DATA3001 - TEAM 11

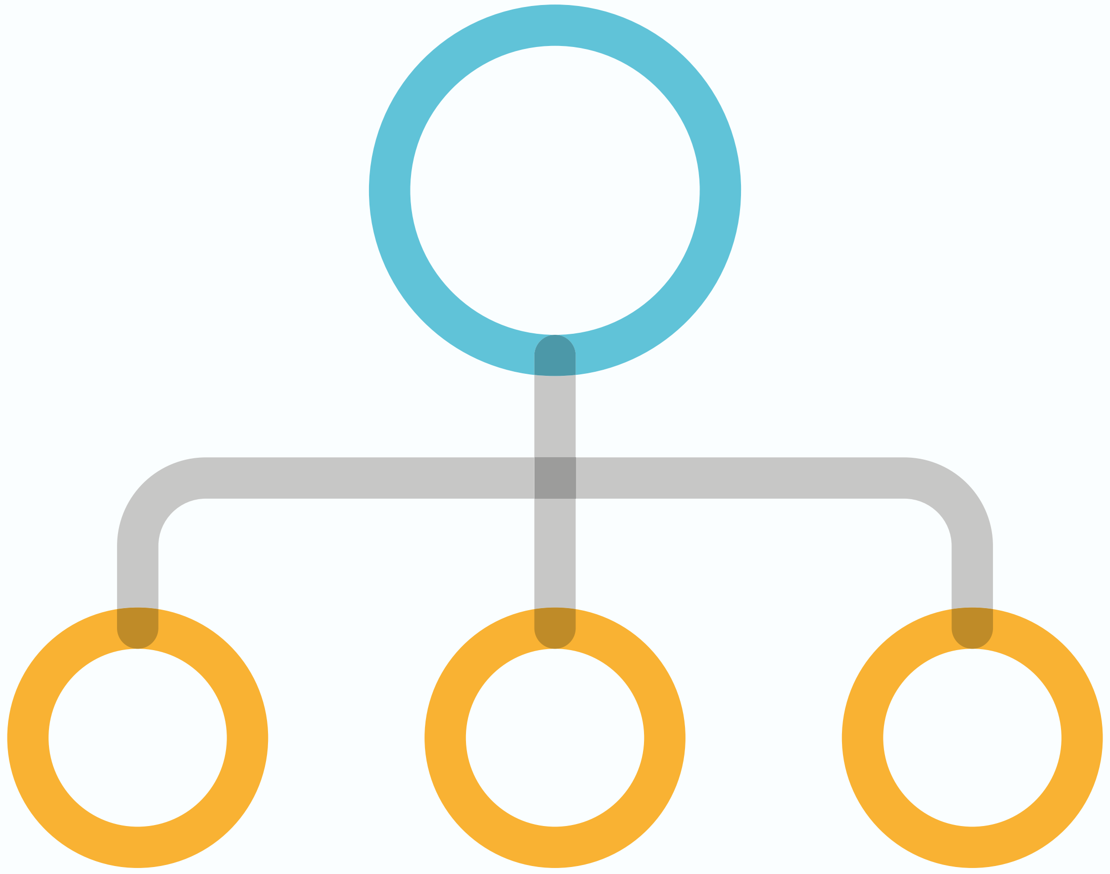
AEMC – How do you train a battery algorithm to make as much money as possible?

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

OVERVIEW

- Context
- Problem Statement
- Our Aim

CHOICES



Context

PROBLEM STATEMENT

Batteries play a role within the energy market. They are provided with data on the price of energy. It would be ideal to maximise profit given this data.

OUR AIM

We wish to develop a dynamic model that forecasts future prices, and develop an effective optimisation model on our forecasts.

OVERVIEW

- Electricity Pricing Models
- Optimisation and Bidding

Electricity Pricing Models

1. STATISTICAL MODELS

- Regression Models
- Autoregressive Time Series Models (AR)
- Autoregressive Time Series Models with Exogenous Factors

2. AI MODELS

- Feedforward Neural Network
- Support Vector Machines (SVM)

3. MULTI-AGENT MODELS

Optimisation and Bidding

METHOD

By implementing a bidding algorithm that can create optimal bids, we can achieve optimal battery performance in the energy market.

RELEVANT WORK

TESLA:

- Tesla's autobidder software currently used in their batteries
- Successfully implemented in The Hornsdale Power Reserve in South Australia

OUR OBJECTIVE

- Develop forecasted dispatch price
- Algorithm formulates a linear function
- Cost coefficients are decided to optimise for objective function (profit)

1. Introduction
2. Literature Review
3. Material and Methods
4. Exploratory Data Analysis
5. Analysis and Results
6. Discussion
7. Conclusion

OVERVIEW

- Software
- Description of Data
- Pre-processing
- Data Cleaning
- Assumptions
- Modelling Methods

Software

python

GUROBI
OPTIMIZATION

Description of Data

Datasets

- dispatchprice_fy2009-2019 (**DispatchPrice**)
- dispatchregionsum_ rrponly- _fy2009-2019 (**DispatchRegionSum**)

Description

- Observations **every 5 mins** from 12am of 1st July 2008 to 12 am 1st July 2019
- **56** features (variables) recorded in total

Main Variables:

Variable Name	Format	Description
SETTLEMENTDATE	Date	market date and time
DISPATCHINTERVAL	Integer	dispatch interval identifier
INTERVENTION	0 or 1	if intervention by organisation is required
RRP	$/Mwh	regional reference price for respective dispatch period
REGIONID	geographic state	indicates the location of respective dispatch
TOTALDEMAND	MWh	electricity demand at the time

Pre-processing

Several different datasets were constructed to form various analysis.

- With **daily averages** over the 10 years of data
- With average prices at different **times during the day**
- With the **different regions** included in the data

Visualising initial
data that was provided …

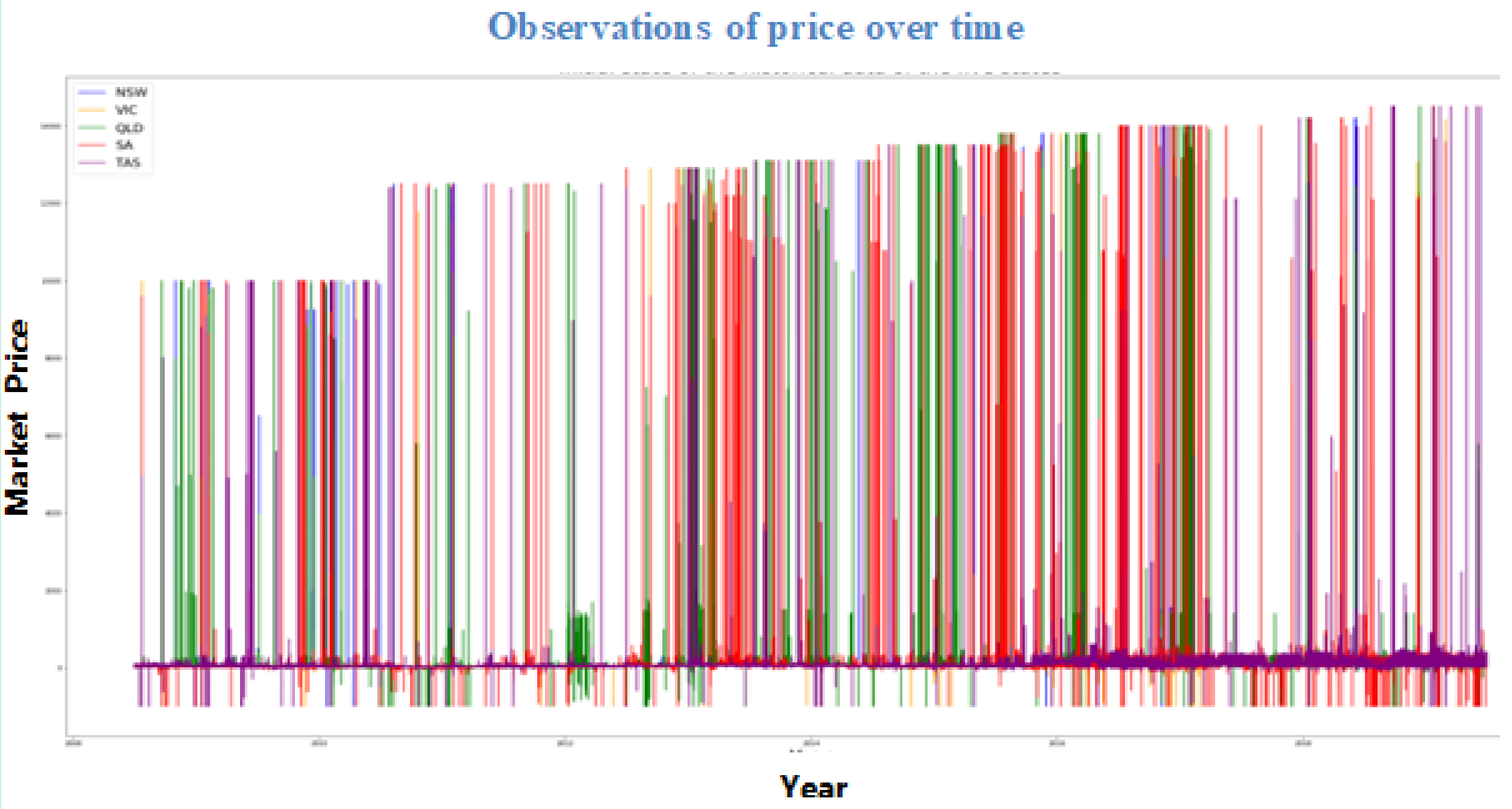


Figure 3.3.1
This figure illustrates the unorganised state of the data prior to our processing

Processed data
with daily averages ...

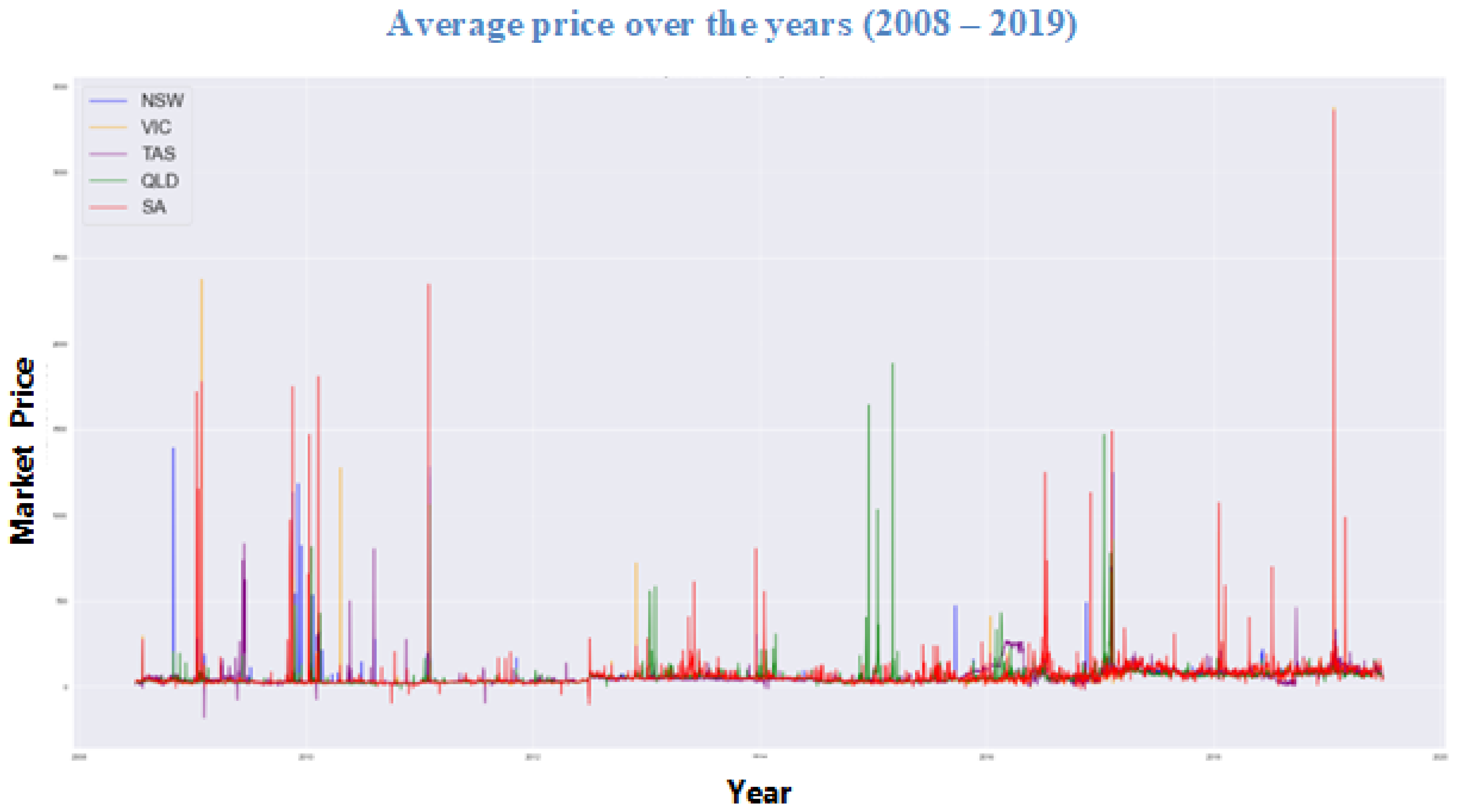
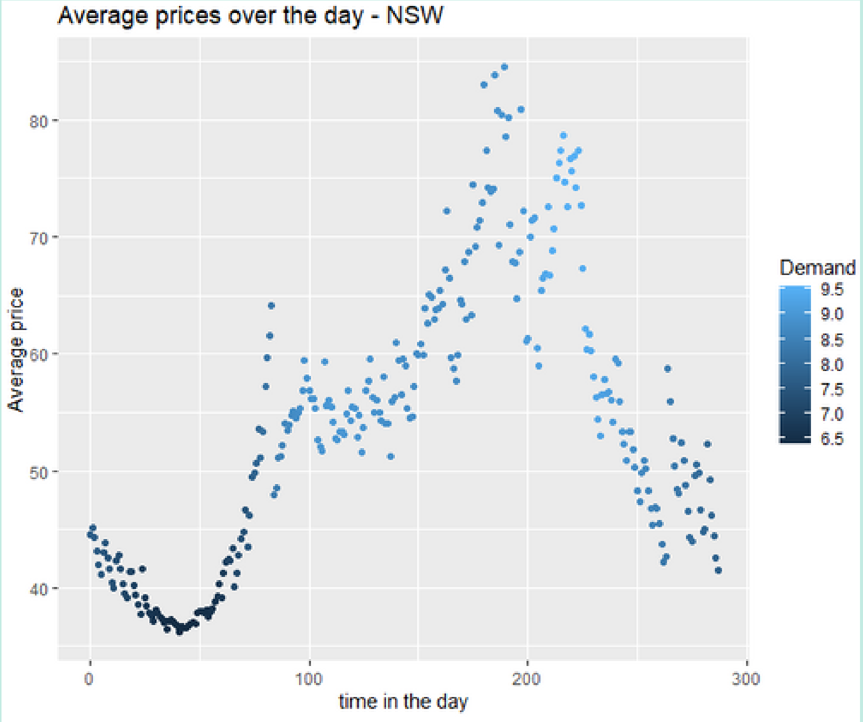
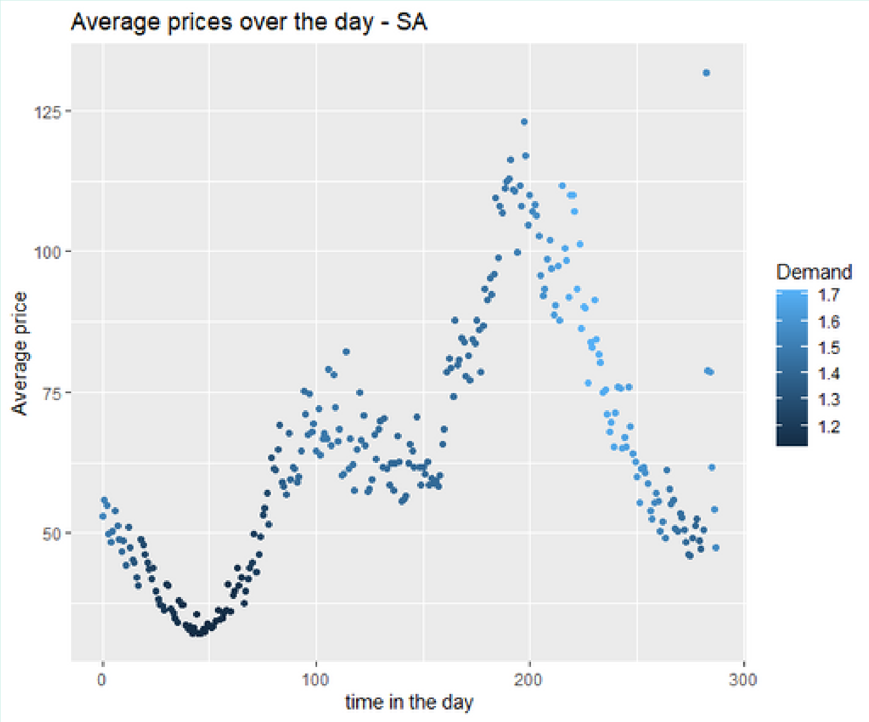
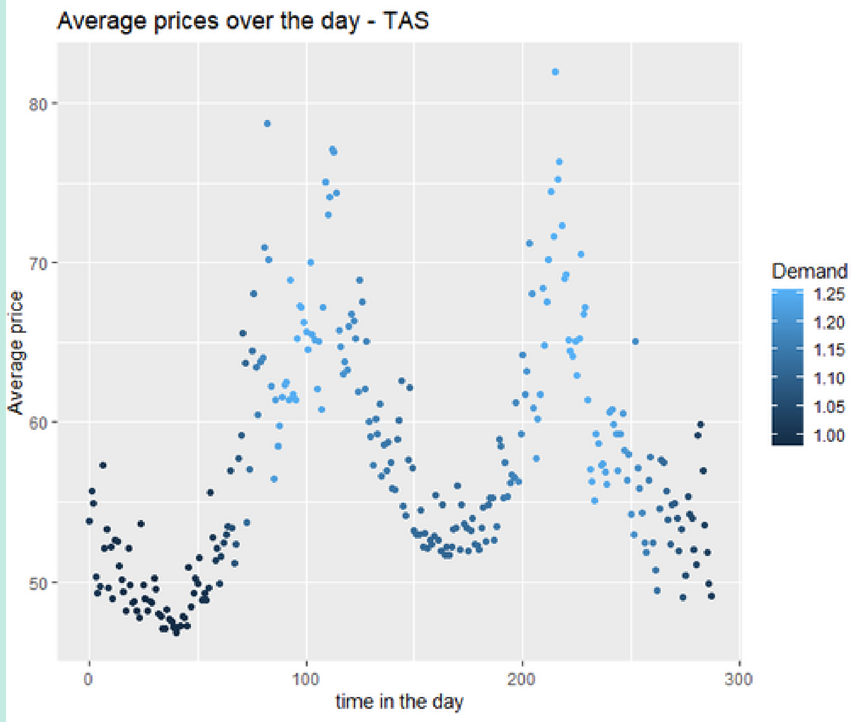


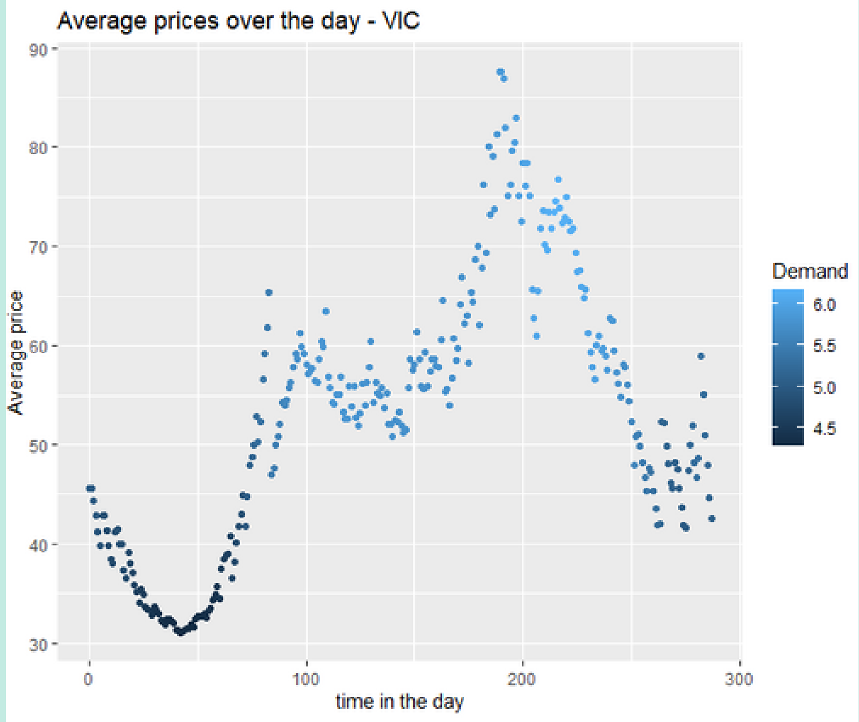
Figure 3.3.2
Visual representation of the change in prices of all 5 regions, from 2008 to 2019

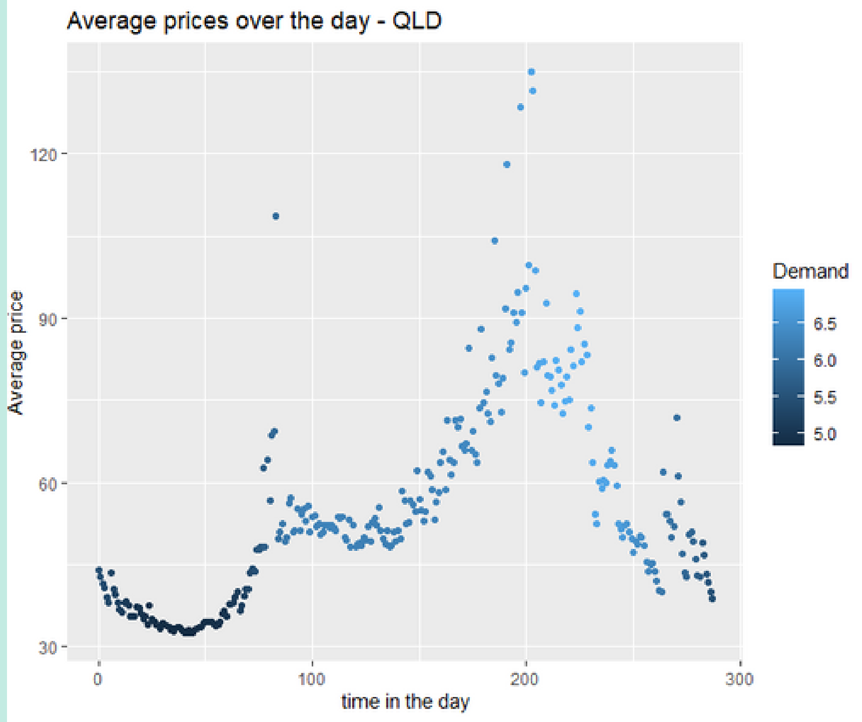
Price trends throughout the day …











Data Cleaning

- The data provided to us from AEMC was already relatively **clean**.
- **R** and **Python** were utilised for initial data cleaning.
- A **unique data point** was observed with the the region listed as 'SNOWY'. This observation was **removed.**
- Any observations with an INTERVENTION value **other than 0** were removed.

Assumptions

1. In reducing high volumes of data and **creating a smaller dataset**, we assume this will not have any significant impact for profit maximisation.
2. In our models utilising **Multiple Linear Regression** (MLR), it is assumed that the conditions required to model the data using MLR are held, and that MLR will provide the best model.
3. An assumption is made that **splitting the dataset into N bags in Gurobi** is a reliable representation of the entire days' worth of data.
4. It is assumed that the objective **profit maximisation function in Gurobi is a linear function** of the dispatch prices that occur at 5 minute intervals.

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. Introduction
2. Literature Review
3. Material and Methods
4. Exploratory Data Analysis
5. Analysis and Results
6. Discussion
7. Conclusion

OVERVIEW

- Trends
- Assessment of Trends

TRENDS

- All the five states have the price peak during both the morning and evening
- 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 hours

TRENDS

- Lower and upper bounds of the price value are between -$1000.00 and $14,500.00 as opposed to being between -$15,000.00 and $15,000.00
- This could be a direct result of the price management system 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 this project

TRENDS

	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

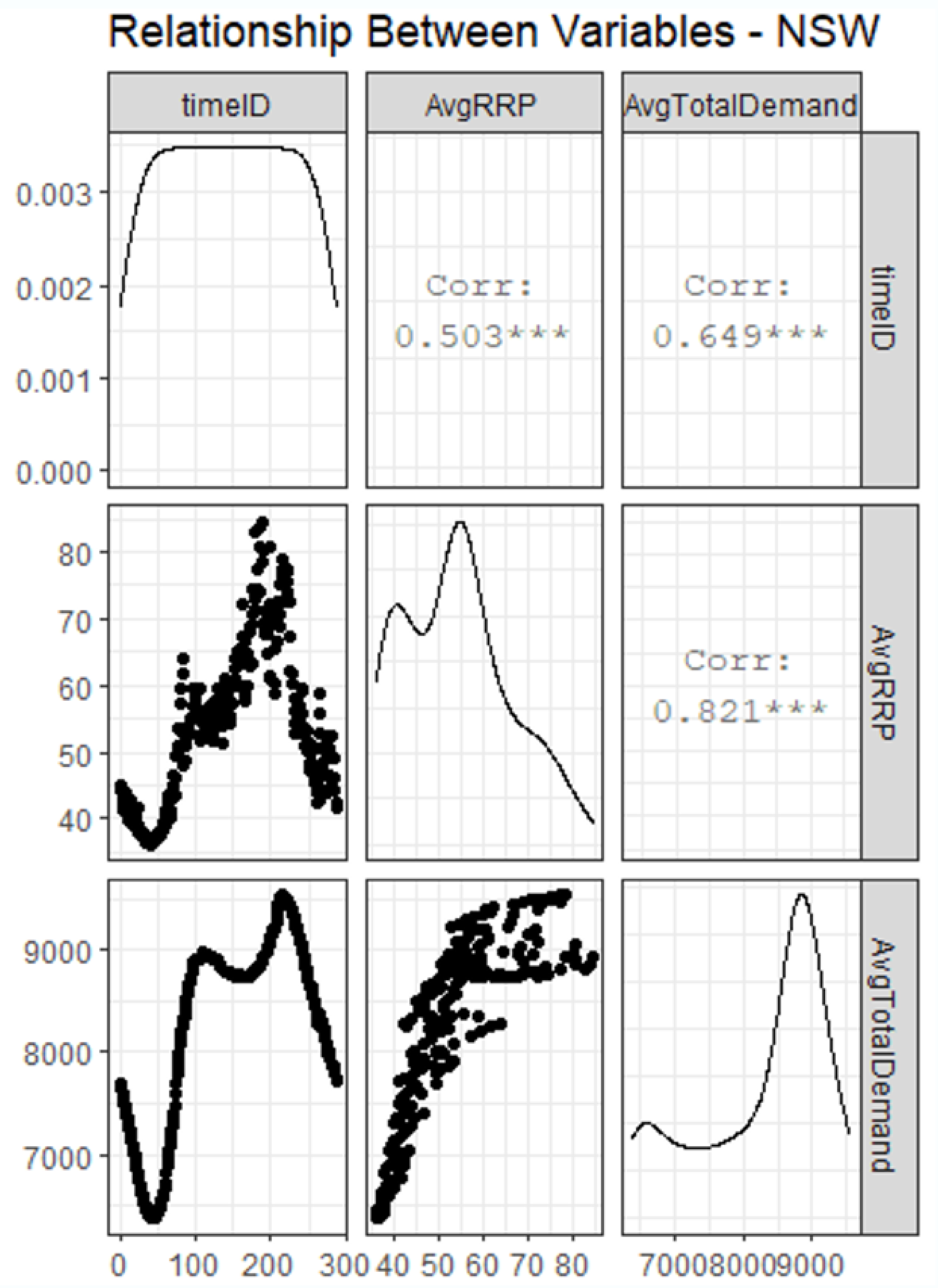
Distribution of data for the different regions reveals roughly a normal distribution, since 50% of the data is concentrated within the range of 27.30 and 60.11.

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.

TRENDS



A few notable differences unique to each region were the difference in average demand in different regions, the spread of price of electricity in the distinct states.

Once the dataset was separated into separate states and analysed, there were also notable similarities, peaks in price and demand occurred at the same time.

Since 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.

It was also noted that higher values of demand resulted in an increased rate of price, suggesting that a transformation of the demand predictor may be required to create an efficient regression model.

OVERVIEW

- Prediction of Prices
- Optimisation Frameworks
- Live Demonstration

As stated earlier, our task was to predict the prices for the upcoming day using the initial dataset.

Three models were created;

- Simple linear regression
- Multiple linear regression
- SARIMAX

Multiple Linear Regression

Strengths

- incorporate an array of different predictor variables

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

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

Upon further inspection…

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

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

With some more inspection...

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

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 ***
```

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

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 *
```

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

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 ***
```

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

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 ***
```

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

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
```

SARIMAX

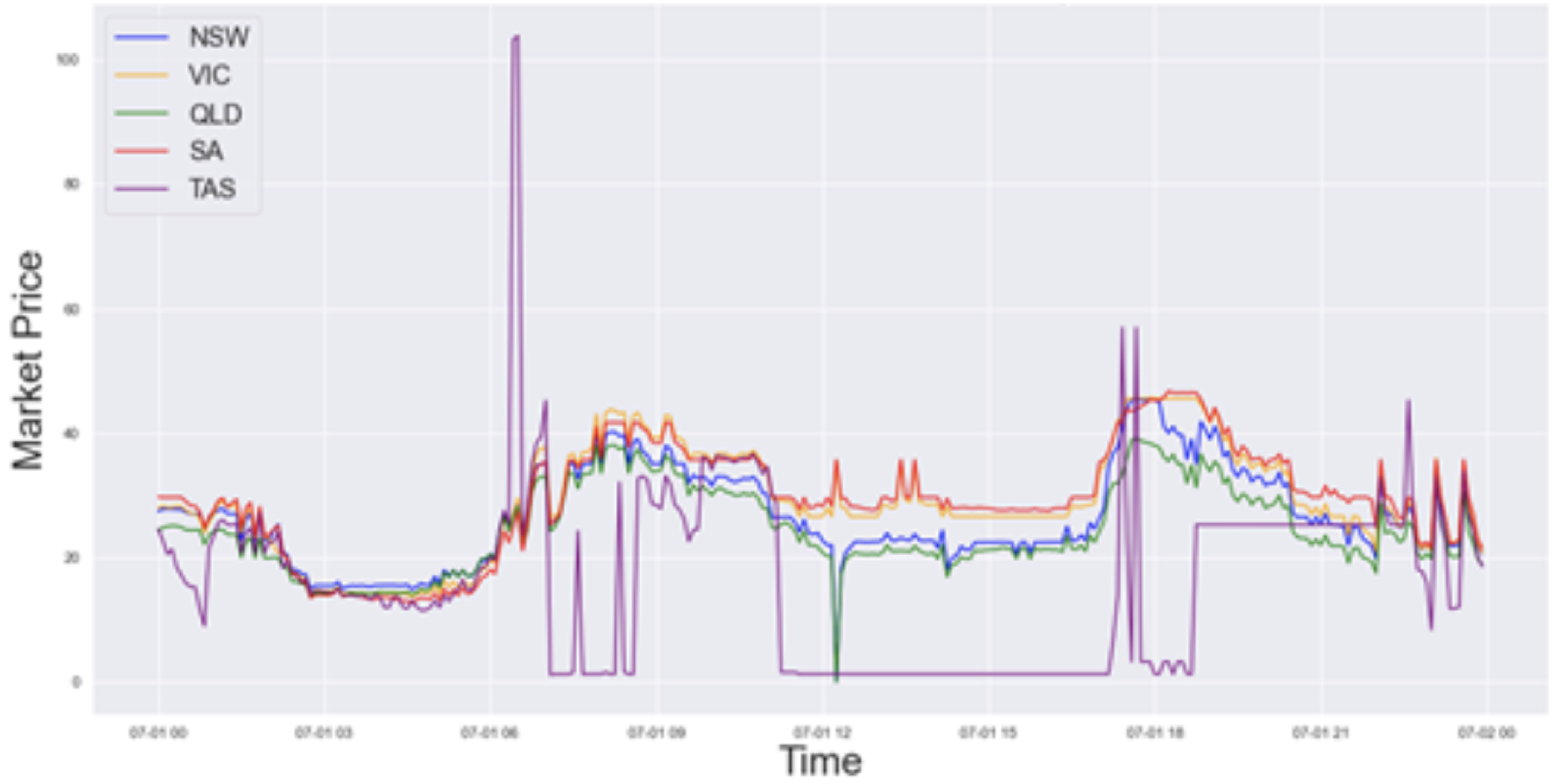
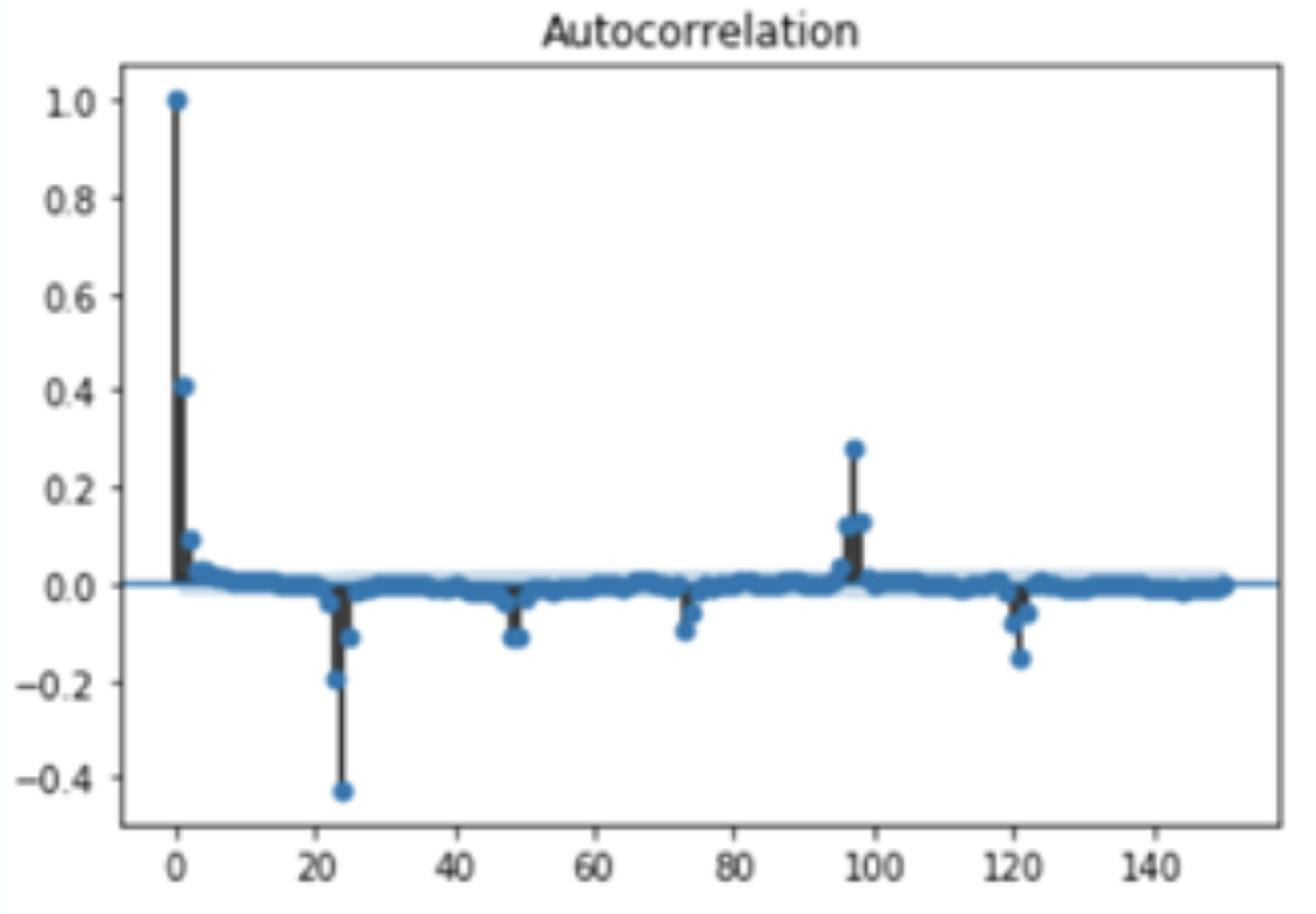
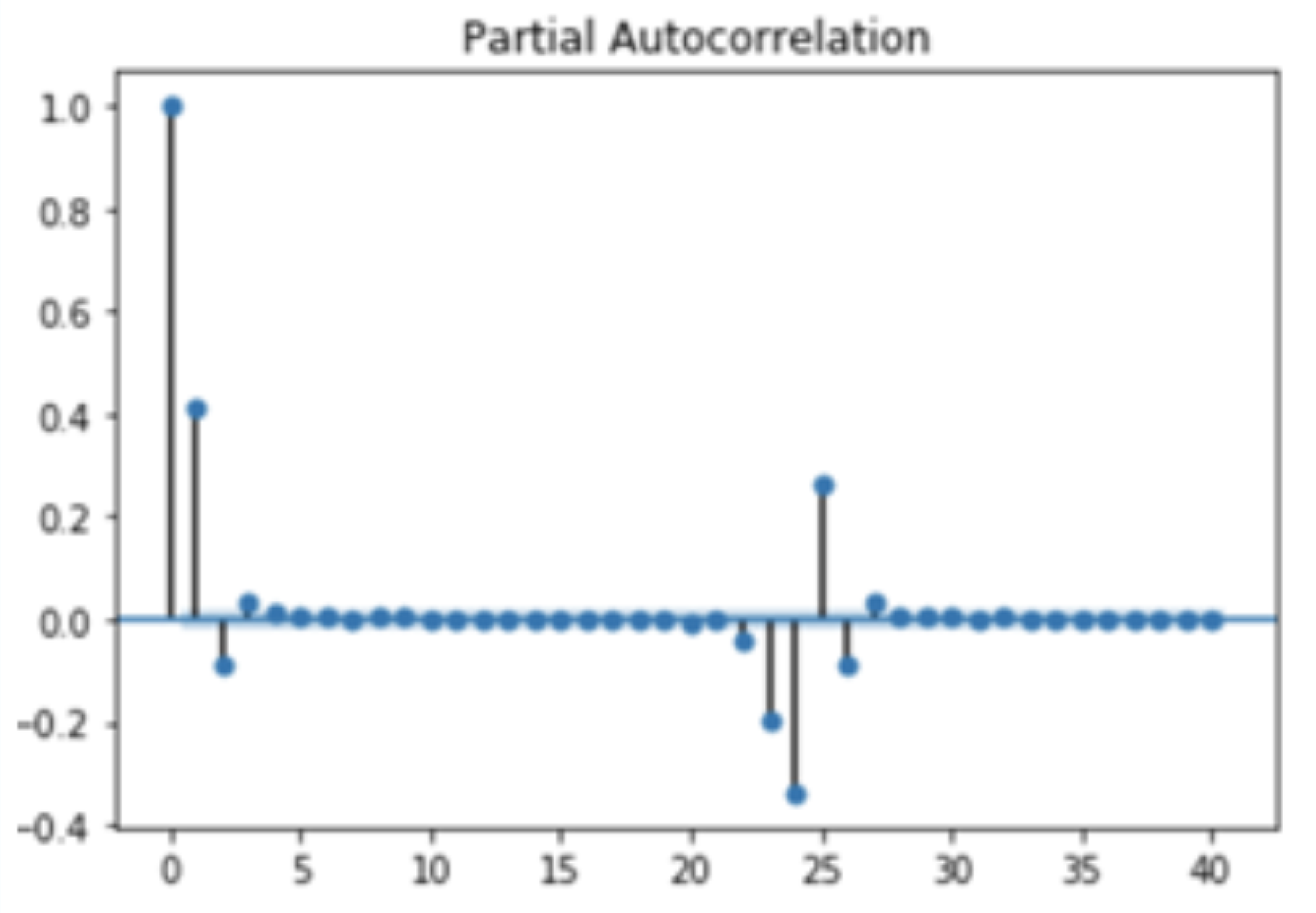


Figure 4.1.1

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



Autocorrelation and
Partial Correlation Plots



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

Summary Statistics for SARIMAX Model

Gurobi - Linear Programming (LP) Optimization

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.
    - $\sum_{h \in H} Dispatch\_Revenue\_h - Dispatch\_Cost\_h\ for\ all\ h\ in\ H.$

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

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

OVERVIEW

- Initial Approach
- New Approach
- Limitations
- Improvements

OUR APPROACH

Initial Approach

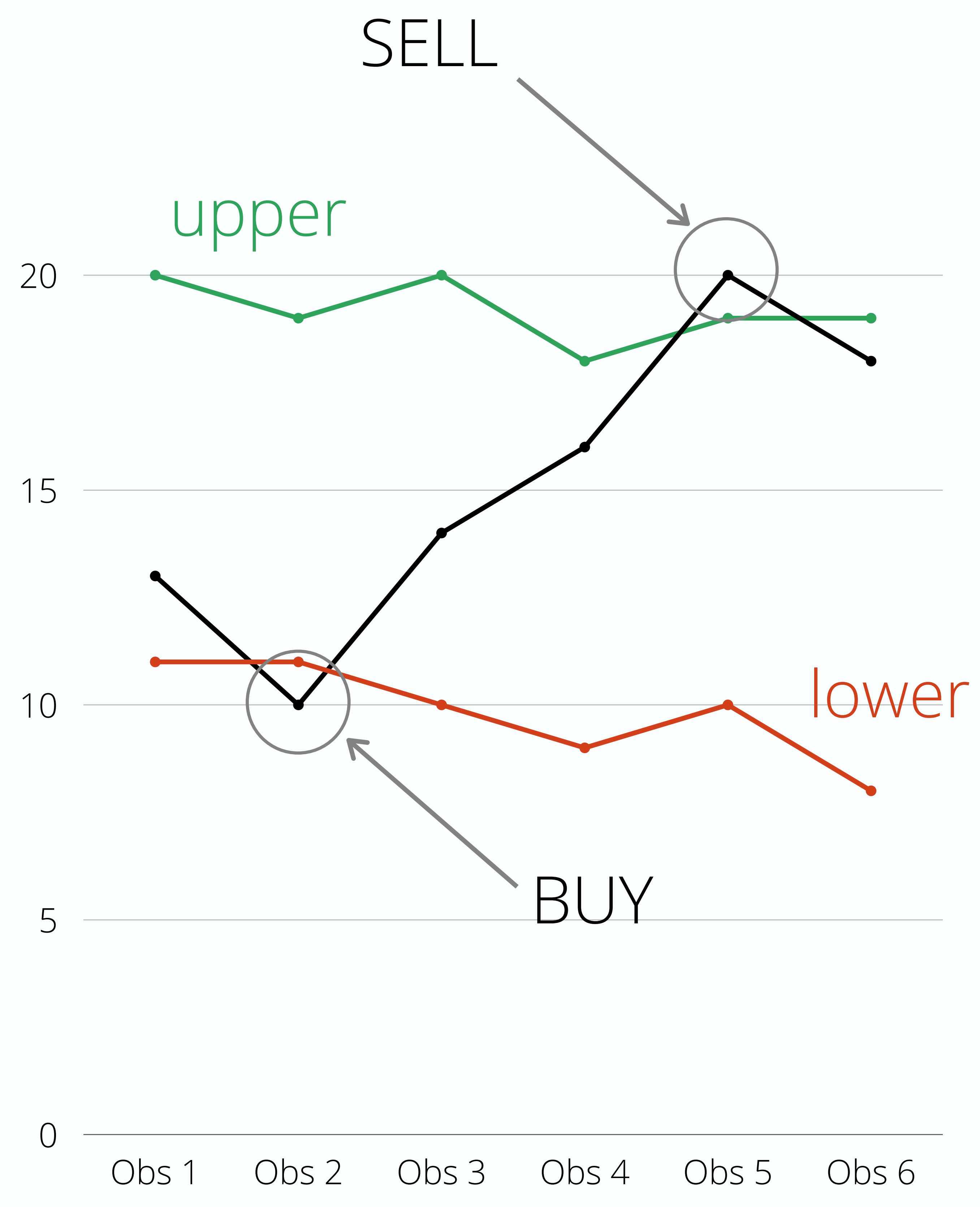
- intervals for our price data
  - if price was above upper bound = SELL
  - if price was below lower bound = BUY
- need to determine the most optimal bounds

Limitations

- static methodology of optimising
- incapable of adapting to incoming stochastic data

New Approach

- form predictions of prices using forecasting models
- utilise linear programming optimisation on forecasted prices



LIMITATIONS

GUROBI

- Chargebattery2[h] and the Dischargebattery2[h] variables hold values close to 0
  - negative value indicates charging
  - positive value indicates discharging

- Querying the optimal values of the decision variables results in values of 0.0, -1.0 or 1.0

- Desirable to increase the significant figures of the returning values to determine exactly what state the battery should be in

SARIMAX

- Unable to handle complete data set
  - Dataset was reduced significantly
  - Frequency of updates reduced drastically

- Two price peaks were observed
  - 8:00 AM and 6:00 PM
  - aided in making model stationary
  - improved performance with SARIMAX

IMPROVEMENTS

Stronger Hardware

Allow for us to run and operate our models over longer periods of time with all data points.

Alternatives

Long Short-Term Memory Networks (LSTMs) may have been used to formulate models.

Various other optimisation frameworks to operate on forecasted prices.

OVERVIEW

- Insights
- Further Issues

INSIGHTS

Prediction of Prices

- Simple linear regression > Multiple linear regression
- Simple linear regression > SARIMAX

Optimisation of Prices

- Linear Programming

Others

- Quadratic Programming
- Quadratically Constrained Programming

stronger hardware and processing power

increase in frequency of decision making

increase in computational possibilities for an optimisation engine

possibility to increase profits

FURTHER ISSUES

Problem

- Excessive amount of information
  - 288 readings per day
  - 365 days per year
  - 11 years
  - 5 regions
- Complex computation and calculations became very taxing on hardware
  - observing correlation plots with lags
  - computation of ARIMA and SARIMA models with lags

Our Solution

- Shift from Excel to R and Python
- Averaging daily prices to reduce number of observations significantly

Most Optimal Solution

- Access to powerful computational resources

Thank You