Electricity usage and the COVID-19 pandemic

Nathaniel Bloomfield

October 17, 2020

1 Aims

Comparing electricity usage patterns between years gives us an idea of how COVID-19 may have affected energy usage, but these changes in usage could also be due to other factors, such as differences in weather. Building a model based on previous years of electricity usage gives us a way to take these differences into account and compare the usage we'd expect given what we've observed previously and the usage that we are currently seeing during the COVID-19 pandemic. This will provide insights into how COVID-19 has affected the economy.

2 Data

To build a model we assembled a half hour electricity usage dataset that spanned from the start of 2015 to the 14th of October, 2020. Finding data with the required historical timeseries as well as being current to only the last couple of days was a challenge, but we were able to find temperature and solar radiation data and link these to the electricity usage data.Our electricity data was obtained from the Australian Energy Market Operator (AEMO) website, and provided total electricity demand for each state over half hour time periods.

Given we have only state level aggregates for electricity usage, we found that we obtained better results with temperature data across each state, in addition to data from the capital cities. The Integrated Surface Database was able to supply hourly temperature data from weather stations to achieve this through the `worldmet` R package, but these stations had large numbers of missing values. We picked stations that had on average one record every six hours, and then linearly interpolated gaps of at-most six hours. Larger gaps of at-most a month were filled by taking a moving average individually for each hour of the day, and stations with larger data gaps were excluded. To match the electricity data, we linearly interpolated our hourly temperature data back to half-hourly values. The coverage of the stations overlayed with population density for NSW and VIC are shown in Figure 2.

We also obtained gridded daily solar radiation data from SILO, and extracted the values at each of the weather stations to provide reasonable coverage of each state. To approximate solar radiation at a half-hourly level, we used times for sunrise, noon and sunset from the `suncalc` R package based on the centroids of the weather station coordinates to first create an estimate of solar intensity throughout the day using a piece-wise sine curve based on these values. We then multiplied the daily solar radiation values at each weather station by this solar intensity approximation to obtain a half-hourly approximation. This ignores the impact of cloud cover, but data with a higher temporal resolution was not available.
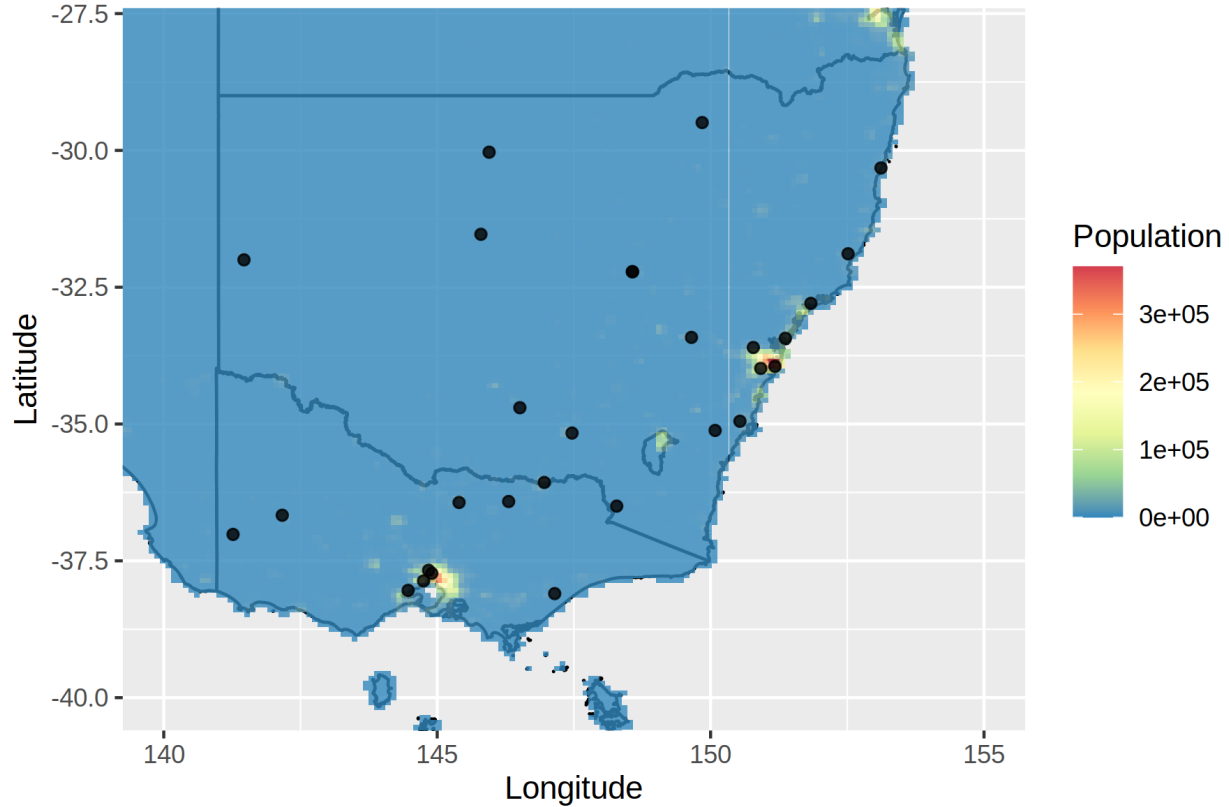
Figure 1: Black dots are the location of weather stations used for building models predicting electricity usage.

3 Model building

We used `xgboost` with the `tidymodels` framework in R to predict electricity usage. We grouped our predictor variables into five sets, and fit these in turn to see how they impacted the quality of the model. Date variables included the time of day, month, day of the week and a boolean variable describing if the day is a weekday. For holidays, we used the `holiday_aus` function from the `tsibble` package, and again used a boolean value to describe them. We also included a variable that described the number of days to the closest holiday, previous or future. To describe temperature, we used the half-hourly temperature values from each weather station in the state as a variable. For solar radiation, we included both our approximate half-hourly solar radiation values at each weather station, as well as the piece-wise sine curve based on timing of sunrise and sunset. We also computed daily and weekly rolling averages of temperature and solar radiation.

The change in model performance as these variables are added is shown in Figure 3 using mean absolute percentage error. Simply using dates to predict electricity usage provides a reasonable model and the error is reduced by 20-30% when all the variables are included, with the largest impact being when temperature is included. Figure 3 shows some comparisons between actual usage and predicted usage for NSW in 2019, and shows that the log of total demand does not vary greatly over the day, staying between 8.7 and 9.3. For further results in this document, 2019 predictions are based on the model trained with data from 2015-2018, and 2020 predictions use a model trained on data from 2015-2019.
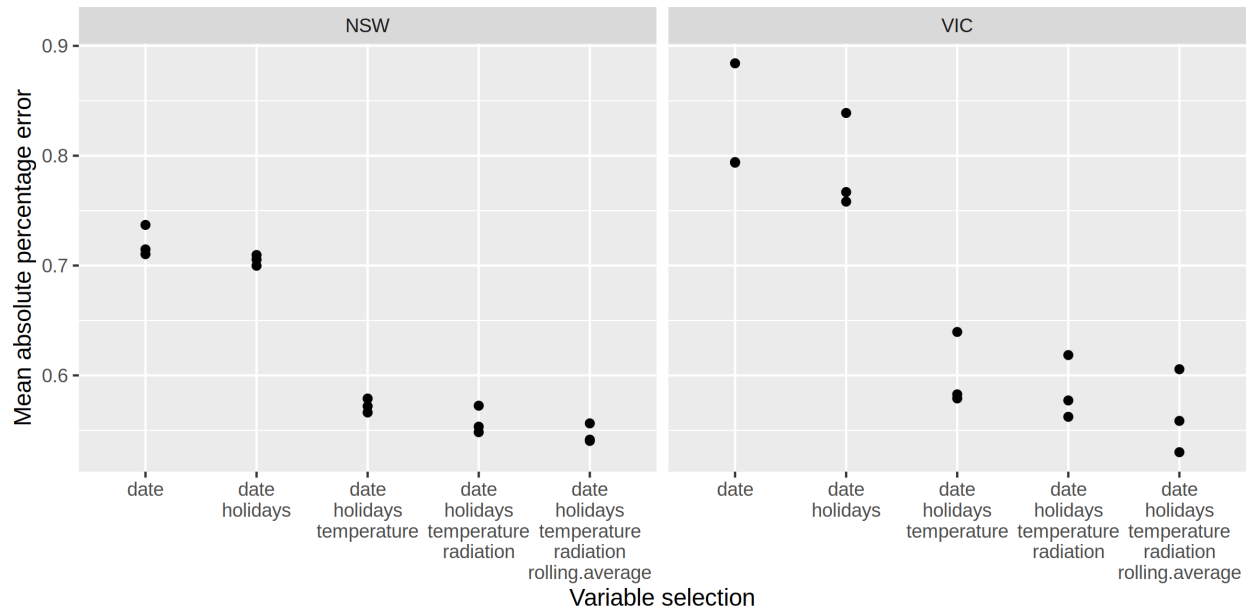
Figure 2: Model performance for VIC and NSW using different variable combinations. The three points are models tested on 2017, 2018 and 2019 data, but trained on all the preceding years.
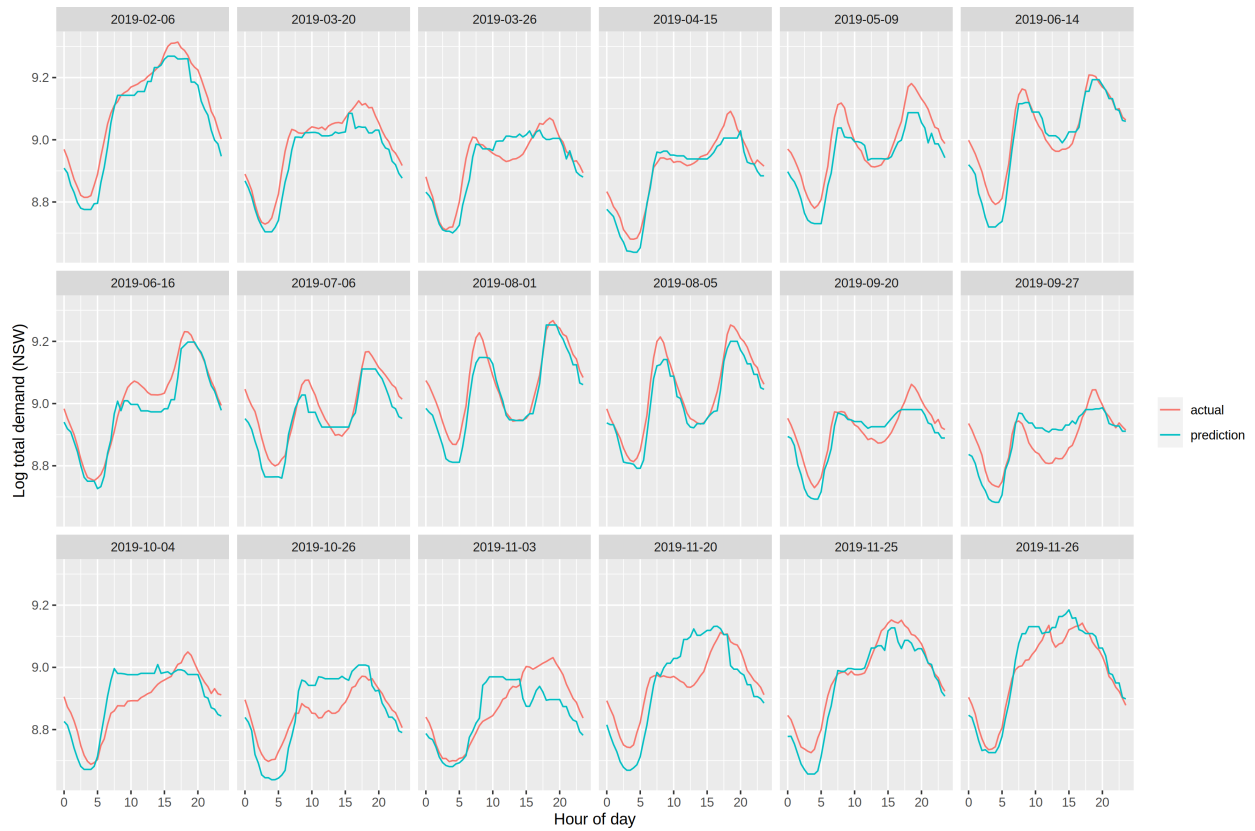

Figure 3: Raw results of model trained on all predictors for NSW.
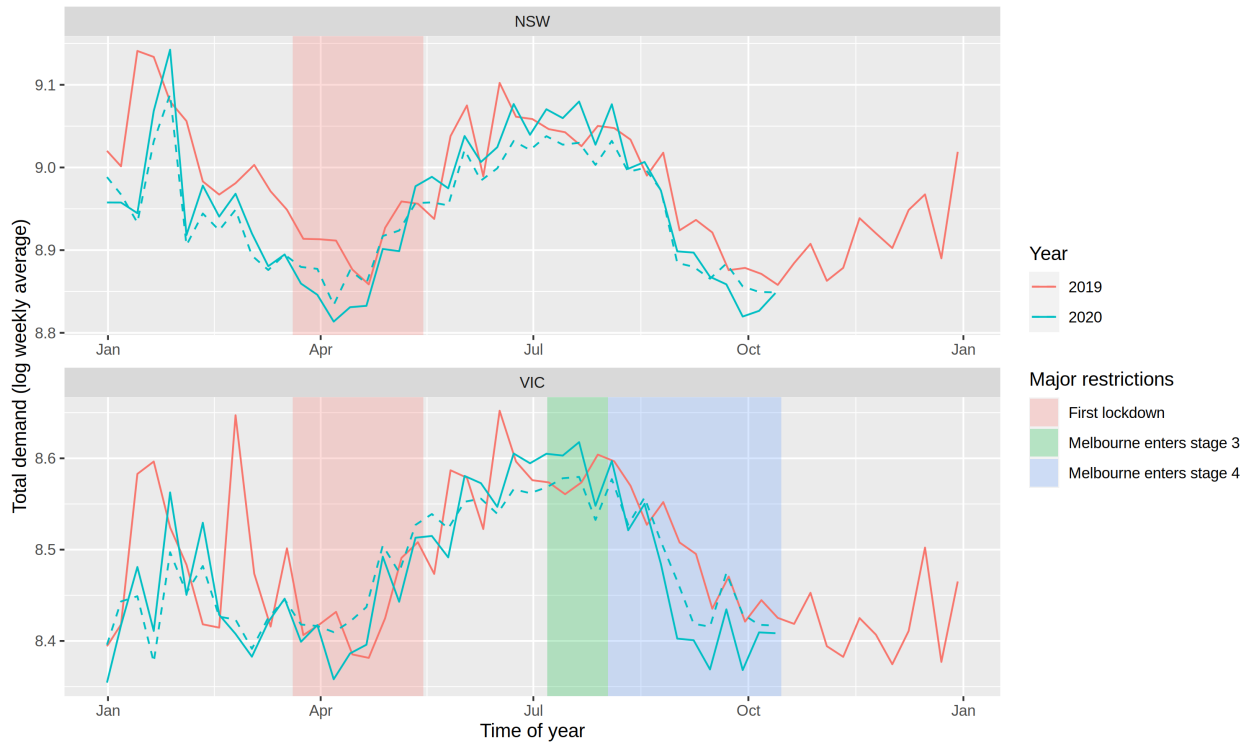
4 Results

Figure 4: Actual demand for 2019 and 2020 with the prediction for 2020 included as the dashed line, overlayed with approximate timing of major restrictions.
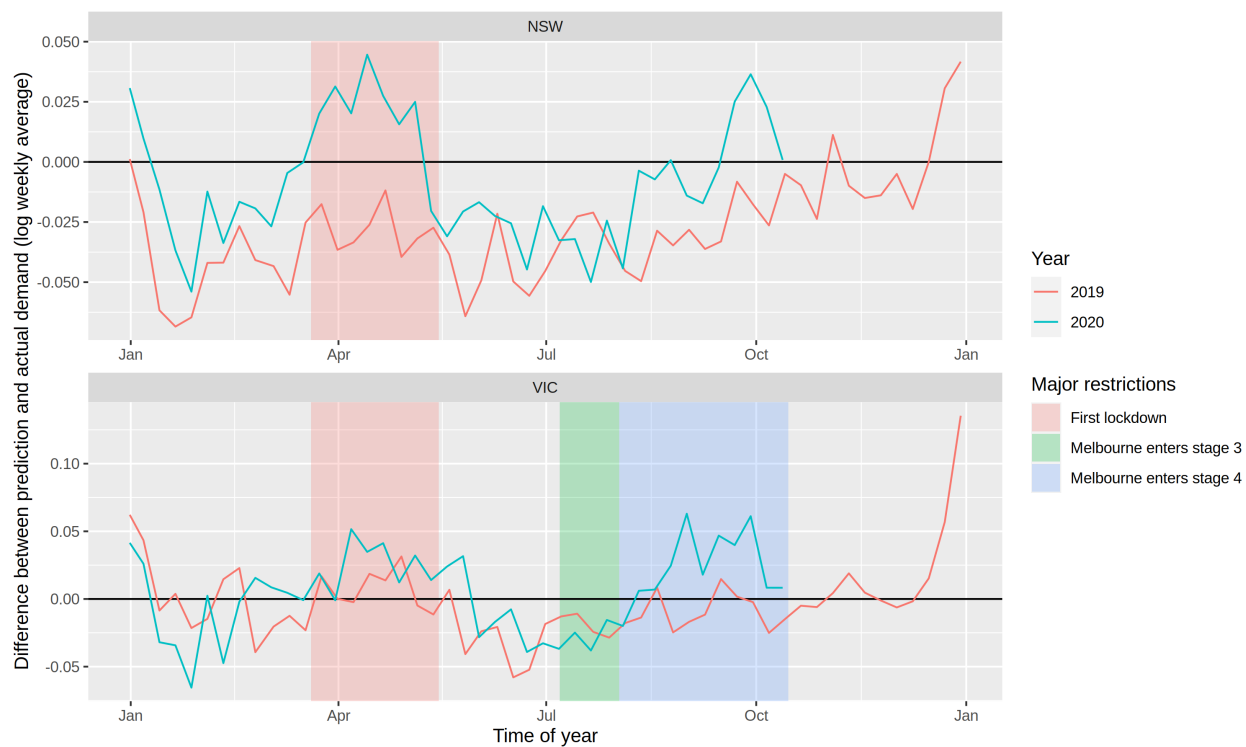
Figure 5: Difference between predictions and actual demand for 2019 and 2020, overlayed with approximate timing of major restrictions.
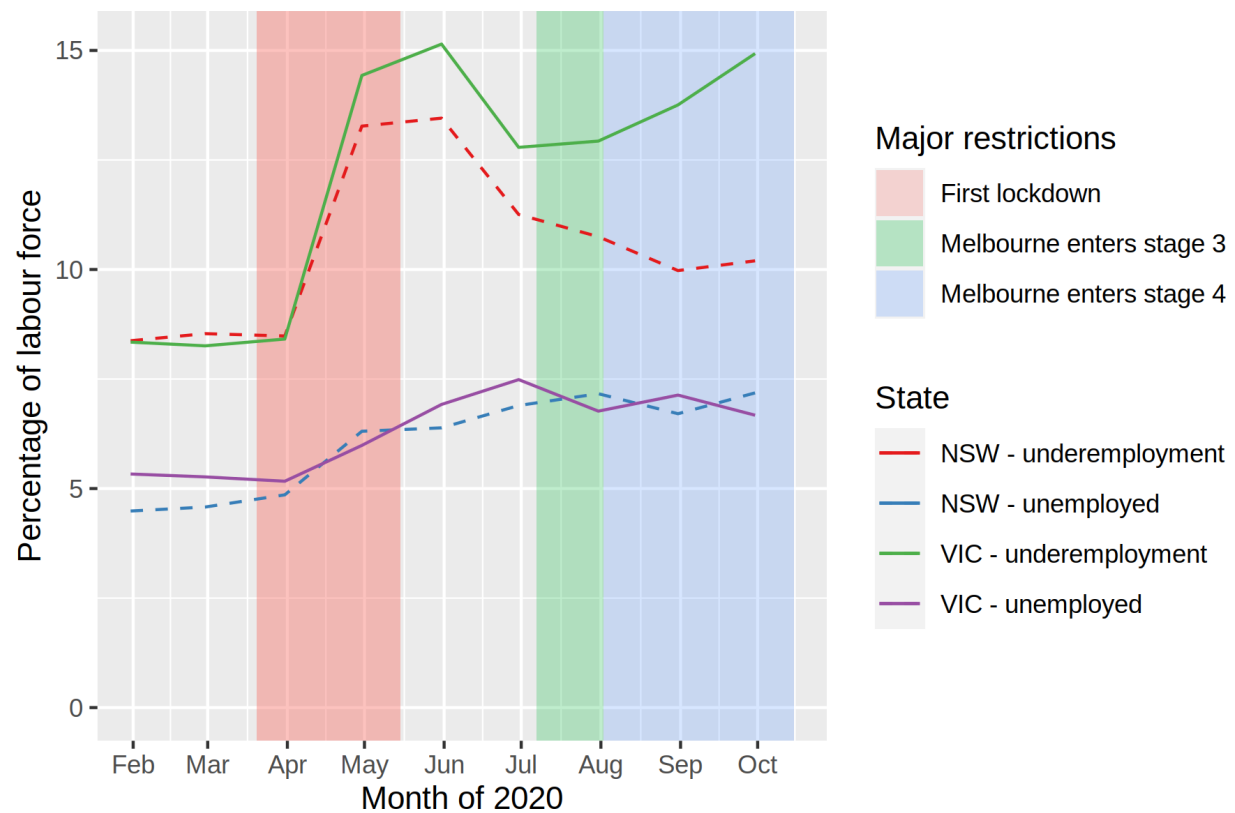
Figure 6: Unemployment data from the Australian Bureau of Statistics overlayed with the timing of major restrictions. Dashed lines are for NSW, and solid are for Victoria.