GROUP REPORT BY TEAM 11

AEMC - How do you train a battery algorithm to make as much money as possible?

Alif Khondaker (z5112689)
Fawaz Mohamed Fazeel (z5205497)
Thazmeel Mohamed Rabeek (z5209070)
Riyaz Mohamed Rabeek (z5258129)
Syed Warisi (z5259749)

Term 3 2020

SUBMITTED IN PARTIAL FULFILMENT OF THE
REQUIREMENTS OF THE CAPSTONE COURSE DATA3001

Plagiarism Statement

I declare that this thesis is my own work, except where acknowledged, and has not been submitted for academic credit elsewhere.

I acknowledge that the assessor of this thesis may, for the purpose of assessing it:

- Reproduce it and provide a copy to another member of the University; and/or,
- Communicate a copy of it to a plagiarism checking service (which may then retain a copy of it on its database for the purpose of future plagiarism checking).

I certify that I have read and understood the University Rules in respect of Student Academic Misconduct, and am aware of any potential plagiarism penalties which may apply.

By signing this declaration, I am agreeing to the statements and conditions above.

Signed: Alif Khondaker Date: 21st Nov 2020

Signed: Fawaz Mohamed Fazeel Date: 21st Nov 2020

Signed: Thazmeel Mohamed Rabeek Date: 21st Nov 2020

Signed: Riyaz Mohamed Rabeek Date: 21st Nov 2020

Signed: Syed Warisi Date: 21st Nov 2020

Acknowledgements

Many thanks must go to our lecturers for the continuous aid in the completion of this report.

Many thanks must also be given to Oliver Nunn for the guidance and clarity in helping us tackle this problem.

Abstract

Australia's electricity market experienced a period of structural reform through the introduction of competitive markets and deregulation. As a result, a wholesale electricity market was established to facilitate the trade of electricity in Australia. There have been various emerging new technologies in the electricity sector to maintain system reliability and security. In particular, batteries are an emerging industry in Australia's energy market which play an important role in the energy market by both getting in energy and distributing it out. Opportunities have emerged for technical development in this area, such as the development of bidding algorithms for batteries, so that retailers can optimize their bidding strategy to ultimately maximize their profits or minimize risk in the wholesale electricity market.

This article analyses Linear Regression, Multiple Linear Regression and SARIMAX in the context of predicting energy dispatch prices in MWh. The requirements of this task stems from determining whether a battery should charge or discharge at a certain time interval to maximize profits.

The optimization considers a linear programming framework to determine the recommended decisions for each time interval. The decisions are outputted in the form of a number between and including -1 and 1. A positive value means it should be charging and a negative value means a state of discharging.

We eventually are shown that the simple linear regression model is better than the SARIMAX in a forecasting framework. In addition to this, increasing the frequency of decision making for the battery leads to an increase in the computational possibilities an optimization engine can work on, which consequently allows for the possibility to increase profits.

In essence, investing in improving predictions to unravel subtle trends and patterns alongside the development of an optimization engine is a profitable prospect in the long run. Though this report focuses on optimizing in a Linear Programming framework using Gurobi, other optimization frameworks like Quadratic programming (QP), and Quadratically constrained programming should also be considered to analyse data from different constraints.

Table of Contents

Chapter 1
Introduction

Electricity pricing is a key issue vital to both the competitiveness of Australian products abroad, and to supplying an essential service to homes nationwide. Since the 1990's, Australia's electricity market experienced a period of structural reform through the introduction of competitive markets and deregulation. Subsequently, the National Electricity Market (NEM) was established, which is a wholesale market to allow for the trade of electricity in Australia between generators and retailers (energy.gov.au, 2017). The NEM is responsible for keeping the power grid balanced as energy must be produced and used simultaneously and must ensure changes in demand are handled in real time (NERA, 2007).

Batteries play an important role in the energy market by taking in, storing and distributing energy. Batteries may either purchase energy when charging, or sell energy when discharging. They will generate energy at a certain cost $X / MWh, and will offer energy at a price of $Y / MWh. These auctions occur every 5 minutes. If the bids are priced substantially high, the optimal strategy is to sell energy, i.e. discharge the battery. On the contrary, when the bid prices are low, it is then best to buy energy, i.e. to charge the battery. This is our fundamental principle to implement in the battery algorithm to derive profits. The frequency of auctions paired with the fluctuations of the price of energy enables the possibility of incurring losses or minimising profits if batteries do not regulate their decisions on when to charge and discharge. Therefore, it would be beneficial to develop algorithms which allow batteries to maximise their profit through their contribution to the market (AER, 2018).

As such, electricity pricing models which attempt to predict the spot price of electricity in the wholesale market are a vital component of an organisation's decision-making processes, as it allows for a much more informed dscison. Due to the nature of batteries and the bidding process, by being able to forecast price to a certain degree of accuracy, retailers can optimize their bidding strategy to ultimately maximize their profits or minimize risk in the wholesale electricity market. In chapter 2 of this report, we will look at the extant methodologies and literature on electricity price forecasting.

Our intuition is to develop a dynamic model that forecasts the bid prices for some hours into the future by accounting for error rate, among other statistics, of our prediction - which we can determine by comparing our model's predicted bid prices and that of real-time data.

Chapter 2
Literature Review

This section will provide a brief outline of existing knowledge regarding electricity price modelling and the use of auto-bidding technology to maximize profit. These are the core processes which will be used to develop a methodology to answer our problem. This literature review will cover two main themes. The first of these will be an overview of electricity pricing models, which will give an insight into the current methods to forecast the spot price of electricity in the wholesale market. The second will be on the optimization and development of auto bidders to determine a firm's optimal bidding strategy.

Electricity Pricing Models

There are 3 main types of models relevant to our work: Statistical Models, Artificial Intelligence (AI) Models, and Multi-Agent Models

1. Statistical Models

Statistical methods/econometric analysis for predicting price use a combination of previous price and exogenous variables. Some of these variables may include weather factors and consumption factors such as temperature, rainfall, demand, and various others. Statistical modelling methods of forecasting price are desirable as the physical interpretation of the variables involved give us an insight into the understanding of their behaviour and how it may relate to the endogenous variable. However, the accuracy and efficiency of statistical models are dependent on the data and its quality. Statistical models are also less efficient when there is a lot of volatility in the data i.e. existence of spikes (Ventosa et al., 2005). Some types of statistical methods are going to be discussed below.

a. Regression Models

Multiple Linear Regression (MLR) is a very common statistical technique which aims to explore the relationship between several explanatory variables on the dependent variable. MLR is estimated through Ordinary Least Squares (OLS) which minimizes the sum of squared distance between the predicted and observed values. Under the Gauss-Markov assumptions, the OLS estimates are linear, unbiased and have minimum variance. The standard MLR model is given by:

$$y_i = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n + \varepsilon_i$$

Where y_i is the dependent variable, x_1 ... x_n are the explanatory variables, ε_i is the error term containing unobservables, β_0 is the intercept and β_1 ... β_n are the coefficients. MLR are often used as foundational models and can be built upon to create more sophisticated models. However, MLR itself is still also a reliable and popular statistical method for forecasting electricity price (Weron, 2014).

b. Autoregressive (AR) Time Series Models

The underlying concept behind autoregressive time series models is to predict something based on past values of that same thing. In an AR model, the forecast of the dependent variable depends on the linear combination of a stochastic term and lagged variables of the dependent variable. An AR model of order p can be given by:

$$y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$

Where c is the constant term, y_{t-1} ... y_{t-p} are the model's parameters and ε_t represents the white noise. Another useful time series model is the autoregressive moving average (ARMA) model, in which the current price depends on the linear combination of the variable's own lagged values and the error term and its lag/s. ARMA models are useful for modelling a weakly stationary stochastic time series, and hence is a useful model for predicting future values of electricity price (Weron, 2014). An ARMA model of order p and q is given by:

$$y_t = c + \varepsilon_t + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

Where the 1st half of the equation is the AR model we saw above, and the latter half is the moving average component of the model which are the lags of the error term. The ARMA model can be generalized to account for non-stationarity, and this model is called the autoregressive integrated moving average (ARIMA). It eliminates non-stationarity by adding a differencing step. However, one of the limitations of the ARIMA model is that it poorly accounts for data that is seasonal and differencing at lag-1 or greater may still not remove seasonality. It is important to obtain a model that can capture seasonality, as electricity prices are likely to show season signals (Weron, 2014). An extension of the ARIMA model which captures seasonality is the seasonal autoregressive moving average model (SARIMA).

c. Autoregressive Time Series Models with Exogenous Factors (ARX)

The AR models discussed previously are time series models which are dependent on previous values of itself. However, we mentioned that electricity prices are likely to be affected by previous and current values of itself and other exogenous factors. The AR models discussed previously can be extended to capture the relationship between electricity price and exogenous variables by simply adding these variables to the AR time series models. As such, the AR, ARMA, ARIMA and SARIMA models become ARX, ARMAX, ARIMAX and SARIMAX when exogenous variables are added to the models (Hyndman and Athanasopoulos, 2016).

2. Artificial Intelligence Models (AI)

AI techniques have been developed to deal with problems that statistical methods are limited by. AI models combine several computational techniques such as machine learning, fuzziness, and evolution to create models that are capable of dealing with non-linearity and complex dynamic systems. The key strength of these models is their flexibility and ability to handle complex data, recognize intricate patterns and ultimately gain better forecast accuracy. As such these models have been favoured for electricity pricing models in extant studies. One limitation is that there are various AI methods to model electricity price and it is difficult to determine the most optimal solution and to compare between the different AI models (Weron, 2014). Some AI models to forecast electricity prices include:

a. Feedforward Neural Network

Feedforward neural networks are a type of artificial neural networks (ANN) where the connection between nodes does not create a cycle. A neural network model for forecasting electricity prices is through a seasonal autoregressive neural network (SAR-NN) which is a

dynamic feedforward ANN. This model can be used to not only gain forecasts for electricity price but also a prediction interval, as well as this the model robustly handles seasonality and non-linearity issues which are prevalent in standard ANN models (Saâdaoui, 2016).

b. Support Vector Machines (SVM)

SVM models are machine learning models which use learning algorithms to analyse data and to ultimately be used for regression/classification analysis. These models non-linearly map data to a higher-dimensional space, then the SVM training algorithm to test new data and create a hyperplane for classifying the data into different spaces. Once the data has been classified it can be used for forecasting, SVM models are especially effective in the application regarding non-linear regression. SVM models are usually a component of hybrid models, in the context of forecasting electricity prices, SVM models can be used to classify electricity price data and then forecast prices within each cluster (Weron, 2014).

3. Multi-Agent Models

Another modelling approach for electricity price is multi-agent models, which involves forecasting price by matching supply and demand in a simulated market through the interaction of heterogeneous agents with each other. Multi-agent models are catered towards qualitative analysis such as how agents may react to changes in price. As such, these models would not be suitable for our study as we are after quantitative forecasts of price in order to ultimately determine optimizing building strategies for batteries (Karakatsani and Bunn, 2008).

Optimization and Bidding

Achieving optimal battery performance in the energy market is through the implementation of a bidding algorithm that creates optimal bids. An example of this technology is Tesla's autobidder software used in their batteries, which allow batteries to autonomously dispatch energy and ultimately maximize revenue (Tesla, 2019). Tesla's autobidder consists of a library of algorithms developed through machine learning and classical statistical principles. Currently the most advanced autobidder, it provides; price forecasting, load forecasting, generation forecasting, dispatch optimization and smart bidding functionalities (Tesla, 2019). Tesla's batteries and autobidder software have been successfully implemented in The Hornsdale Power Reserve in South Australia. This large-scale battery storage project has driven down energy prices for the consumer and has improved system security within South Australia (Aurecon, 2018).

From the perspective of this project, the objective is mapped into Linear Programming program problem, given the forecasted dispatch price of electricity at time interval 'h' into the future, the algorithm formulates a linear function in which it considers linear multiples of the variables with cost coefficients to optimize for an objective function. This objective function is the profit function. Where the profit is the revenue gained through selling electricity or discharging electricity into the market minus the cost of charging the battery at given intervals. Profit = $Revenue - $Cost.

Chapter 3
Material and Methods

3.1 Software

To develop the model that can predict the fluctuation of prices, Python and R were used for data scrubbing and data exploration primarily due to their ability to process large amounts of data as opposed to the more limited capabilities of Microsoft Excel. We used packages such as NumPy, pandas for numerical analysis and packages such as matplotlib and Plotly and seaborn for data visualization. When it comes to building machine learning models, we used python programming language libraries called scikit-learn and statsmodels where we imported packages for the ARIMAX and SARIMAX model. Furthermore, we used R for statistical analysis for multiple linear regression to explore the usefulness of the features we developed. Finally, jupyter notebooks and the R programming scripts were used to manage, store and organize the code.

3.2 Description of the Data

Two datasets that were most relevant in producing models to determine price forecasts were the dispatchprice_fy2009-2019 (*DISPATCHPRICE*) and dispatchregionsum_rrponly-_fy2009-2019 (*DISPATCHREGIONSUM*). *DISPATCHPRICE* records five-minute dispatch prices for energy and denotes whether an intervention like price override (for example the Administered Price Cap) has occurred. It updates when price adjustments occur, where the new price is written to the *RRP* field, and the old price to the *ROP* field as an audit trail. The initial data was too complicated in order to account for all 56 features. There was a disproportionately large number of features in reference to the status of the different types of intervention that may occur during auctions like the Market Price Cap, Market Price Floor, Administered Price Cap and Administered Price Floor etc. The features that were most important in predicting the price fluctuations were the *SETTLEMENTDATE, REGIONID, DISPATCHINTERVAL, INTERVENTION* and *RRP*. The data had observations from 7/1/2008 12:00 am to 6/29/2010 4:10 am. The CSV file *DISPATCHPRICEREGIONSUM* also contained these variables, but on top of that, we acquired the respective demand from *TOTALDEMAND* in MWh of energy every five minutes for each other states. The following are the main variables:

- '*SETTLEMENTDATE*' (format: Date) - which provides the market date and time,
- '*DISPATCHINTERVAL*' (format: integer) - Dispatch interval identifier 001 to 288 in format YYYYMMDDPPP.
- '*INTERVENTION*' (format: Boolean) - Intervention by the organisation in rare circumstances for electricity generation.
- '*RRP*' (format: float, units: $AU per MWh) - Regional Reference Price for the respective dispatch period. RRP is the price used to settle the market.
- '*REGIONID*' (format: string, categorical variable) - Denotes whether the observation is from NSW, QLD, VIC, SA or from TAS.
- '*TOTALDEMAND*' (format Number (15,5), units: MWh)) - from both of the datasets combined.

Using this information, the aim is to predict the dispatch prices ($/MWh) with lower and upper bounds ranging from -$1000.00 to $14500.00 for all regions as per our analysis upon the minimum and maximum values and the distribution of historical data.

3.3 Pre-processing Steps

Starting from the initial state of possessing two datasets, we isolated variables that we found held significance in predicting prices for the future. Then by using $SETTLEMENTDATE$, $REGIONID$ and python's .merge() method, the datasets were merged to create a panel data consisting of price. This provided two variables; time and demand in order to predict the prices. The decision was then to determine whether to have the $REGIONID$ as categorical variables and use it as a feature to train the model or to split the dataset into smaller sets by their respective states. The method chosen was the latter due to the assumption that splitting would remove the mixture of irregularities from other states, which in effect removes unnecessary noise while also improving the accuracy of the model. This decision was also made to the possibility of extrapolating new variables characteristic to the analysis of the data of each state, as it would be more insightful than looking at all the observations at the same time for all states.

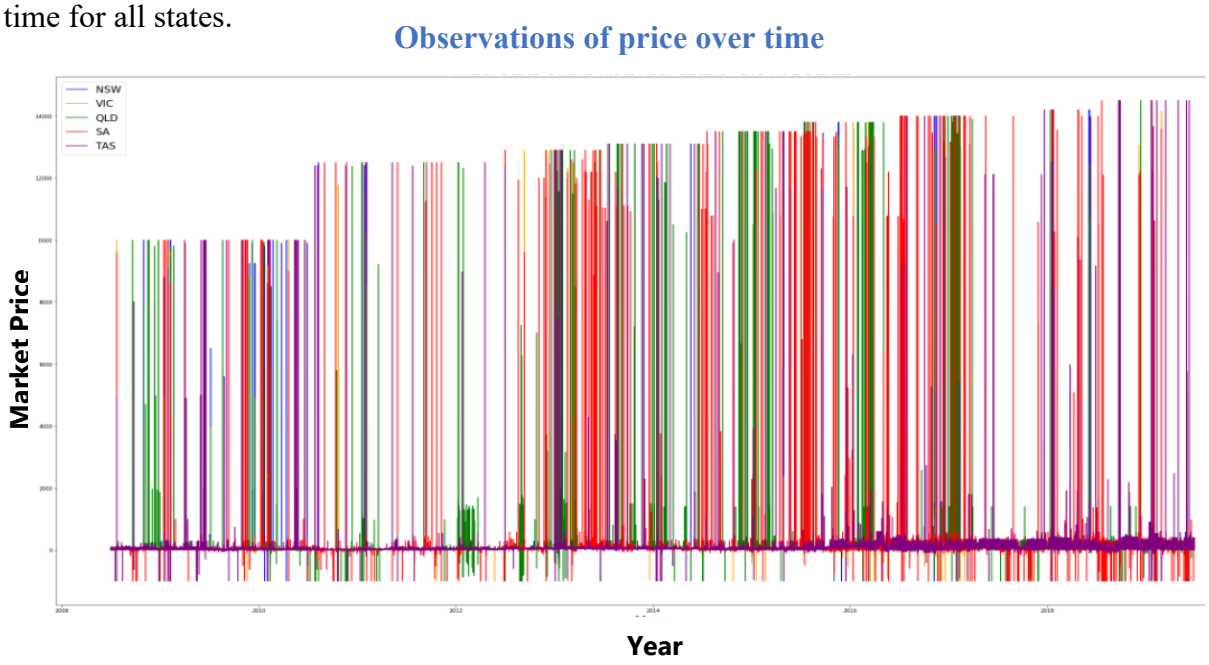

Figure 3.3.1
This figure illustrates the unorganised state of the data prior to our processing

The next step was to remove instances that possessed an $INTERVENTION$ value of 1 (forcing $INTERVENTION = 0$) to delete instances where the market price was overridden to insert an alternative. This was primarily a suggestion by Oliver Nun to lower down the complexity of the model. However, while analysing what type of effect this could have on our model, we observed that there were no instances in our dataset with $INTERVENTION = 1$, suggesting a net impact of zero upon our analysis from this "transformation". Subsequently, we removed the entire column as it provided no value for our prediction.

Furthermore, as a result of high volumes of data, there was excessive information to derive useful insights into the daily, monthly or yearly trends from the graph. Hence, to observe

yearly trends, a new price was formed by performing daily averages. This way, 288 observations - which is the total number of observations per day - were reduced to 1 - representing the daily average price. The mean statistic was an optimal choice in this step to incorporate all the data within a single day and map to a single value.

Figure 3.3.2
Visual representation of the change in prices of all 5 regions, from 2008 to 2019

Moreover, casting SETTLEMENTDATE variable to type Datetime aided in formulating our time series data. This created the training set for SARIMA() and ARIMA() models. Whereas for the regression model, a new variable called "timeID" was created to locate the n^{TH} time interval throughout the day, where n [0,287], with 0 being the first and 287 the last 5 minute interval of the day. This was primarily done as a means to overcome the problem of being unable to implement a datetime variable based feature in accounting for the nature of prices to be dependent on the time of day it is situated in. Hence, an integer-based feature ($timeID$) was developed to fit into the linear regression analysis.

A new dataset for each region was also created for the regression model containing the variable for time of the day ($timeID$) and the average price ($AvgRRP$) and the average demand in thousands ($AvgDemand$) at that particular time in the day. In the following steps of our report, this dataset was a huge improvement in allowing us to visualise trends in price and demand throughout the day for each state, and also allowed for the analysis and regression process to be much more efficient, as we could now work with a much smaller number of observations of the data while also capturing the trends that they follow.

Finally, feature extraction was performed. Categorical features such as $hour\_group$, $month\_group, IsWeekend, IsWeekday$ were extracted. When all the possible features were finalized, the feature column was normalized by subtracting it by its mean and dividing by its standard deviation. This is to ensure that any column that exists in the dataset has the

same scale and their predictive power only comes from the information they hold, not how big their values are.

3.4 Data Cleaning

R and Python were utilised for initial data cleaning and merging multiple datasets due to their ability to handle high volumes of data. During our inspection we noticed a unique data point, an observation included '$SNOWY$' as its $REGIONID$. This instance was removed as there were no records following up the market price fluctuation for electricity in this region. The data provided to us from AEMC was already relatively clean, most of the efforts before the stages of analysis were primarily focused on pre-processing and thereby little emphasis in cleaning was placed other than the minor type-casting used to format the dataset in order to be processed by models.

3.5 Assumptions

a. In reducing high volumes of data and creating a smaller dataset, we reduced 10 years' worth of data into one year where each observation in the data frame corresponds to 1 hour. We assume going from a gap of 5 minutes per observation to 1 hour per observation will not have any significant impact in optimization to make a profit.

b. In our models utilising Multiple Linear Regression (MLR), it is assumed that the conditions required to model the data using MLR are held.

c. It is assumed that using the average price and demand throughout different times in the day will provide meaningful insight into the data, and allow us to still come up with an efficient model after estimating the mean. Although we are predicting the mean, this prediction will give us an insight into the trend that the data follows throughout the day.

d. The number of output variables from the gurobi optimization model are dependent upon the frequency of decision making (whether to charge or discharge) the battery is made to do. If the decision is made every 5 minutes then the output would be the amount of profit gained over the day and 288 decisions based on the predictions of our model one day into the future. However, the frequency is increased to 30 minutes, 60 minutes or 120 minutes, the model is only making decisions either 48, 24, or 12 times over the day respectively. Hence, a reliable mechanism is required to map 288 values to either 48, 24 or 12 values depending on the frequency of decision made while also representing all the observations. An assumption is made that splitting the dataset into N bags (N is equal to 288 divided by the number of 5 minute intervals there are within the time span we take in making decisions with the model) is a reliable representation of the entire days' worth of data.

e. The gurobi optimizer operates in a Linear Programming framework, there by it is assumed that the objective profit function is a linear function of the dispatch prices that occur at every 5 minute interval.

3.6 Modelling Methods

There were 3 main modelling methods utilised in this study for the modelling and predicting of price:

1. Multiple Linear Regression
2. ARIMA
3. SARIMAX

1. Multiple Linear Regression

Multiple linear regression was performed on the variables in the created average price and demand data set, which included a variable for the time of day ($timeID$), the average price at that particular time ($AvgRRP$) and the demand of electricity at that specific time in thousands ($AvgDemand$). It was necessary to create a smaller dataset for 2 main reasons:

1. The immense amount of data provided would slow down our analysis and visualisation of the data considerably, as even running some simple regressions with that amount of data may take several hours. This would not allow us enough time to try multiple different models and decide on the best possible model to describe the data.
2. The observations in the provided dataset were very volatile - there were many observations in the provided data that were potential outliers, and due to the immense number of observations provided it was difficult to determine which particular ones were outliers and which were significant observations.

We started with an initial model:

$$Average\ price \sim Average\ Demand + Time\ in\ the\ day$$
$$AvgRRP \sim AvgDemand + timeID$$

The summary of the results of this model for each region can be seen in the figure below.

Region	Adjusted R^2	RMSE
NSW	0.67	6.73
QLD	0.57	6.75
SA	0.50	15.45
TAS	0.65	4.20
VIC	0.80	5.95

Figure 3.6.1
Adjusted R^2 and root Mean Squared error values of model the data of the different regions using the model:
$$AvgRRP \sim AvgDemand + timeID$$

Exploring the bivariate relationships between the variables allowed us to determine which transformations of the predictors would result in a better model. After transformations, best subset selection and stepwise regression - both forward and backward selection - were used to determine which predictors would give rise to the best model, according to the lowest AIC.

The model was then revised after transformations to the dependent variable (price). The models for each region were then judged according to adjusted R^2 , with the best model being jusdged as the one with the highest adjusted R^2.

2. ARIMA and 3. SARIMAX

ARIMA, which stands for Auto Regressive Integrated Moving Average, is a model that predicts future values given time series based on its own past values and its own lags. It is capable of modelling non-seasonal time series data and data that is not randomly distributed. It is characterized by three terms: AR term, MA term and differencing term required to make time series stationary.

If the time series dataset has seasonal patterns, then the SARIMAX model is preferred and a seasonal term is added.

Chapter 4
Exploratory Data Analysis

4.1 - Python

In peaking onto the first day of the dataset, we noticed an affirming trend of the general intuition that revolves around price fluctuation of electricity. It is clear that all five states have a strong representation of the prices to peak during both the morning and evening peak hour while simultaneously experiencing a reduction in prices between the peaks. It is observable that the second peak is almost always higher in magnitude than the first - indicating the demand for electricity is lower during the morning peak hours (6:00 am to 9:00 am) in comparison to the evening 5:30 pm to 8:00 pm peak periods.

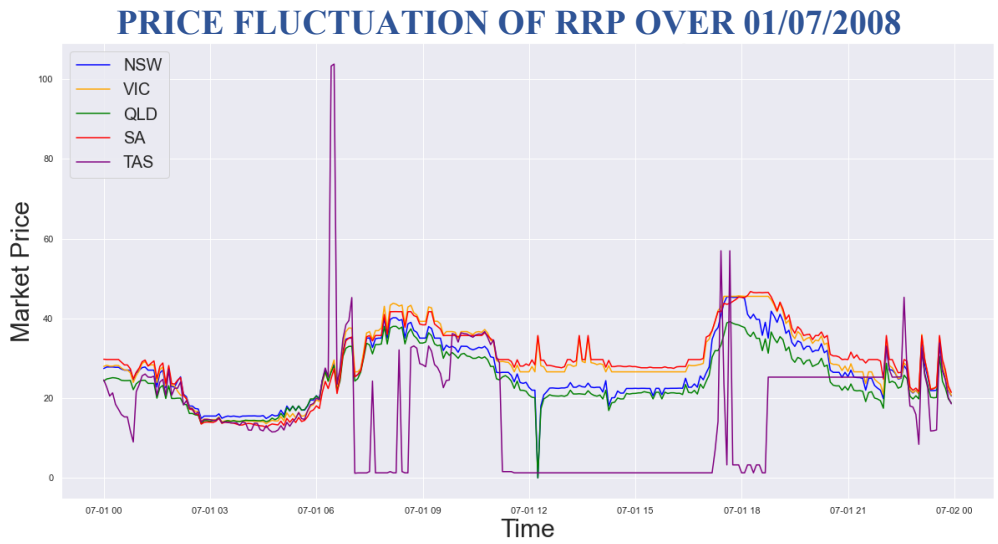

Figure 4.1.1
Change in price over the course of the first day in the given data set. Prices recorded at 5 minute intervals from 12am – 11:55pm on 1 July 2008.

This provided the intuition for the developing features that accounted for the prices being in a particular period/interval of time within a day, to be specific two features regarding the two peaks of the day in the Multiple Linear Regression model.

Further exploratory analysis revealed the lower and upper bounds of the data [-$1000.00, $14,500.00], which was contrary to the bounds relayed to us by the sponsor, which was [-$15,000.00, $15,000.00]. This could be a direct result of the market price management systems that apply price caps and floors in response to the activation of certain flags. Incorporating these additional variables were initially considered but were later dropped due to the loaded complexity it brings to the objectives of the project.

In comprehending the extremities of the data, one must ask, how come electricity generators are willing to accept losses or pay the AEMC to offload their electricity? What does it mean to have negative prices as for $dispatchprice$ (RRP)? What does it mean to have extremely high prices?

Negative dispatch prices could occur for many reasons. It is perhaps cheaper for a battery to continue the operation of electricity charging and discharging in comparison to a shutdown. Many renewable energy resources are being implemented into the grid to lower the demand on coal-powered generated electricity. From the below data and from external research, we can infer the fact that states like Tasmania, ACT and SA have the highest proportion of renewable electricity implemented within their grid (Climate Council, 2018). This is a clear indicator of the competition of renewable energy in having the ability to lower down the cost of electricity per MWh.

Region	Number of Negative cases
NSW	120
VIC	2391
SA	9332
QLD	1095
TAS	12057

Figure 4.1.2
Number of negative observations of price recorded across the different regions

DISTRIBUTION OF PRICE DATA FOR DIFFERENT REGIONS

	RRP Distribution				
	NSW	VIC	QLD	SA	TAS
Observations	1156896	1156896	1156896	1156896	1156896
Mean	54.14	53.23	55.59	65.27	57.51
S.D.	193.77	216.17	281.29	354.77	172.25
Min	-1000.00	-1000.00	-1000.00	-1000.00	-1000.00
25%	27.30	25.04	25.10	27.46	28.28
50%	43.17	39.37	41.50	43.32	40.93
75%	60.11	59.23	59.73	68.08	68.80
Max	14500.00	14500.00	14500.00	14500.00	14500.00

Figure 4.1.2
Summary statistics of electricity price across all states.

On the other hand inspection of the distribution of data for the different regions reveals roughly a normal distribution. Since 50% of the data is concentrated within the range of 27.30 and 60.11. This means, making the data stationary would probably not be required due to the property of normal distributions having constant mean and constant variance.

$$If\ X \sim N(\mu, \sigma^2)\ then\ E(X) = \mu\ and\ Var(X) = \sigma^2$$

However, when the data of the five states were analysed through autocorrelation, it was evident that the data lacked the property of constant variance due to high fluctuation of the

data and significant divergence from the lag axis. Differencing the RRP column once eliminated the initial heavy divergence from the axis. An illustration of the kernel density estimation of the data is shown in Figure 4.1.3.

4.2 - R

Average price and demand trends throughout the day

The data for each state had to be separated and analysed on their own, as initial summary plots of the average of the variables showed that although the general trend of the 3 main variables (price, time of day and demand) were similar for some states, there were still, in some instances, a few notable differences unique to each region. Some examples of this were the difference in average demand in different regions, as well as the spread of the price of electricity in the distinct states. However, there were also notable similarities, such as the times throughout the day where peaks in price and demand occurred. These observations are also relatively intuitive, as it would be expected that each separate state may have different factors influencing the price and demand in the region, such as population size, cost of getting electricity to homes, and other factors. In regards to the common times where the prices and demands peak, it can be explained by the fact that most people use up electricity in the evening and night for lighting, heating, cooling or electronic devices, which corresponds to increased demands and thus increased prices at these times.

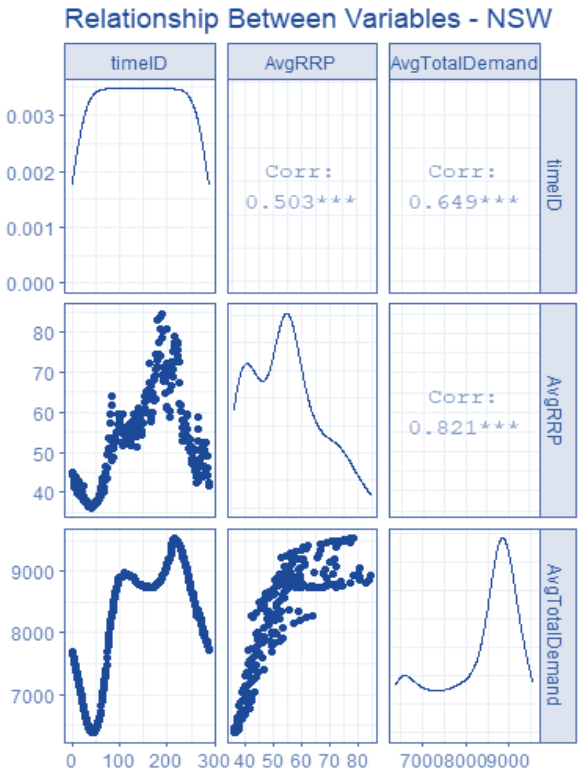

Figure 4.2.1
Relationships between the variables for time throughout the day (timeID) and the average price (AvgRRP) and average demand (AvgTotalDemand) of electricity at each particular time in the day for NSW. Graphs for other regions can be found in Figure 4.2.2.

Observations of potential predictors

From Figure 4.2.1 and 4.2.2, we can observe that there is a similar trend in the peaks of electricity price and demand throughout all the regions. The main peak for all states occurs around 4-5 pm, with the exception of Tasmania, which has two main peaks, one around 8-9 am and the second at approximately 4-5 pm, similar to the other states.

From this, it is fairly obvious that the time of day has a significant impact on the average price. The price vs time of day graph is also polynomial in nature, taking a rough form between a quadratic and quartic shape, with different regions having a shape closer to one of these more than the other. Due to these reasons, we decided to include the time of day ($timeID$) as a predictor in our regression models, noting that a quadratic or other polynomial linear transformation may be needed to get a better fit of the data.

It can also be seen that the average demand curve follows a similar pattern to the average price throughout the day, with two main peaks around the same time as the peaks in price.

This seems to suggest that as demand for electricity increases, the price also increases. The price vs demand curve confirms this notion as well, as it shows the same result. This concept is easy to understand and accept, as the effect of demand on price can be seen in almost every part of the economy. This was one of the main reasons for which we included the demand as a predictor variable in our regression models to predict the price. It was also noted that higher values of demand resulted in an increased rate of price increase, suggesting, similar to the $timeID$ predictor, that a transformation of the demand predictor may be required to create an efficient regression model.

Chapter 5
Analysis and Results

1. Predicting price using average price dataset

These models involve the dataset containing the average values of price and demand throughout the day, at 5-minute intervals.

Initial models

The initial model, as stated above in section 3.6, was:

$$(1a) \qquad avgPrice \sim avgDemand + time$$

The results of this regression on the data can be seen in Figure 3.6.1. However, after running a best subset selection test and stepwise regression, both forward and backward selection, the results showed that a model with just the Average Demand variable as a predictor was superior to this initial model for some states, as this model had lower AIC values.

$$(1b) \qquad avgPrice \sim avgDemand$$

Using model $(1b)$, the regions NSW, SA, and VIC all had lower AIC values and similar values to model $(1a)$. The p values for the time predictor in these regions were also very high, suggesting that the time of day predictor was not significant for the data in these states. For QLD and TAS, the initial model had lower AIC values, and the p values of the time of day predictor were low. However, the adjusted R^2 and RMSE values across both models were relatively similar for all states. The preferred model for each state is summarised below:

$Model\ (1a)$: NSW, SA and VIC

$Model\ (1b)$: QLD and TAS

Models with transformed predictors

After studying the bivariate relationships between the different variables in this dataset, it was noted that the time of day and average price variables may have a quadratic relationship, while the average demand may have an exponential or quadratic relationship with average price. A few more sample models were created, making use of this information.

$(2a)$ $avgPrice \sim avgDemand + avgDemand^2 + time + time^2$
$(2b)$ $avgPrice \sim avgDemand + time + time^2$
$(2c)$ $avgPrice \sim avgDemand + avgDemand^2 + time$
$(2d)$ $avgPrice \sim avgDemand + avgDemand^2$
$(2e)$ $avgPrice \sim time + time^2$

Stepwise forward and backward selection was used to determine the best model for each region, given the possibility of utilising the predictor variables $avgDemand$, $time$, $avgDemand^2$ and $time^2$. The best models for each state are summarised in the table below.

Region	Model	Adjusted R^2	RMSE
NSW	$price \sim avgDemand + avgDemand^2 + time + time^2$ $(2a)$	0.72	6.20
QLD	$price \sim avgDemand + avgDemand^2 + time + time^2$ $(2a)$	0.61	11.55
SA	$price \sim avgDemand + time + time^2$ $(2b)$	0.65	12.88
TAS	$price \sim avgDemand + time + time^2$ $(2b)$	0.66	4.16
VIC	$price \sim avgDemand + demand^2 + time$ $(2c)$	0.81	5.72

Figure 5.1.1 - *Adjusted R^2 and RMSE values for models with transformed predictors*

KEY:
$price$ – $RRP\ estimate$
$avgDemand$ – $Avg\ Electricity\ demand\ at\ the\ time$
$time$ – $time\ of\ day\ (number\ between\ 0\ and\ 287, representing\ 5\ min\ intervals\ starting\ from\ 12\ am)$

Models with transformed price and predictors

The best overall model for each state was determined by the best fit, according to the highest adjusted R^2 value. The best model for each transformation of price was determined by stepwise forward and backward selection regression.

As can be seen in Figure 5.1.2 containing the adjusted R^2 values for each regression, the inverse transformation seems to produce the best results for every state. For this reason, the inverse transformation of price was chosen for the model for every region.

Final models of price for each region

NSW:

$$1/price \sim avgDemand + time + time^2$$

QLD:

$$1/price \sim avgDemand + time + time^2$$

SA:

$$1/price \sim avgDemand + avgDemand^2 + time + time^2$$

TAS:

$$1/price \sim avgDemand + time$$

VIC:

$$1/price \sim avgDemand + avgDemand^2 + time^2$$

The summary statistics for all models are shown in figures 5.1.3 A – E.

2. Other MLR models

The dataset was split into a training and a testing set where the test set size was 33% of the whole dataset. Then, a Simple Linear Regression model was built out of feature TOTALDEMAND and target RRP. This is mainly to be used as a benchmarking model to compare against other models. Using root mean squared error as the evaluator, we got 57.6.

Then, a multiple linear regression model following similar steps as above was built. The additional feature columns were hour_group, month_group, week_group, weekends. Following similar procedures to above, 58.04 was obtained as the root mean squared error.

So, the simple linear regression model seemed to perform better than the multiple linear regression. Meaning, the additional columns worsened the performance of the model.

3. SARIMAX model

The next model tried was the SARIMAX model. SARIMAX is a complex model and is time-consuming to fit to all the 10 years' worth of data. So, the dataset was grouped so that each observation corresponds to one hour from any day. All data before the year 2015 was also removed to reduce the time taken to compute the model and also because the data from 2015 onwards is more relevant for time moving forward. The dataset was also then grouped by hour, day, month by averaging RRP and TOTALDEMAND. In the end, there were 8784 observations.

Then, some data visualization, including a line plot of RRP with time on x-axis revealed a very interesting trend. Each day has two valleys and two peaks, one peak at 8 am and another peak at 6 pm. So, this concluded that our data is seasonal. The line plot also confirmed that the dataset is not stationary. Meaning, constant mean and variance are violated.

In order to make the dataset stationary, each observation was subtracted by 24 observations before as there is a pattern every 24 hours. Then, a hypothesis test using adfuller was

conducted to make sure the dataset is stationary. It rejected the null hypothesis that the dataset is not stationary. Hereafter, the Auto-correlation plot and partial auto-correlation plot was drawn to determine the hyperparameter terms AR, differencing and MA terms. Through inspection, the AR term is 1, MA term is 1 and differencing is also 1. The seasonality hyperparameter was set to be 24 due to the pattern present in the data every 24 hours. Finally, the model was fit on the training set and predicted on the test set. The root mean squared error attained is 59.5.

Statespace Model Results

Dep. Variable:	RRP	**No. Observations:**	720
Model:	SARIMAX(1, 1, 1)x(1, 1, 1, 24)	**Log Likelihood**	-3269.732
Date:	Sat, 21 Nov 2020	**AIC**	6549.464
Time:	14:20:13	**BIC**	6572.183
Sample:	07-02-2019	**HQIC**	6558.249
	- 07-31-2019		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.4115	0.014	30.195	0.000	0.385	0.438
ma.L1	-1.0000	20.763	-0.048	0.962	-41.694	39.694
ar.S.L24	-0.4159	0.009	-46.243	0.000	-0.433	-0.398
ma.S.L24	-0.9994	4.162	-0.240	0.810	-9.157	7.158
sigma2	607.4993	1.22e+04	0.050	0.960	-2.33e+04	2.45e+04

Ljung-Box (Q):	65.18	**Jarque-Bera (JB):**	77925.95
Prob(Q):	0.01	**Prob(JB):**	0.00
Heteroskedasticity (H):	0.02	**Skew:**	0.84
Prob(H) (two-sided):	0.00	**Kurtosis:**	54.85

Figure 5.3.1-Summary Statistics for SARIMAX model

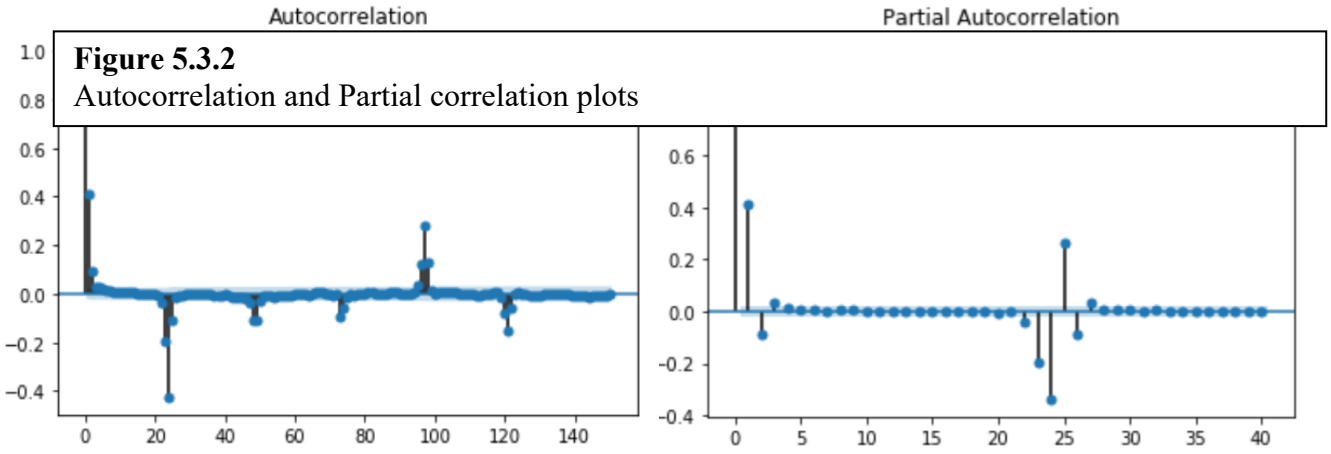

Figure 5.3.2
Autocorrelation and Partial correlation plots

4. Optimisation using Gurobi

This is an example of a profit maximization problem where the objective is to determine if at the current point in time the battery should be charging or discharging electricity. By choosing to charge a cost is incurred upon the battery, whereas by choosing to discharge revenue is gained. After forecasting the price one day into the future, 288 data points need to be processed in order to maximize the profit function. The problem statement is framed as a Linear Programming optimization (LP problem).

The optimizer also takes into account the frequency of decision making. If instructed to make a decision every 5 minutes, given a cost matrix that consists of 288 observations (as there 288 five-minute intervals in a day), with some constraints upon the decision variables along with other structural constraints and objective function to maximize or minimize, the optimizer outputs the chosen decisions for 288 intervals along with profit made over the day. Nevertheless, if the decision making frequency was every one hour, since the output would only consist of 24 values (as there are 24 one hour interval within a day to make a decision for), the task is to choose a reliable function to map 288 observations into 24 values that represents the general distribution of the data.

The assumption we imposed upon the model was that splitting the observation into N partitions (N = (24 hrs)/(frequency_of_update hrs)) and finding the averages of the N partitions to make N observations is a reliable representation of the data to make optimization for.

For the following variables, Price_h is defined as the cost matrix - the dispatch price of electricity in MWh.

Decision Variables

- $DispatchGen[h]$ $(units: MW, Lower\ Bound\ =\ -1, Upper\ Bound = 1)$
- $DischargeBattery2[h]$ $(units: MW\ Lower\ Bound = 0, Upper\ Bound = 1)$
- $ChargeBattery2[h]$ $(units: MW, Lower\ Bound =\ 0, Upper\ Bound = 0)$
- $EnergyStorage[h]$ $(units: MWh, Lower\ Bound = 0, Upper\ Bound\ =\ 2\ MWh)$
- $DispatchCost[h]$ $(units: \$AU, definition:\ -1 * ChargeBattery2[h] * Price[h] * (1/60) * frequencyOfUpdate)$
- $DispatchRevenue[h]$ $(units: \$AU, definition: (DischargeBattery[h] * Price[h] * (1/60) * frequencyOfUpdate)$

Objective function

- $Dispatch\_Revenue\_h$ $(units: \$AU, Lower\ Bound\ =\ -infinity, Upper\ Bound\ =\ infinity, definition)$
  - $Discharge2[h] * Price[h]\ *\ (1/60) * (frequencyOfUpdate)$
- $Dispatch\_Cost\_h$ $(units: \$AU, Lower\ Bound\ =\ -infinity, Upper\ Bound\ =\ infinity, definition)$
  - $Charge2[h] * Price[h]\ *\ (1/60) * (frequencyOfUpdate)$
- This constraint is profit.
  - $\sum_{h\epsilon H}\ Dispatch\_Revenue\_h\ -\ Dispatch\_Cost\_h\ for\ all\ h\ in\ H.$

- Where $ChargeBattery2[h]$ is negative or 0.

Structural Constraints

- $EnergyInStorage[0]\ =\ 0$
- $EnergyInStorage[h+1]\ =\ EnergyInStorage[h]\ -\ 0.81 * ChargeBattery[h] * Price[h]\ * (1/60) * (frequencyOfUpdate)\ - DischargeBattery[h] * Price[h] * (1/60) * frequencyOfUpdate$

GUROBI'S PROFIT ESTIMATION FOR NSW FROM MODEL 5 MINUTE UPDATES

Output from Model	Gurobi's profit estimation

sarimaxNSW	1.729174600e+02
MLRNSW	9.805204000e+01
LRNSW	1.165710400e+02

GUROBI'S PROFIT ESTIMATION FOR NSW FROM MODEL 1 HOUR UPDATES

Output from Model	Gurobi's profit estimation
sarimaxNSW	1.480758600e+02
MLRNSW	9.805173973e+01
LRNSW	1.086252926e+02

Chapter 6
Discussion

Initially, our primary objective was to project dispatch prices 1 day into the future and use these projections to identify 2 boundaries (say b+ and b-) of bidding prices of which if b+ were surpassed, it would provide us with an indication to sell. On the other hand, if the spot price falls below b- buying electricity would be a better job. Coming up with these boundaries (b+ and b-) would be likely through formulating an optimization equation for profit and incorporating methods in likeness to maximum likelihood to observe which pair of boundaries of our boundary space generates the highest profit for the current dataset. However, this methodology of optimisation is static and thereby does not suit well to the needed criteria of being able to adapt to incoming data. To account for this nature of input stream of data, Gurobi was used to form a Linear Programming optimization task with constraints.

Our metric to determine which model is better is to inspect the optimal objective value from the output of the optimization. This informs the maximum profit that could be achieved if manoeuvred the battery decisions with optimal recommended suggestions by the linear programming optimizer. By this criterion, we can infer that the SARIMAX model is able to generate the maximum profit out of the three. However, from our initial root mean squared error analysis, the simple linear regression was the best. This may seem confusing at first as to how the best model is not able to generate the best profit, but the optimization and predictions are completely disjoint entities. This means it does not matter if we our model outputs lots of profits. What matters in terms of forecasting data into the future is ensuring that the accuracy of prediction is to a reliable standard. Other methodologies of optimization should be considered during the analysis of dispatch price fluctuations to determine which makes the highest profit given a set amount of predicted data. Where our analyses rely only upon the usage of Linear programming optimization framework.

Furthermore, in analysing how the variation of the frequency of decision making affects profits, it is as expected that when we move from a longer time interval to a shorter time span, which means high frequency updates to low frequency updates, we notice an increase in the profits produced. This is primarily due to the increase in the degrees of freedom the model can now operates, meaning the optimizer now has more possibilities to compute and compare with to determine maximum profit. Initially, when in the framework of 1-hour updates, our profits were \$148.08, \$98.05174 (5.d.p), \$108.63 per day from the three models developed. When moved to the 5-minute framework, we see these profits increase to

$172.90, $98.05204 (5.d.p), and $116.57. Hence, a suggestion can be made that if the computational resources can meet the requirements of processing data every minutes, then certainly it should be considered as an objective of maximizing profits.

A technical issue with Gurobi outputs at the moment is that since the Chargebattery2[h] and the Dischargebatter2[h] variables holds value close to 0, querying the optimal values of the decision variables from the optimal model after the inbuilt hyperparameter tuning results in values that are either 0.0, -1.0 or 1.0. Where a negative value indicates charging, and a positive value indicates discharging. Although, the optimization produces the estimate profits for the upcoming day, the decision it is making to achieve this outcome due to technical difficulties is not entirely clear. There needs to be a method researched on increasing the significant figures of the returning values to determine exactly what state the battery should be in.

From the development of the models, we noticed that the SARIMAX model proved too complex to fit 10 years' worth of data. So, the dataset had to be reduced to 1 year and each observation had to go from every 5 minute to every 1 hour. The two price peaks observed each day were very instrumental in making the model stationary and improving its performance with the SARIMAX model. One peak was at 8 am and the other peak was at 6 pm. Both are expected as people tend to use a lot of electricity before they leave to office or school and again people use more electricity when they're back home after work.

Another interesting discovery is that, when utilising all the data, simple linear regression was a better fit that a multiple linear regression model when target RRP is regressed with predictor $TOTALDEMAND$ This proved that the categorical columns created were making the model perform worse and further care and inspection should be given to properly segregate each class in these new features or remove these features if they are redundant.

Changing the AR, MA and differencing terms during hyperparameter surprisingly did not affect the models performance. The summary of the model showed this in figure 5.3.1. However, setting too high a value can cause the model to crash by making it run for too long.

Improving the model:
Due to the computational restrictions posed by processing limits of our physical hardware, running the models over long periods of time is not ideal. However, if it met the necessary requirement, performing estimates would not have to be performed to leverage data for computational capabilities from our initial data, which would have allowed for our model's learning capability to be enhanced.

Moreover, we could have used Long Short-Term Memory Networks (LSTMs) Neural Network in formulating our models to use the latest technology at hand and perhaps make additional gains upon predictions. They provide support for input data

Chapter 7
Conclusion and Further Issues

Conclusion:
What are the main conclusions? What are your recommendations for the "client"? What further analysis could be done in the future?

In conclusion, based on our analysis the simple linear regression model is better than the SARIMAX in a forecasting framework. Furthermore, we have identified that increasing the frequency of decision making for the battery leads to an increase in the computational possibilities an optimization engine can work on, which consequently allows for the possibility to increase profits. This is clear as an inspection of this phenomenon pointed out gains of up $24.83 for one day in the NSW region. Thus, investing in improving predictions to unravel subtle trends and patterns alongside the development of an optimization engine is a profitable prospect in the long run. Though this report focuses on optimizing in a Linear Programming framework using Gurobi, other optimization frameworks like Quadratic programming (QP), and Quadratically constrained programming should also be considered to analyse data from different constraints.

Issues faced:
One of our problems was having too much information, we had 288 readings every day from each of the regions. Performing analysis over the entire year or the entirety of 11 years becomes substantially hard due to the processing limit imposed by the hardware at hand. This produced the issue of observing correlation plots with ideal lags mere impossible. A memory Error was generated in attempting to compute the ARIMA model with lag 288 and SARIMA model with ideal lags. Hence, when trying to extrapolate yearly trends, we decided to use the average of every day – providing us 365-366 observations every year for each region. This transformation allowed us to visualize yearly trends - giving us a comprehension of the importance of the 1-year lag incorporated by the Auto regressive models.

Another initial bottleneck during the data exploration was having only 2 years' worth of data to derive useful insights. It was problematic in terms of training the models. However, after using python and R, there was no cap on the amount of information that was available to us. Though processing them was always slow.

References

AEMC. (n.d.). *About Us*. [online] Available at: https://www.aemc.gov.au/about-us [Accessed 21 Nov. 2020].

AER (2018). *Wholesale electricity market performance report*. [online] *aer.gov.au*, pp.1–78. Available at: https://www.aer.gov.au/system/files/Wholesale%20electricity%20market%20performance%20report%20-%20December%202018_0.pdf.

Aurecon. (2018). *Hornsdale Power Reserve Impact Study - Battery storage's role in a sustainable energy future*. [online] Available at: https://www.aurecongroup.com/markets/energy/hornsdale-power-reserve-impact-study.

energy.gov.au (2017). *National Electricity Market (NEM) | energy.gov.au*. [online] Energy.gov.au. Available at: https://www.energy.gov.au/government-priorities/energy-markets/national-electricity-market-nem.

Hyndman, R.H. and Athanasopoulos, G. (2016). *Forecasting: Principles and Practice*. [online] Otexts.com. Available at: https://otexts.com/fpp2/.

Karakatsani, N.V. and Bunn, D.W. (2008). Forecasting electricity prices: The impact of fundamentals and time-varying coefficients. *International Journal of Forecasting*, 24(4), pp.764–785.

NERA (2007). *The Wholesale Electricity Market in Australia A report to the Australian Energy Market Commission*. [online] *aemc.gov.au*, pp.1–20. Available at: https://www.aemc.gov.au/sites/default/files/content/a6470fbb-bbeb-41d8-948e-a5aff27ee09f/The-Wholesale-Elec-Market-in-Aust-NERA.pdf.

Saâdaoui, F. (2016). A seasonal feedforward neural network to forecast electricity prices. *Neural Computing and Applications*, 28(4), pp.835–847.

Tesla. (2019). *Autobidder*. [online] Available at: https://www.tesla.com/en_AU/support/autobidder.

Ventosa, M., Baíllo, Á., Ramos, A. and Rivier, M. (2005). Electricity market modelling trends. *Energy Policy*, [online] 33(7), pp.897–913. Available at: https://www.sciencedirect.com/science/article/pii/S0301421503003161 [Accessed 14 Nov. 2019].

Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, [online] 30(4), pp.1030–1081.

Louis Brailsford, Professor Andrew Stock,Petra StockGreg Bourne . (2018). POWERING PROGRESS: STATES RENEWABLE ENERGY RACE . Available: https://www.climatecouncil.org.au/wp-content/uploads/2018/10/States-Renewable-Energy-Report.pdf. Last accessed 21/11/2020.

Appendix

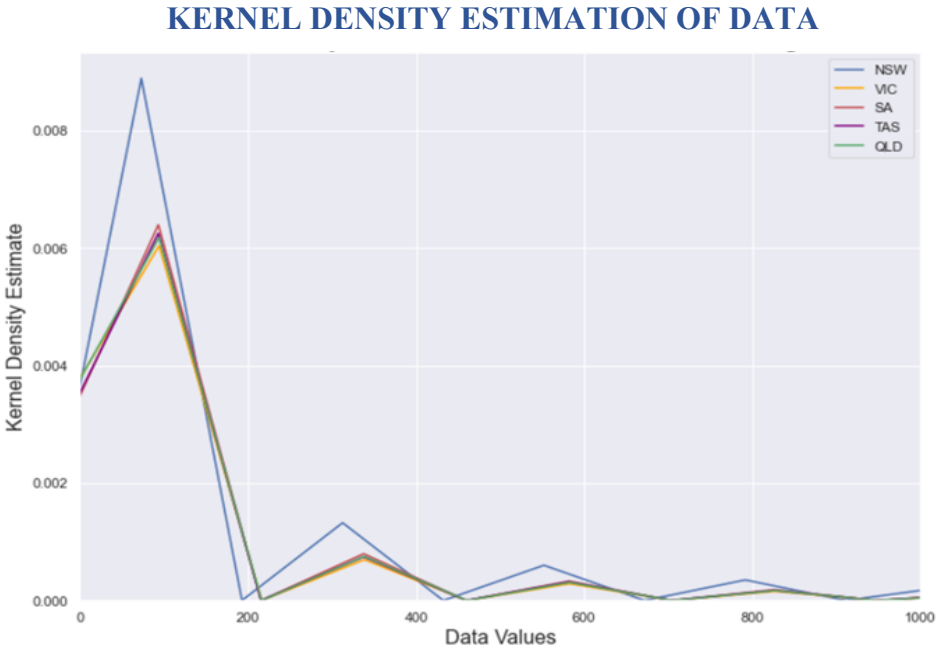

Figure 4.1.3
Kernel density estimation of data for all regions

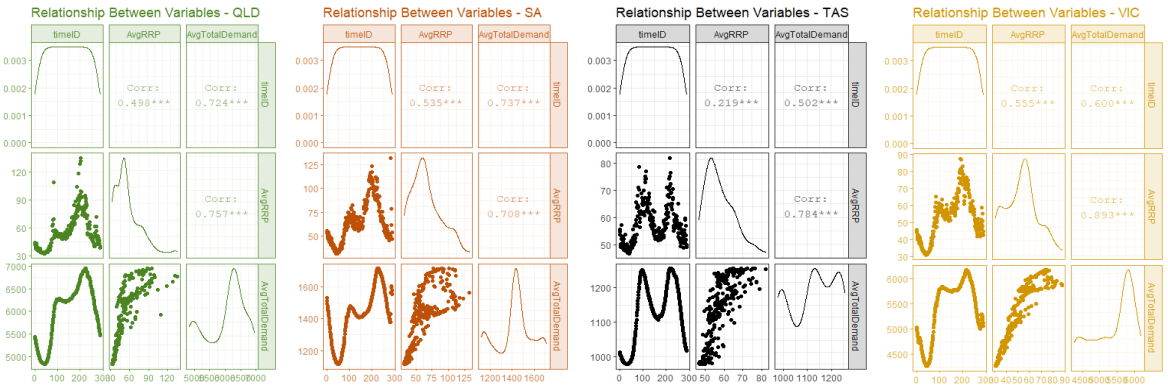

Figure 4.2.2
Relationships between the variables for time throughout the day (timeID) and the average price (AvgRRP) and average demand (AvgTotalDemand) of electricity at each particular time in the day for QLD, SA, TAS and VIC (left to right)

NSW

Transformation on 'price'	Model	Adjusted R^2
-	$price \sim avgDemand + avgDemand^2 + time + time^2$	0.72
log	$\log(price) \sim avgDemand^2 + time + time^2$	0.79
Square root	$sqrt(price) \sim avgDemand + avgDemand^2 + time + time^2$	0.76
Inverse	$1/price \sim avgDemand + time + time^2$	0.85

QLD

Transformation on 'price'	Model	Adjusted R^2
-	$price \sim avgDemand + avgDemand^2 + time + time^2$	0.61
log	$\log(price) \sim avgDemand + avgDemand^2 + time + time^2$	0.73
Square root	$sqrt(price) \sim avgDemand + avgDemand^2 + time + time^2$	0.68
Inverse	$1/price \sim avgDemand + time + time^2$	0.82

SA

Transformation on 'price'	Model	Adjusted R^2
-	$price \sim avgDemand + time + time^2$	0.65
log	$\log(price) \sim avgDemand + avgDemand^2 + time + time^2$	0.78
Square root	$sqrt(price) \sim avgDemand + avgDemand^2 + time + time^2$	0.72
Inverse	$1/price \sim avgDemand + avgDemand^2 + time + time^2$	0.86

TAS

Transformation on 'price'	Model	Adjusted R^2
-	$price \sim avgDemand + time + time^2$	0.66
log	$\log(price) \sim avgDemand + time + time^2$	0.69
Square root	$sqrt(price) \sim avgDemand + time + time^2$	0.67
Inverse	$1/price \sim avgDemand + time$	0.70

VIC

Transformation on 'price'	Model	Adjusted R^2
-	$price \sim avgDemand + avgDemand^2 + time$	0.81
log	$\log(price) \sim avgDemand^2 + time$	0.89
Square root	$sqrt(price) \sim avgDemand + time + avgDemand^2$	0.85
Inverse	$1/price \sim avgDemand + avgDemand^2 + time^2$	0.92

Figure 5.1.2
The Adjusted R^2 values of the best regressions (according to stepwise regression) for all transformations of the $price$ variable. Green values are the highest for the region, suggesting that the corresponding model is the best for that region

```
Call:
lm(formula = invAvgRRP ~ AvgDemand + timeIDsq + timeID, data = AvgNSW)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0045865 -0.0011679  0.0001871  0.0011182  0.0036219

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.935e-02  1.224e-03   40.34  < 2e-16 ***
AvgDemand   -3.278e-03  1.938e-04  -16.92  < 2e-16 ***
timeIDsq     1.596e-07  2.214e-08    7.21 5.09e-12 ***
timeID      -5.022e-05  7.506e-06   -6.69 1.19e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.001633 on 284 degrees of freedom
Multiple R-squared:  0.8514,    Adjusted R-squared:  0.8498
F-statistic: 542.2 on 3 and 284 DF,  p-value: < 2.2e-16
```

5.1.3 A
Summary Statistics for final model of price - NSW

```
Call:
lm(formula = invAvgRRP ~ AvgDemand + timeIDsq + timeID, data = AvgQLD)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0108125 -0.0013223  0.0002477  0.0016649  0.0053802

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.672e-02  2.162e-03  30.862   <2e-16 ***
AvgDemand   -7.604e-03  4.732e-04 -16.068   <2e-16 ***
timeIDsq     9.875e-08  3.440e-08   2.871   0.0044 **
timeID      -2.766e-05  1.209e-05  -2.289   0.0228 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.002495 on 284 degrees of freedom
Multiple R-squared:  0.8175,    Adjusted R-squared:  0.8156
F-statistic: 424.2 on 3 and 284 DF,  p-value: < 2.2e-16
```

5.1.3 B
Summary Statistics for final model of price - QLD

```
Call:
lm(formula = invAvgRRP ~ AvgDemand + AvgDemandsq + timeID + timeIDsq,
    data = AvgSA)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0095958 -0.0014979  0.0002087  0.0017102  0.0039545

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.334e-01  8.150e-03  16.373   <2e-16 ***
AvgDemand   -1.320e-01  1.174e-02 -11.237   <2e-16 ***
AvgDemandsq  3.723e-02  4.165e-03   8.939   <2e-16 ***
timeID      -9.762e-05  6.622e-06 -14.742   <2e-16 ***
timeIDsq     3.337e-07  2.108e-08  15.833   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.002137 on 283 degrees of freedom
Multiple R-squared:  0.8662,    Adjusted R-squared:  0.8643
F-statistic:   458 on 4 and 283 DF,  p-value: < 2.2e-16
```

5.1.3 C
Summary Statistics for final model of price - SA

```
Call:
lm(formula = invAvgRRP ~ AvgDemandsq + timeID, data = AvgTAS)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0035502 -0.0007061  0.0001899  0.0007931  0.0024838

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.876e-02  4.321e-04  66.571  < 2e-16 ***
AvgDemandsq -9.201e-03  3.719e-04 -24.738  < 2e-16 ***
timeID       4.539e-06  9.142e-07   4.965 1.18e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.001124 on 285 degrees of freedom
Multiple R-squared:  0.7045,    Adjusted R-squared:  0.7025
F-statistic: 339.8 on 2 and 285 DF,  p-value: < 2.2e-16
```

5.1.3 D
Summary Statistics for final model of price - TAS

```
Call:
lm(formula = invAvgRRP ~ AvgDemand + AvgDemandsq + timeIDsq,
    data = AvgVIC)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0046735 -0.0009103  0.0002050  0.0010626  0.0030387

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.132e-01  9.594e-03  11.799  < 2e-16 ***
AvgDemand   -2.643e-02  3.714e-03  -7.117 9.05e-12 ***
AvgDemandsq  1.690e-03  3.545e-04   4.768 2.98e-06 ***
timeIDsq    -6.696e-09  4.154e-09  -1.612    0.108
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.001533 on 284 degrees of freedom
Multiple R-squared:  0.9196,    Adjusted R-squared:  0.9188
F-statistic:  1083 on 3 and 284 DF,  p-value: < 2.2e-16
```

5.1.3 E
Summary Statistics for final model of price - VIC