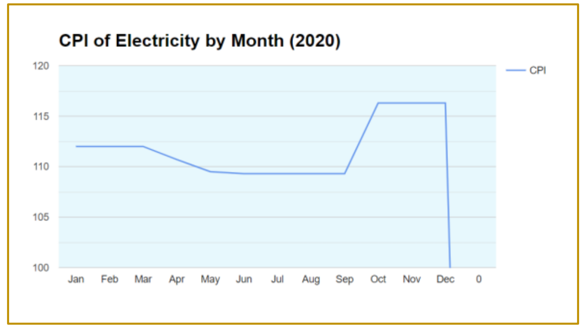
A Regression-Based Forecasting Model for Irish Electricity Price Inflation, Using The Consumer Price Index (CPI) as a Measure

INTRODUCTION

The ability to forecast electricity prices is essential for companies to budget the cost of their electricity use. Moreover, it important for consumers to understand how much electricity will cost. The purpose of this project is to develop a novel forecasting model which accurately predicts the cost of electricity in the Republic of Ireland, using regression-based forecasting techniques. For purposes of simplicity and accuracy, I will be using CPI as a measure of electricity prices.
The forecasting model is based on three attributes: time, demand, commodity prices and production costs.
This model is trained and tested on a data from July 2016 – December 2020.

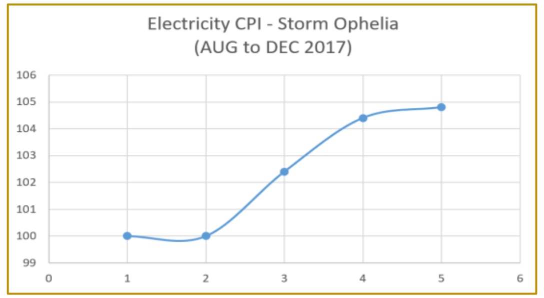
OBSERVATIONS

This forecasting model notes some fundamental patterns in electricity CPI trends. Investigation into CPI data shows that electricity price inflation tends to be cyclical and seasonal.
There is a positive trend in seasonality. Electricity prices are consistently higher in winter months than they are in Summer months.
This is demonstrated by an increase in the CPI of electricity in following figure for the year 2020.



Further investigation into the data reveals that there are other inherent trends which inform electricity prices.

Special circumstances, such as weather events, e.g. heat waves or cold periods can increase the demand for electricity (oftentimes for heating and cooling purposes). The following figure uses an example to demonstrate the impact of *special circumstances* on electricity prices (Hurricane Ophelia 2017).



It can be observed that electricity prices spiked following Hurricane Ophelia (which occurred in October 2017).

Method & Assumptions

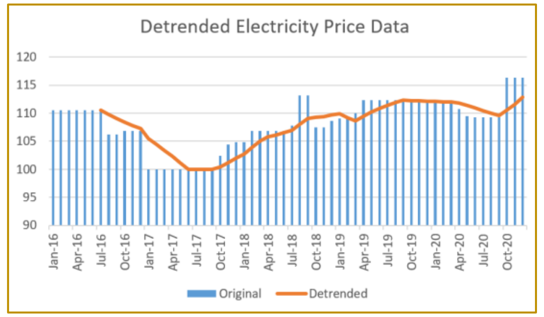
This forecasting model is regression-based, and hence relies on the identification if and investigation of specific attributes or *regressors* which are selected as a result of their relationship with a change in CPI. The method used to fulfil this purpose is as follows:

1. Detrending Data
2. Identifying the effects of *fuel inputs* and *demand* on changes in the electricity price.
3. Quantifying the effects of *time* and *special circumstances* on the electricity price.

Data Sources:
Electricity CPI Data – Central Statistics Office
Fuel Inputs WPI Data – Central Statistics Office
Electricity Demand Data – Sustainable Energy Authority of Ireland

1. Detrending Data

In order to avoid false correlations, we must remove long-term trends in the data that are not relevant to our investigation. For this, we detrend the data using the *moving average* method. The average of each window of values was then calculated for each point in the time series. The moving-average values are then subtracted from the original values in order to get the detrended data.



2. Identify the effects of *fuel inputs* and *demand* on changes in the electricity price.

Firstly, *fuel inputs* are quantified by WPI (Wholesale Price Index) data for oil and gas.

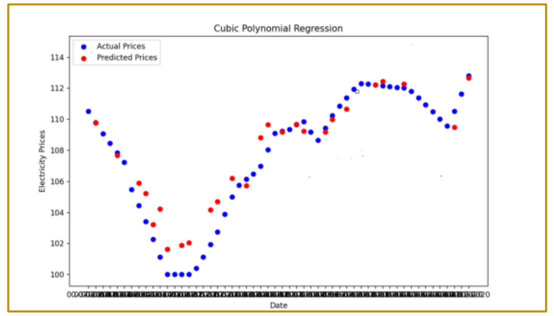
In order to identify the effects of fuel inputs and demand on changes in the electricity price, we use a cubic **polynomial regression** algorithm

This function has the form:

$$y = b_0 + b_1x + b_2x^2 + b_3x^3 + b_4x^4 + b_5x^5$$

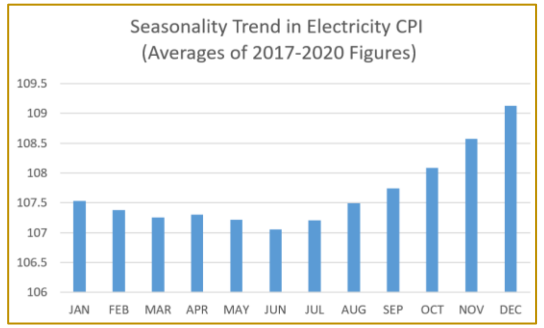
Here, y is the dependent variable, which is the electricity price. X is the independent variable, which denotes the demand and fuel input values. Furthermore, a, b, c and d are coefficients, which are determined by the model during training.

In the following figure, the cubic polynomial's predictions can be seen in comparison to the true electricity prices.



3. Quantifying the effects of *time* and *special circumstances* on the electricity price.

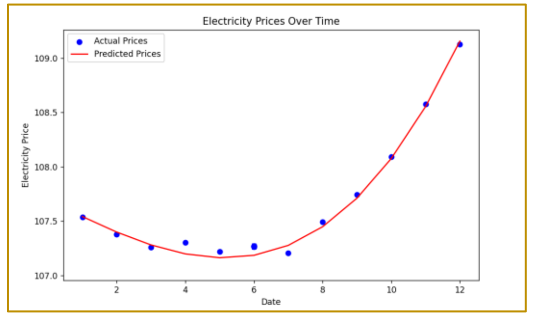
As observed earlier, there is a seasonal trend in electricity prices.
The below figure clearly displays a positive trend in seasonality.



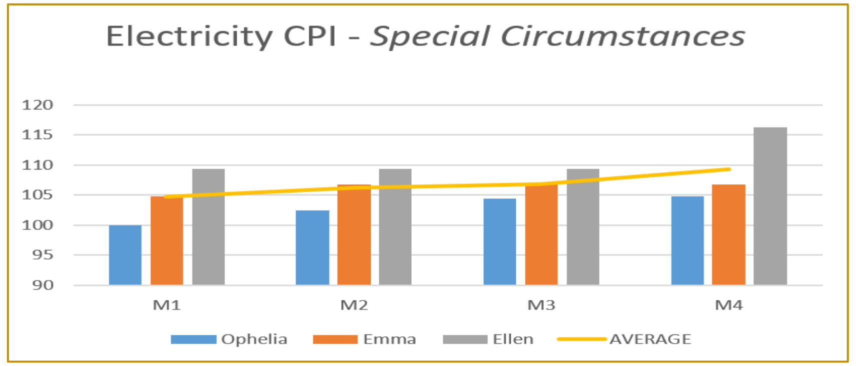
CPI tends to increase in Winter months.
This is seen by plotting the monthly averages of CPI data from 2017-2020.
In order to quantify this phenomenon, we again apply a cubic **polynomial**.

This function has the form: $y = b_0 + b_1x + b_2x^2 + b_3x^3$

In the following figure, the algorithm's predictions can be seen in comparison to the true electricity prices.



As observed earlier, *special circumstances* (weather events) can cause an increase in the price of electricity. In order to investigate this, we take three weather events from within out dataset and examine how the CPI changes as a result. A positive trend following each event can be observed in the following figure.



The exact increase of the CPI following such an event is dependent on the degree to which electricity supply was affected by said event.
In order to determine if electricity CPI increases reliably, we calculate the percentage increases of the Electricity CPI and form three different groups (one for each event) and perform **One-way ANOVA Regression**.

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_1 k$$
$$H_A: \mu_1 \neq \mu_2 \neq \mu_3 = \mu_1 k$$

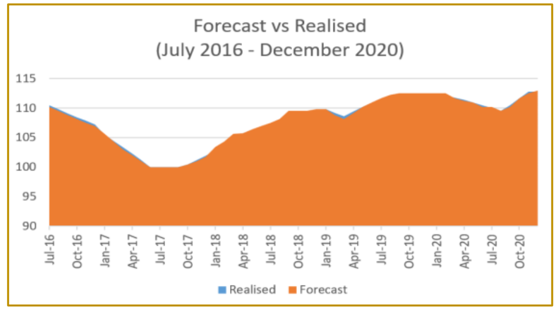
The observed P value is less than 0.05. Hence, we reject the null hypothesis that *there is no significant difference in the percentage increase between the groups*. The electricity price increase following such an event cannot be determined without the consideration of external factors.

Conclusion

The detailed regression-based forecasting model is implemented in R Studio, the statistical programming framework.
Electricity CPI data from July 2016 - December 2020 was taken as the data set. In order to reduce overfitting and acquire more accurate evaluation metrics for the model, **K-fold cross validation** was used in the initial training and testing of the dataset. The data-set was separated into 10 folds, then trained 10 times, each time using a different fold as the test dataset and the remainder of folds a the training datasets.

The following graph demonstrates the model's performance on the same data set (July 2016 – December 2020).

The returned **Mean Absolute Percentage Error (MAPE)** is 6.1%.
Furthermore, an **R-Squared** value of 0.84 was returned.



Although the model is evaluated to be reasonably accurate, it would benefit from being tested on other datasets and through the consideration of additional regressors. These measures will be investigated in the future will be investigated in the future.