

note

85 views

Power usage and payments have a time history

There has been a few discussions on classification of customers based on static data. However, the power usage and payment are dynamic and change with time.

Since time history of usage records are available to the utility company, can we use time series analysis like triple smoothing by Holt-Winters and/or ARIMA PDQ, or CUSUM to detect usage and payment changes? Note that these time based analyses are advantageous to the utility company because its competitors don't have this data and therefore they cannot provide the same service.

The payment history can also be analyzed with GARCH to check if there is any time dependent heteroskedasticity- that is if the payment irregularities have similar patterns now relative to last year. GARCH is widely used for value and risk estimations in the stock market- I wonder if the same concept can be applied to utility customers. Since zip code cannot be used as a classification, perhaps we can group the customers based on power usage and apply GARCH on different group sizes and see if there is any useful pattern there.

hw8

Updated 14 hours ago by Alexa Langford and Sandip

followup discussions for lingering questions and comments

☒ Resolved ☐ Unresolved**davethewave** 15 hours ago

I also considered using the GARCH for estimating the volatility of payments and taking this output as an input for the clustering method when considering which locations to be shut off. The higher the volatility estimated the higher the risk that the losses for the company might accumulate over time, thus prioritizing these one first when doing the optimization task.

☒ Resolved ☐ Unresolved**Tony ElHabr** 9 hours ago

To add to y'all's thoughts, I think it is possible to use time-series models for all customers, regardless of how long they have been with the company. The videos propose a regression approach to model customers without much data because time-series models would not work without some historical information. However, it is reasonable that the customers could be classified into groups/clusters based on other variables, which would allow us to use imputation to generate estimates of their historical power usage which could then be used for time-series models.

Additionally, if we were to do some kind of clustering-imputation combination like this, we could apply the same framework to create time-series models for small groups of customers who are alike in all ways except for perhaps location of residence. Fitting time-series models for each customer individually might be overkill, while fitting a single model for all customers together seems naive. A cluster-based subsetting of customers into groups, each calling for their own models, seems like a good compromise.