ELECgrid: Higher Accuracy in Forecasting Electrical Consumption Using Small Area Estimates

Edwin Lim Jun Yun
Lee Kong Chian School of Business
Singapore Management University
Singapore, Singapore
edwinlim.2016@business.smu.edu.sg

Maegan Joyce Wu
Lee Kong Chian School of Business
Singapore Management University
Singapore, Singapore
maeganwu.2016@business.smu.edu.sg

Wong Ming Sen
Lee Kong Chian School of Business
Singapore Management University
Singapore, Singapore
mswong.2016@business.smu.edu.sg

Abstract - Singapore is rolling out its plan for the privatisation of the electricity market [1]. Currently, there are as many as 12 electricity retailers who typically charge a price lower than that set by Singapore Power – the de facto energy retailer. These retailers purchase electricity in bulk from electricity generating companies. At the moment, electricity retailers are unable to forecast electricity consumption accurately with the kind of data available publicly (*average electricity consumption per postal code*). This inevitably means that retailers are not buying a close enough amount of electricity to meet the actual electricity demand of their customers. Resources and potential revenue are therefore being wasted and lost because of this. This project therefore utilises Small Area Estimate to improve accuracy of forecasting by combining the already available data on *average electricity consumption per postal code* and other auxiliary information.

I. Introduction

Just last year in 2018, Singaporean households have been able to choose their electricity provider instead of solely getting it from Singapore Power -- a government electricity provider. This recency caused an inevitable pursuit amongst private retailers to capture as much market share as they can. To do so, private retailers are aggressively lowering their prices, just to get residents to patronise them.

However, apart from increasing their revenue through acquisition of more customers, it is imperative to consider the cost of their operations as well. As most private retailers do not own power plants, they procure electricity in bulk from generating companies instead. The amount of electricity they purchase is typically found through Direct Estimation using the monthly average electricity consumption per postal code data available in the Energy Market Authority website. However problem arises due to its very low accuracy. The average value assigned to each postal code does not coincide with the true values associated for each unit in the block. Not all units in the blocks are surveyed every month. Hence it fails to take into account that there might be differing electricity consumption in different months.

Ultimately, our project's objective is to reduce electricity retailers' cost and wasted resources by more accurately estimating the total monthly electricity consumption. Our team wanted to undertake this since cost-savings for retailers meant lower prices for consumers. This is of great opportunity to look into as various consumers even our families are shifting into private-provided electricity.

II. Related Works

There has been a very recent research study by Yong Ying Joanne Tan and Dr Tan Kim Seong called *Exploring and Visualizing Household Electricity Consumption Patterns* *in Singapore: A Geospatial Analytics Approach.* The research revolves around the use of local indicators of spatial association (LISA) to cluster households based on electricity consumption patterns, and to find out the variation across different dwelling types in the same postal code. Similarly, the research aims at allowing electricity retailers make better informed business decisions, such as choosing the variety and pricing of plans to offer consumers. However, this differs from our primary goal as our project focuses on providing a more accurate forecasting.

III. Analytical Method

A. Small Area Estimate (SAE)

SAE allows us to obtain accurate and reliable estimates from a sampling data, at area or unit level, compared to using direct estimates. This is because SAE allows us to calculate the EBLUP, which is a combination of direct and regression-synthetic estimators [2].

However, our project does not require sampling because the data provided is comprehensive (parameter is known) [3].

B. Data Selection and Preparation

We retrieved the Data, *Average Monthly Household Electricity Consumption per Postal Code Jan-Dec 2016* from ema.gov.sg. This is because the data is recent, provides the required information to perform SAE, and is stratified into different dwelling types. Ultimately, we want to measure the electricity consumption by the different dwelling types as we recognize that difference amongst these types vary largely.

Next we retrieved the shapefile *Subzone_HDB_Postal* from data.gov.sg which contains the information on the number of units for each dwelling type in a postal code.

The data imported was in .xls format and was extracted using the readxl package. The purrr package was then used to efficiently combine the monthly electricity consumption across all excel sheets together into one table. Subsequently, the data *Subzone_HDB_Postal* was imported and both the tables are joined together using the dplyr package. The data is cleaned and any missing or unnecessary information were removed.

Following that, we multiplied the number of units in each dwelling type with the *"Average Electricity Consumption per Dwelling Type in a Postal Code"* in order to compute and find the "*Total Electricity Consumption / Dwelling Type in a Postal Code"*. This is done because "total" data are more accurate compared with "average" data.

C. Direct VS SAE Approach

Step 1) We first grouped the postal codes into their respective subzones. We then compute the direct estimates in each subzone for each dwelling type for each month. This is done by finding the mean of the "Total Electricity Consumption" in the subzone [4]. Subsequently, the variance and covariance is computed. It is important to note that direct estimation is a naive approach in estimating the electricity consumption in a subzone as compared to the SAE approach.

Step 2) With the direct estimate calculated, the linear mix model is built by adding the auxiliary variable "Total Population per dwelling type housed in a Subzone". This done so that the model can link all areas through a common parameter so as to “borrow strength” from related areas [2].

Using the SAE approach, we can find the EBLUP and MSE in their respective subzones. The restricted maximum likelihood (REML) is used to fit the model as it accounts for the degrees of freedom to reduce random effects and errors. The default number of iterations used is 100 and the precision of the fisher-scoring algorithm optimized the result within an acceptable range of error. Note that if the precision used is above 0.01, the results will have a confidence interval equal or lower than 99% and the results becomes less reliable. Subsequently, the MSE is used to compute the EBLUP covariance.

Step 3) We then plot and compare the results using the ggplot2 package.

Results How the geospatial analytics tool developed help to discover new understanding from the data.

IV. User Interface Design

The dashboard (refer to image 1) is supported by R Shiny, with the base logic coded entirely in R. As seen from the left panel, users are able to choose a specific dwelling type, month and a subzone. The map then shows a visualisation of the subzone chosen along with the EBLUP value of total electricity consumption.

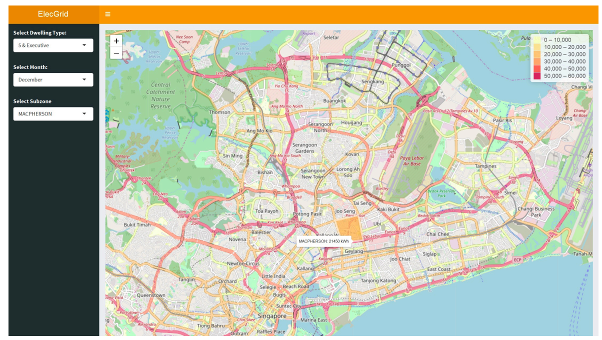
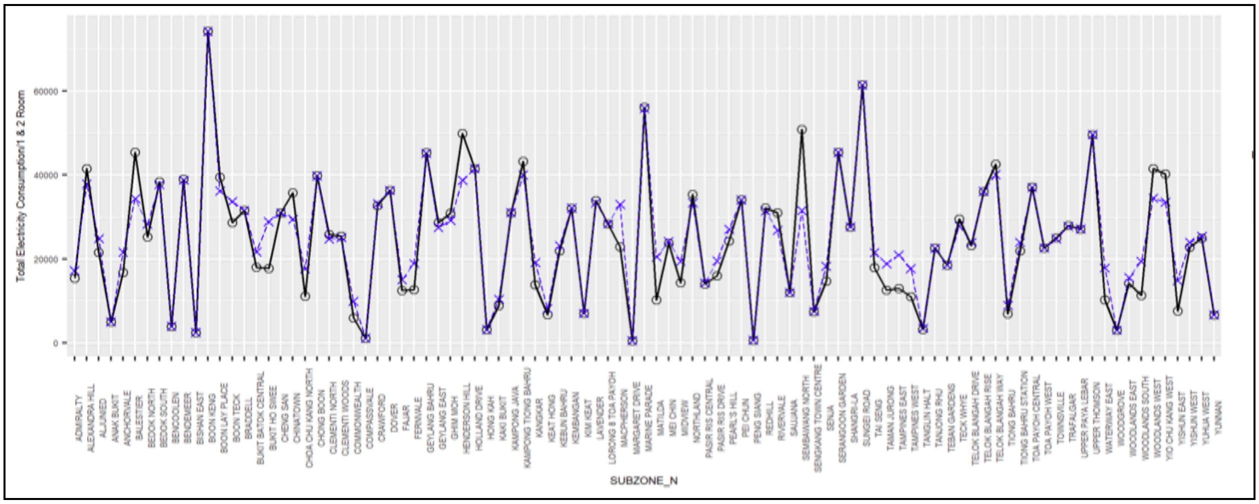


Image 1: R Shiny Dashboard

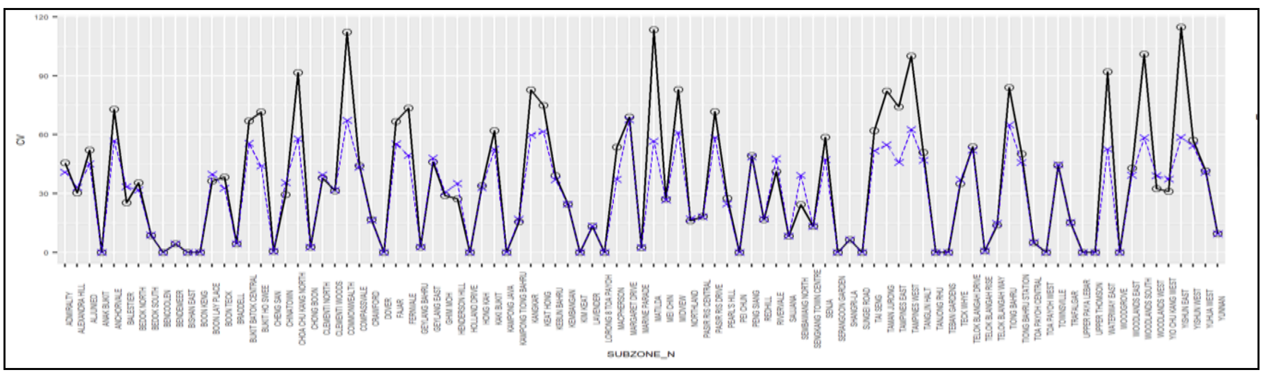
V. Discussion

The charts below illustrates differences between Direct Estimates (black) and EBLUP (blue) in terms of Total Average Electricity Consumption (Chart 1) and Coefficient of Variation or CV (Chart 2). Based on the analysis, we can observe that the total electricity consumption values for EBLUP are less volatile and more stable as compared to the direct estimates.

Furthermore, the CV of EBLUP is generally lower than that of Direct Estimates, which means that the errors are generally lower compared to the Direct Estimation approach. Therefore, electricity retailers should use the EBLUP information to calculate and forecast the electricity demand in a subzone as they are more reliable as compared to direct estimates.



Eg: Averaged Total Electricity Consumption in 1&2 room flat June (Chart 1)



Eg: Coefficient of Variance for 1&2 room flat June (Chart 2)

VI. Future Work

Since SAE makes use of auxiliary information to improve the accuracy of estimates, introduction of more auxiliary variables such as demographics of the households (income level, number of children and working adults et cetera) will improve the usability of our model.

It is important to note that our project's domain is centered around individual subzones only due to lack of access to unit-level data regarding electricity consumption. Due to the lack of budget and the very nature of this being a school project, we are not able to acquire these commercially available data. Nonetheless, making use of such data will undeniably improve the accuracy and usefulness of our model as well.

Aside from spatial information, we can also look into temporal using the Dynamic Time Warping (dtw) algorithm [5]. It is not uncommon that there are various unpredicted significant events that caused critical changes in electricity consumption. An example is the significant fall in electricity consumption in the UK during Princess Diana's funeral. The use of dtw overlooks these significant events to achieve consistency in trend across the years.

VII. Acknowledgement

The authors would like to thank Dr Kam Tin Seong at Singapore Management University for the kind support and guidance.

VIII. References

[1]Brand Studio (2019). Open Electricity Market: A new world of choices. Channel News Asia.
Retrieved 14 April 2018 from:
https://www.channelnewsasia.com/news/advertorial/open-electricity-market-a-new-world-of-choices-11053226

[2]Molina I and Marhuenda Y (2015). *sae: An R Package for Small Area Estimation.* The R Journal. Vol. 7/1. Pp.81-98,
Retrieved 14 April 2018 from:
https://journal.r-project.org/archive/2015-1/molina-marhuenda.pdf.

[3]T. Lumley (2012). survey: Analysis of Complex Survey Samples, 2012.
Retrieved 14 April 2018 from:
http://CRAN.R-project.org/
package=survey

[4]Molina I and Marhuenda Y (2015). Basic direct and indirect estimator in sae package.
Retrieved 14 April 2018 from:
https://cran.r-project.org/web/packages/sae/vignettes/sae_basicdirect.pdf

[5]Toni Giorgino (2009). *Computing and Visualizing Dynamic Time Warping Alignments in R: The dtw Package.* Journal of Statistical Software, 31(7), 1-24,
Retrieved 14 April 2018 from:
doi:10.18637/jss.v031.i07.