

# Rate Setting

James Woods

3/3/2017

# Overview

- ▶ The idea is to create a tariff a rate structure that,
  - ▶ Will satisfy the revenue requirements until the next rate case.
  - ▶ Balances efficiency, in the dead weight loss sense, with other objectives, e.g., equity.
  - ▶ Includes at least a nod to marginal cost pricing at some level.

# General Rate Patterns

- ▶ Two-Part
- ▶ Decreasing block tariff
- ▶ Increasing block tariff
- ▶ Peak pricing
- ▶ TOU

The big question? What do you mean by marginal cost.

## Variations for Other Purposes

- ▶ Life-line or low-income
- ▶ Electric heat or electric vehicle

# Data Overview

- ▶ Appliance surveys for bottom up
- ▶ System and rate class use time series for top down

Both are used to create models of future consumer behavior and forecast how the consumers will react to rates so that you can verify if the revenue requirement is satisfied.

## But First, Lets Talk Weather

How you model weather depends on the frequency of the load data you are working with.

- ▶ Weather data is commonly from NOAA and it was a pain to get with FTP servers.
- ▶ I used to have a cron job to get new California data every month and every day.
- ▶ NOAA has better APIs these days. For example to pull daily temperature data for PDX is

```
meteo_pull_monitors("USW00024229", date_min =  
"2016-01-01", date_max = "2016-12-31")
```

- ▶ A Year of hourly data is

```
isd(usaf="726980", wban="24229", year =2016)
```

## In General

- ▶ Try to get weather data that has similar frequency as your data.
  - ▶ If you are working with hourly load data, get hourly weather data.
  - ▶ If you are working with daily or larger aggregations of use, get daily weather and aggregate up.
- ▶ Working with hourly data can be hard
  - ▶ Not all, and not a majority, of the NOAA data will have hourly observations.
  - ▶ Daily summaries are more common and you can get closer geographic matches.
- ▶ Make trade-offs
  - ▶ You have fewer 1st order hourly stations ([http://www.weather2000.com/1st\\_order\\_wbans.txt](http://www.weather2000.com/1st_order_wbans.txt)).
  - ▶ Often does not cover all the territory.
  - ▶ You sometimes have to make up your own data for isolated places using nearby 1st order stations.

# What NOAA gives you

Metric!!

Core Values ([https:](https://www1.ncdc.noaa.gov/pub/data/ghcn/daily/readme.txt)

[//www1.ncdc.noaa.gov/pub/data/ghcn/daily/readme.txt](https://www1.ncdc.noaa.gov/pub/data/ghcn/daily/readme.txt))

- ▶ prcp: Precipitation, in tenths of mm
- ▶ tavg: Average temperature, in tenths of degrees Celsius
- ▶ tmax: Maximum temperature, in tenths of degrees Celsius
- ▶ tmin: Minimum temperature, in tenths of degrees Celsius
- ▶ awnd: Average daily wind speed, in meters / second
- ▶ wsfg: Peak gust wind speed, in meters / second

There are variables about gusts, clouds and other weather. Except for dew point, which is used for relative humidity, I have stayed away from the others.



## Warning about hourly weather

- ▶ These are mostly airports
- ▶ Higher wind
- ▶ Fewer trees
- ▶ They tend to be cooler than the non-airport. Keep in mind if you use a mix of hourly and daily measurements.

## Using Hourly Weather with Hourly Load

- ▶ Not as easy as just putting in the temperature
- ▶ Load for systems and individual households have an asymmetric response to temperatures.
- ▶ The load tends to increase with temperature when temperatures are high.
- ▶ Loads increase as temperature falls with low temperatures.
- ▶ The cut between low and high temperature is different by region and household and often called the set-point or base temperature.
- ▶ If you have a specification that allows for an asymmetric, U-shaped or V-shaped response, you should be OK.

# Typical Shape

- ▶ V shaped with an apex at around 68F (20C).
- ▶ Slope of load reaction to high temperatures about x3 the reaction to low temperatures.
- ▶ The low temperature load is either electric heat or air handlers.
- ▶ Techniques
  - ▶ Run model multiple times with different set-point and choose the one with maximum fit.
  - ▶ Estimate the set-point within the model by making the transition U-shaped rather than V-shaped.
  - ▶ Discretize temperature into 5F blocks and estimate effect of each temperature.

# The techniques (Scan Over Setpoints)

- ▶ Set up a load equation with

$$L_i = \alpha + \beta_1(b - T_i)I(b - T_i > 0) + \beta_2(T_i - b)I(T_i - b) + g(X_i)$$

- ▶  $b$  is the set-point
  - ▶  $T$  is the temperature
  - ▶  $I$  is an indicator function
  - ▶  $g(X_i)$  are the other variables
- ▶ Run regression over a range of  $b$  values and pick the one with the highest goodness of fit.

# What is Wrong With This?

- ▶ Everything !!!!
- ▶ Treats the unknown  $b - T$  data as known
- ▶ Estimates  $b$  outside the model even though it is used in the model and correlated with other parameters.
- ▶ Don't get me started on multiple comparisons.

How to fix this, pick a base temp without looking and stick with it.  
Easier and only has an errors-in-variables problem.

# Smooth or Discretize

- ▶ Loess, a non-parametric technique would work for this problem.
- ▶ Polynomial of degree 3 or higher, e.g.,  $\beta_1 T + \beta_2 T^2 + \beta_3 T^3$
- ▶ Discretize temperature
  - ▶ Create a dummy variable for temperatures in the,  $[30,35)$ ,  $[35,40)$  range and so on.
  - ▶ Similar to polynomial and loess in flexibility.
  - ▶ Does not guarantee monotonic response without restrictions.

# How Many Models

- ▶ It is possible for you to have one model, relation between weather and load, per hour per day of week,  $24 \times 7 = 168$ .
- ▶ More than likely you can aggregate some hours, many night time hours can be aggregated.
- ▶ Many days can be treated as the same:
  - ▶ Sometimes Weekday, Weekend
  - ▶ M-Th, F, Weekend.
  - ▶ M, F, T-Th, Weekend.
  - ▶ Some states have different Sunday shapes because of blue laws
- ▶ The models will be part of a system
  - ▶ Often previous hour and previous day are useful.
  - ▶ Often errors within a day have correlated errors, SUR models

# Dealing with Daily and Lower Frequency Data

- ▶ You often have to use daily data, because that is the best or only weather station.
- ▶ You are modeling daily peak load.
- ▶ Don't try averaging over longer periods.
- ▶ You are in the land of degree days.



# What is a degree day?

- ▶ Stole it from agriculture, growing degree day.
- ▶  $HDD(b) = \max(0, b - Avg)$
- ▶  $CDD(b) = \max(0, Avg - b)$
- ▶ Average Temperature is the average of the daily high and daily low.
  - ▶ Not the average of the hours.
  - ▶ With shared base, there are either HDD in a day or CDD but not both.

# The Logic of HDD and CDD

- ▶ Really tries to get at heating and cooling load.
- ▶ Not comparable if the diurnal swing is very different, Hawaii vs Mojave.
- ▶ Not knowing  $b$  has the same trouble as before.
- ▶ Errors in CDD because of unknown  $b$  cause correlated errors in HDD. (More Later)
- ▶ I have tried way more complicated things and XDD measures work.

# Typical Models

- ▶  $kWh = \alpha + \beta_1 CDD + \beta_2 HDD + g(X)$
- ▶ CDD and HDD often, but not always have the same base.
- ▶ There is a scanning methodology for the base and a within model technique

## Problems with CDD and HDD as commonly used

- ▶ People make econometrics mistakes that make their models look better.
- ▶ They treat HDD and the base temperature as a known instead of something that should be estimated.
- ▶ Upshot:
  - ▶ Biased effect of HDD and CDD parameter, one up and one down.
  - ▶ Variance of estimators are smaller than they should be.
- ▶ Use the spline based method, but watch for the point discontinuity.

<http://www.sciencedirect.com.proxy.lib.pdx.edu/science/article/pii/S014098831400139X>

# Heat Buildup and Other Effects

- ▶ Loads at the beginning of a heat wave are typically lower than loads later even when the temperature is the same.
- ▶ Happens when there is not enough time to let things cool off.
  - ▶ Think hot concrete
  - ▶ Attic spaces
- ▶ You often put in a lagged CDD and HDD in your model
$$kWh = \alpha + \beta_1 CDD + \beta_2 HDD + \beta_3 L(CDD) + g(X)$$
- ▶ or even polynomial lags
$$kWh = \alpha + \beta_1 CDD + \beta_2 HDD + \beta_3 \sum \gamma^n L^n HDD + g(X)$$
 and estimate the weight of the lag,  $\gamma$ .