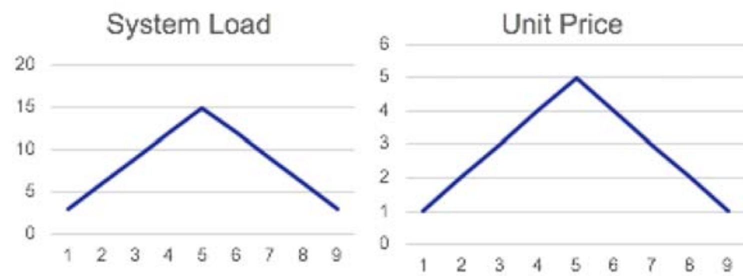




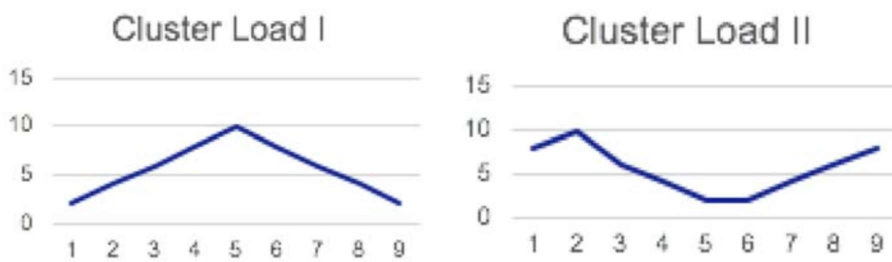
Note that cost is calculated as the product of price and load. As cluster III has a higher coinciding peak load with the price peak, we would expect cluster III to be charged a higher cost than cluster I. Thus, higher peak levels are associated with higher costs on a real-time pricing scheme.

Peak Time

Re-examining the same scenario considered above:



Now, consider two clusters that have the same total daily loads and peak levels. The only variable is their peak time deviation:



As cluster I has its peak load coinciding with the unit price peak, we expect cluster I to be charged a higher price than cluster II. Thus, larger time deviations (TD) are associated with lower costs on a real-time pricing scheme.

Appendix III – Cost Analysis Metrics

Table 8 – Compilation of indifference ratios (DIRs) for the 6/1/2016 – 5/31/2017 billing period.

			C23 – Single family w/o electric space heat		C24 – Multi family w/o electric space heat		C25 – Single family w/ electric space heat			C26 – Multi family w/ electric space heat	
Summer	Weekday	Cluster	0	1	0	1	0	1		0	1
		Indifference Ratio	1.2501	1.2002	1.3617	1.4103	1.8578	1.771		1.8905	1.7974
	Weekend	Cluster	0	1	0	1	0	1		0	1
		Indifference Ratio	1.2484	1.1982	1.398	1.3435	1.8389	1.75496		1.8645	1.7677
Fall	Weekday	Cluster	0	1	0	1	0	1		0	1
		Indifference Ratio	0.9601	0.9741	1.0911	1.1014	1.5726	1.5308		1.5356	1.565
	Weekend	Cluster	0	1	0	1	0	1		0	1
		Indifference Ratio	0.9726	0.9578	1.1026	1.0825	1.5719	1.5249		1.5661	1.5207
Winter	Weekday	Cluster	0	1	0	1	0	1		0	1
		Indifference Ratio	1.0455	1.0378	1.1668	1.1755	1.6529	1.6239		1.651	1.6219
	Weekend	Cluster	0	1	0	1	0	1		0	1
		Indifference Ratio	1.0458	1.0388	1.1784	1.1671	1.6534	1.6224		1.6509	1.6164
Spring	Weekday	Cluster	0	1	0	1	0	1	2	0	1
		Indifference Ratio	0.9259	0.9176	1.0436	1.0463	1.493	1.511	1.5265	1.5115	1.4953
	Weekend	Cluster	0	1	0	1	0	1		0	1
		Indifference Ratio	0.9258	0.9163	1.0354	1.0507	1.5312	1.4901		1.4777	1.52

Table 9 – Compilation of indifference ratio (DIRs) for the 6/1/2017 – 5/31/2018 billing period. Note that DIRs were not computed for spring 2018 as up to date LMP data was unavailable at the time of this report.

			C23 – Single family w/o electric space heat		C24 – Multi family w/o electric space heat			C25 – Single family w/ electric space heat		C26 – Multi family w/ electric space heat	
Summer	Weekday	Cluster	0	1	0	1		0	1	0	1
		Indifference Ratio	1.228	1.2722	1.4496	1.4026		1.9324	2.015	2.0459	1.9541
	Weekend	Cluster	0	1	0	1		0	1	0	1
		Indifference Ratio	1.269	1.223	1.4427	1.3914		1.9076	2.0015	2.028	1.9292
Fall	Weekday	Cluster	0	1	0	1	2	0	1	0	1
		Indifference Ratio	1.0347	1.0153	1.1912	1.1747	1.1638	1.7896	1.7258	1..7832	1.7194
	Weekend	Cluster	0	1	0	1	2	0	1	0	1
		Indifference Ratio	1.0341	1.0156	1.1612	1.19	1.1695	1.7763	1.7191	1.7745	1.7108
Winter	Weekday	Cluster	0	1	0	1		0	1	0	1
		Indifference Ratio	1.0851	1.0784	1.2177	1.2291		1.7399	1.7692	1.7639	1.7319
		Cluster	0	1	0	1		0	1	0	1
		Indifference Ratio	1.0796	1.0858	1.2299	1.2204		1.7704	1.7422	1.7634	1.7381