# Abstract

# Acknowledgements

# Table of Contents

# List of Tables

# List of Figures

# List of Graphs

# Introduction

## Background

The Energy Market Authority (EMA) is a government entity started to promote competition in Singapore’s energy market and to ensure the energy supplied is reliable and secure.

Power generation companies generate electricity from natural gas and oil and sell them to the wholesale market. Electricity retailers then buy the electricity from this wholesale market and sell to their customers.

Since 2001 [1], EMA has slowly opened the energy retail market to other competitors for business consumers. The high electrical consumption by these businesses means that the cost of electricity plays a big part in their running cost and profits. They will benefit from the increase flexibility and choices when choosing their own retailers who offer different pricing plans for different needs.

Then in May 2019, EMA fully opened the energy market to all households and smaller business accounts. Everyone can now benefit from the flexible plans offered by the numerous retailers.

As of March 2020, there are 3 ways of purchasing electricity for households: Fixed price plans, discounted off the regulated tariff plans and wholesale price.

Fixed price plans are like traditional regulated tariff from SP Group where a fixed price agreed beforehand and calculated per kWh is billed every month.

The discounted off the regulated tariff plans offers a fixed discount of the traditional regulated tariff from SP Group and calculated per kWh for billing every month.

Lastly, any consumers can buy from the energy supplier directly in the wholesale market where half-hourly prices are used to determine the cost at time of usage. However, a 10-day lag is imposed in calculating and releasing of the wholesale price.

Having the ability to forecast future prices is important to both electrical suppliers and consumers during the bidding process. Suppliers can optimise their generation of electricity to prevent wastage and consumers can adjust their usage habits, less during high prices and more during low prices.Thus successful predictions can lead to rewarding monetary returns.

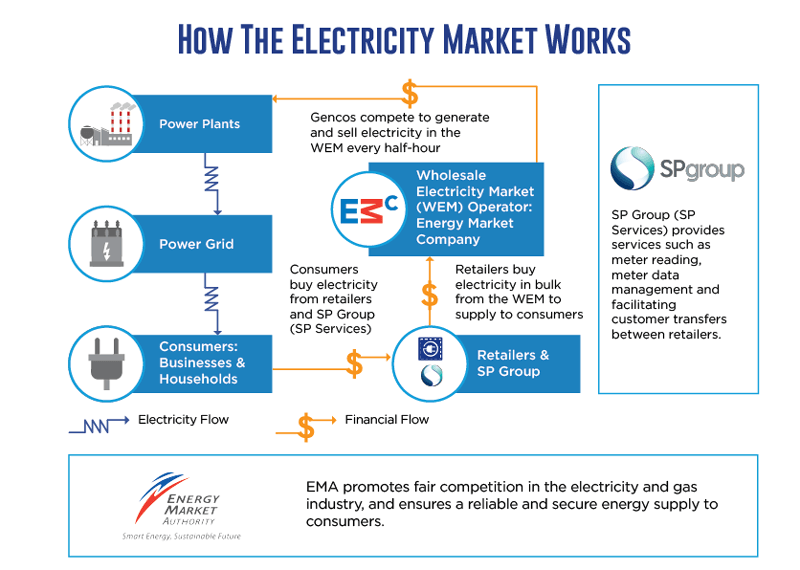


Figure 1. How the electricity market works

## Purpose and scope

Many researches had been done on predicting future electrical price. Both statistical methods and artificial neural network have both achieved certain rates of success in predicting the future.

Statistical methods like Autoregressive Integrated Moving Average (ARIMA) usually involve solely the historical prices to perform the regression while the artificial neural network can involve both historical prices and other factors that can influence the electrical prices.

For this project, we will be focusing on the artificial neural network method to forecast the Wholesale Electricity Price (WEP) in Singapore. The WEP is the cost charged to the consumer upon the time of use, and ultimately the price that the consumer must pay.

Due to the lag time of 10 days in which the WEP is released to the public, we will need to be able to forecast 10 days in advance for the viability of this project.

# Literature Review

## Factors Affecting Electric Prices and Trends

Sanjeev et al. generalise electric price to contain the following attributes: high volume data with small timesteps, constantly changing mean and variance, highly volatile and outliers are common. They deduce that this is due to the non-tangible aspect of electricity where it cannot be easily stored and there must be an equilibrium between the electric load and generators. Also, demand rarely changes over the small timeframe and the electricity market is commonly oligopolistic. Lastly, both load and generation of electricity can be affected by very unpredictable events. Sudden rain and cloud covers can reduce electricity generated via solar while unforeseen dip in temperature can lead to higher load consumed by heaters. [2]

Sanjeev et al. then categorised the probable factors affecting the price of electricity into 5 classes: market characteristics, nonstrategic uncertainties, other stochastic uncertainties, behaviour indices, and temporal effects.

|  |  |
| --- | --- |
| **Class** | **Input variable** |
| **Market Characteristics** | (1) Historical load f(load), (2) System load rate, (3) imports/exports, (4) capacity excess/shortfall (5) Historical reserves (6) Nuclear, (7) thermal, (8) hydro generation, (9) generation capacity, (10) net-tie flows, (11) MRR, (12) system’s binding constraints, (13) line limits |
| **Nonstrategic Uncertainties** | (15) Forecast load, (16) Forecast reserves, (17) temperature, (18) dew point temperature, (19) weather, (20) oil price, (21) gas price, (22) fuel price |
| **Other Stochastic Uncertainties** | (23) Generation outages, (24) line status, (25) line contingency information, (26) congestion index |
| **Behaviour Indices** | (27) Historical prices, (28) Demand elasticity, (29) bidding strategies, (30) spike existence index, (31) ID flag |
| **Temporal Effects** | (32) Settlement period, (33) day type, (34) month, (35) holiday code, (36) Xmas code, (37) clock change, (38) season, (39) summer index, (40) winter index |

Table 1. Factors influencing electric prices

Using these factors, they further classify them into different input variables used by different researchers with their own predictive models. Specifically, majority of the factors used are those of (1) Historical load, (15) Forecast load and (27) Historical prices in Table 1.

## Methodology in Price Forecasting

### Data Pre-processing and Analysis

According to 2 papers [3] [4], removal of price spikes and outliers gives better accuracy in their neural network models with H.Y. Yamin et al improving their models from 39.89% and 15.47% in mean absolute percentage error (MAPE) of their training and testing sets respectively to 7.98 and 13.7%. Furthermore, instead of simply removing the spikes, a price ceiling was implemented, allowing the model to be trained with these spikes still.

In Singapore context, Shrestha and Qiao was able to determine that the available generation capacity has the greatest influence in determining the price of the electricity [5]. The correlation was only relevant when the mean price was calculated over a time period but the spot price during time of usage is more crucial for end users to optimise their electricity usage and reduce cost.

### Neural Network

# Discussion

## Data and Analysis

All data are downloaded and compiled from the Energy Market Company Pte Ltd (EMC) website. They are the middleman between electricity buyer and seller, regulating the market and providing the trading infrastructure for Singapore.

The first five rows of the 2019 data are shown in Figure 2 where the Wholesale Electricity Price (WEP), Uniform Singapore Energy Price (USEP), power demanded, gross power generated and net power generated are compiled against the datetime index.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **DATE** | **PERIOD** | **WEP ($/MWh)** | **USEP ($/MWh)** | **DEMAND (MW)** | **GROSS INJECTION (MWh)** | **NET INJECTION (MWh)** |
| 1/1/2019 0:00 | 1 | 83.33 | 82.7 | 5201.89 | 2555.1 | 2362.296 |
| 1/1/2019 0:30 | 2 | 83.83 | 82.71 | 5150.461 | 2549.41 | 2357.615 |
| 1/1/2019 1:00 | 3 | 83.19 | 82.7 | 5106.794 | 2519.013 | 2327.042 |
| 1/1/2019 1:30 | 4 | 83.13 | 82.69 | 5075.841 | 2492.473 | 2300.457 |
| 1/1/2019 2:00 | 5 | 83.2 | 82.67 | 5044.147 | 2453.576 | 2261.511 |

Figure 2. 2019 Electric Data

The prices of 2019 in Figure 3 intuitively shows large amount of random price spikes and any trends or seasonality are not immediately obvious within a year.

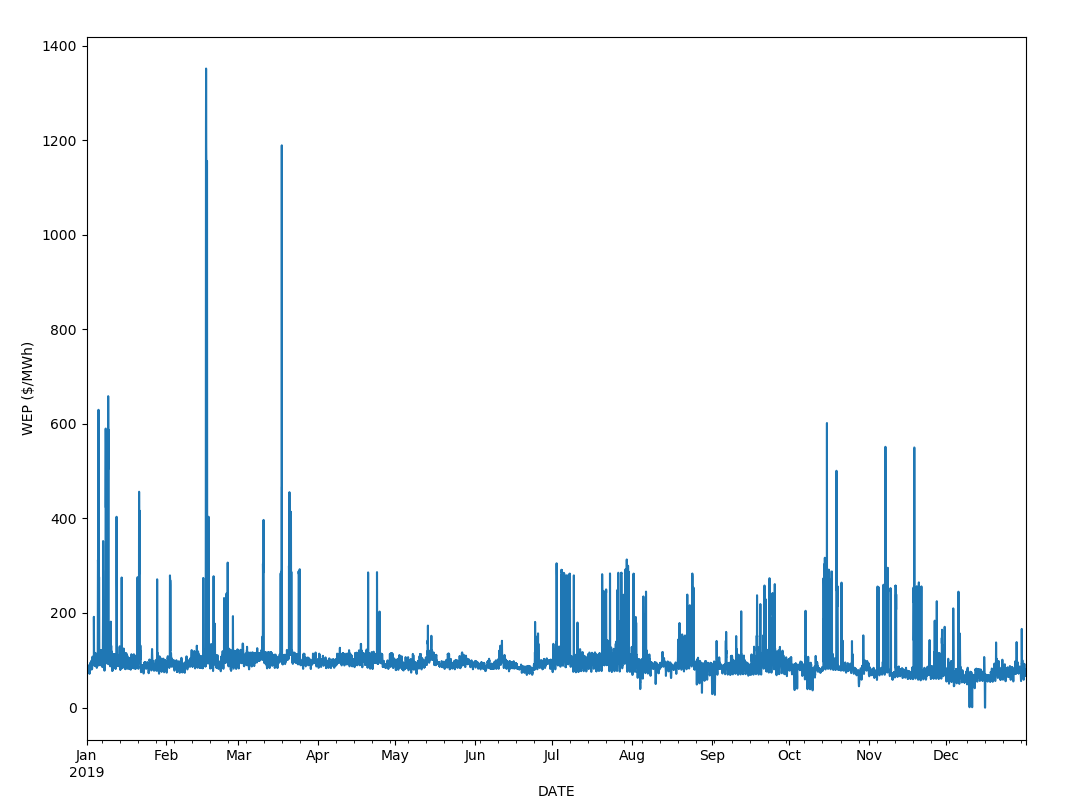


Figure 3. Electric Prices in 2019

In Figure 4, outliers were truncated to within 3 standard deviation from the mean to allow better visualisation of the monthly and daily trends if any.

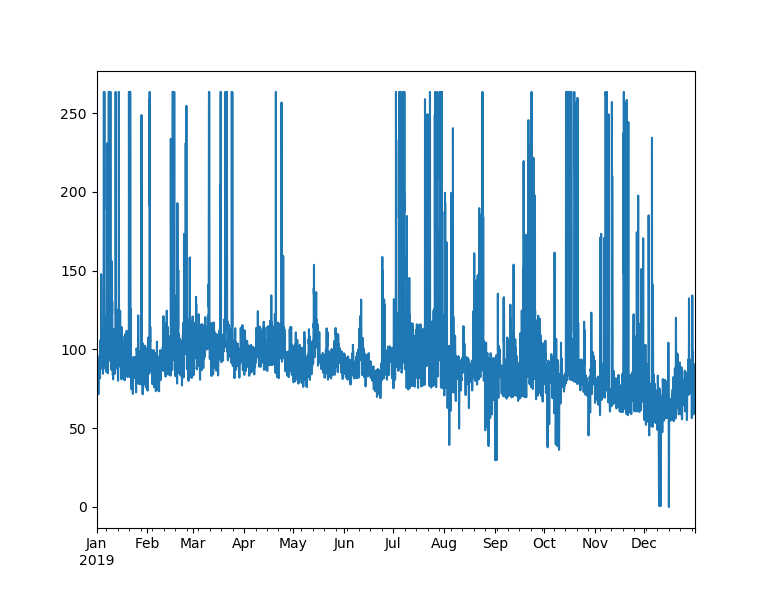


Figure 4. Outliers truncated

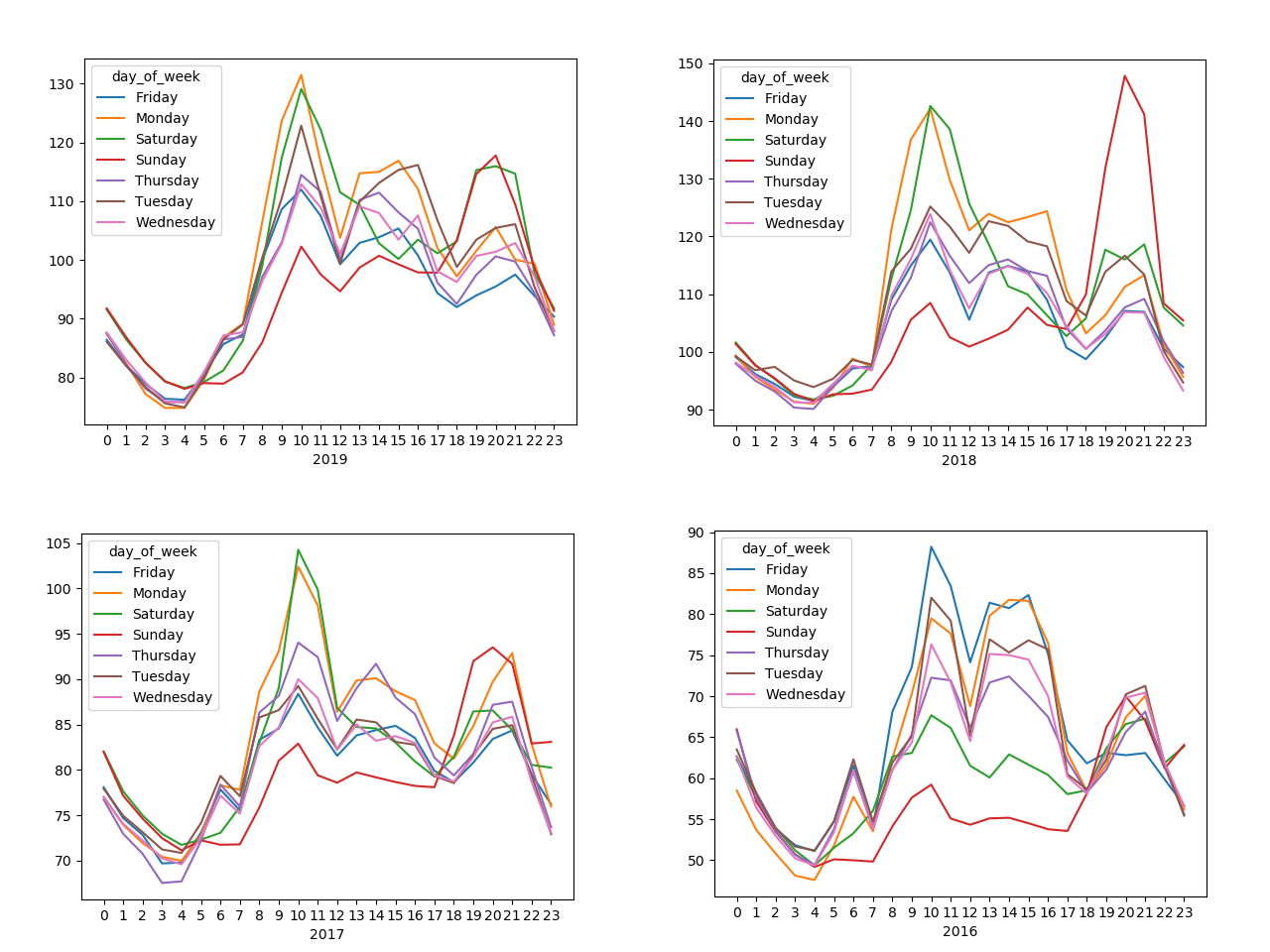
We further analyse the effect of date and time on the WEP in Figure 5. Observation on the daily mean price shows the same trend on the past 4 years of data. The duration between 8 to 10 am and 6 to 9 pm shows large spikes in prices across all the days of the week with Sunday being the lowest during the day but highest during the night. 

Figure 5. Daily mean price data plotted across the hour of the day over the past 4 years

Plotting the correlation matrix in Table 2 shows that neither demand nor supply are highly correlated to the WEP. The high correlation between WEP and USEP is due to WEP being derived from USEP and the different tariffs and administrative costs. From these observations, trying to predict the spot prices simply using demand and supply in Singapore may not be very effective. A timeseries approach will be taken for the prediction.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **PERIOD** | **WEP ($/MWh)** | **USEP ($/MWh)** | **DEMAND (MW)** | **GROSS INJECTION (MWh)** | **NET INJECTION (MWh)** |
| **PERIOD** | 1 | 0.141533322 | 0.143468486 | 0.538817247 | 0.555029534 | 0.56109735 |
| **WEP ($/MWh)** | 0.1415333 | 1 | 0.999900995 | 0.25060302 | 0.256192447 | 0.253536614 |
| **USEP ($/MWh)** | 0.1434685 | 0.999900995 | 1 | 0.254470731 | 0.26015551 | 0.257450558 |
| **DEMAND (MW)** | 0.5388172 | 0.25060302 | 0.254470731 | 1 | 0.995496467 | 0.992566336 |
| **GROSS INJECTION (MWh)** | 0.5550295 | 0.256192447 | 0.26015551 | 0.995496467 | 1 | 0.995956884 |

Table 2. Correlation matrix of electric price data

## Feature Engineering

Since the weekly data shows signs of seasonality and the characteristics of the prices, we propose a 7-day time lag to be used to train and predict the next day price. 7 days of data requires a lag time step of 336 and a day of data needs 48 future time step due to the half-hour pricing in the dataset. Thus, we can create a supervised dataset where the input and output vector will be 336 and 48 respectively, totaling to a size of 384 columns in our training data.

The 2019 WEP price will be used for training and validation while the 2020 WEP price up to March will be used to testing.

### Creating the training set

#### Transform data into timeseries

Generating the training data requires first converting the 2019 data into a timeseries data. The datetime index will help to slice the data during the process of training and validating.

|  |  |
| --- | --- |
|  | WEP |
| 2015-01-01 00:00:00 | 92.055 |
| 2015-01-01 01:00:00 | 90.660 |
| 2015-01-01 02:00:00 | 82.655 |
| 2015-01-01 03:00:00 | 76.625 |
| 2015-01-01 04:00:00 | 75.865 |
| 2015-01-01 05:00:00 | 75.325 |
| 2015-01-01 06:00:00 | 76.685 |
| 2015-01-01 07:00:00 | 77.510 |

Table 3. First 8 items of the WEP Timeseries

#### Building the lag dataset

The first 336 columns *(t-336, t-335, t-334, t-333, …, t-4, t-3, t-2, t-1)* are generated from the price before timestep *t*. The next 47 columns *(t+1, t+2, t+3, t+4, …, t+44, t+45, t+46, t+47)* are generated from the price after timestep *t*.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| DATE | var1(t-336) | var1(t-335) | var1(t-334) | var1(t-333) | … | var1(t+44) | var1(t+45) | var1(t+46) | var1(t+47) |
| 2019-01-08 00:30:00 | 83.83 | 83.19 | 83.13 | 83.20 | … | 115.01 | 103.04 | 93.63 | 105.15 |
| 2019-01-08 01:00:00 | 83.19 | 83.13 | 83.20 | 78.81 | … | 103.04 | 93.63 | 105.15 | 100.07 |
| 2019-01-08 01:30:00 | 83.13 | 83.20 | 78.81 | 78.30 | … | 93.63 | 105.15 | 100.07 | 97.01 |
| 2019-01-08 02:00:00 | 83.20 | 78.81 | 78.30 | 73.55 | … | 105.15 | 100.07 | 97.01 | 97.54 |
| ... | ... | ... | ... | ... | … | ... | ... | ... | ... |
| 2019-12-30 22:00:00 | 80.92 | 75.99 | 71.83 | 67.34 | … | 79.40 | 79.94 | 78.58 | 75.21 |
| 2019-12-30 22:30:00 | 75.99 | 71.83 | 67.34 | 66.62 | … | 79.94 | 78.58 | 75.21 | 72.43 |
| 2019-12-30 23:00:00 | 71.83 | 67.34 | 66.62 | 66.32 | … | 78.58 | 75.21 | 72.43 | 69.58 |
| 2019-12-30 23:30:00 | 67.34 | 66.62 | 66.32 | 58.67 | … | 75.21 | 72.43 | 69.58 | 70.21 |

Table 4. Supervised data generated from 2019 data

#### Building the test dataset

Similarly, the 2020 data must be converted into a supervised dataset but without the need of the 47 future time steps. Only the input vectors will be fed into the model for prediction and the results will be compared to the existing data.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| DATE | var1(t-336) | var1(t-335) | var1(t-334) | … | var1(t-2) | var1(t-1) | WEP |
| 2020-01-08 00:30:00 | 75.50 | 72.69 | 70.36 | … | 65.77 | 53.85 | 69.74 |
| 2020-01-08 01:00:00 | 72.69 | 70.36 | 70.33 | … | 53.85 | 69.74 | 55.00 |
| 2020-01-08 01:30:00 | 70.36 | 70.33 | 69.55 | … | 69.74 | 55.00 | 53.31 |
| 2020-01-08 02:00:00 | 70.33 | 69.55 | 68.62 | … | 55.00 | 53.31 | 52.34 |
| ... | ... | ... | ... | … | ... | ... | ... |
| 2020-03-01 21:30:00 | 79.59 | 76.70 | 73.22 | … | 79.78 | 85.19 | 79.49 |
| 2020-03-01 22:00:00 | 76.70 | 73.22 | 72.83 | … | 85.19 | 79.49 | 73.27 |
| 2020-03-01 22:30:00 | 73.22 | 72.83 | 70.71 | … | 79.49 | 73.27 | 71.14 |
| 2020-03-01 23:00:00 | 72.83 | 70.71 | 70.53 | … | 73.27 | 71.14 | 70.11 |

Table 5. Testing et generated from 2020 data

## Building the Neural Network Model

* Number of epoch
* Batch size
* Number of neurons
* Number of layers
* Lag time
* Forward time

# Conclusion