Electricity usage and the COVID-19 pandemic

Nathaniel Bloomfield

October 18, 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 could also be due to other factors, such as differences in weather and domestic solar power generation. Building a model based on previous years of electricity usage data gives us a way to take these differences into account and compare expected usage with what we are currently seeing during the COVID-19 pandemic, and provide insights into how COVID-19 has affected the economy.

2 Data

To build a model we assembled an 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 that met these requirements and link them together. 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 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. Incorporating solar raidation is important, as the influx of domestic solar demand during the day leads to decrease in usage around noon. We approximated solar radiation at a half-hourly level from the daily values using the times for sunrise, noon and sunset from the `suncalc` R package based on the centroids of the weather station coordinates for each state. This allowed us to create an estimate of solar intensity throughout the day using a piece-wise sine curve, which we then multiplied by the overall daily solar radiation values at each weather station to obtain our half-hourly approximation. This ignores the impact of cloud cover that might occur during some portions of the day, but we were unable to get access to solar radiation data with a higher temporal resolution.

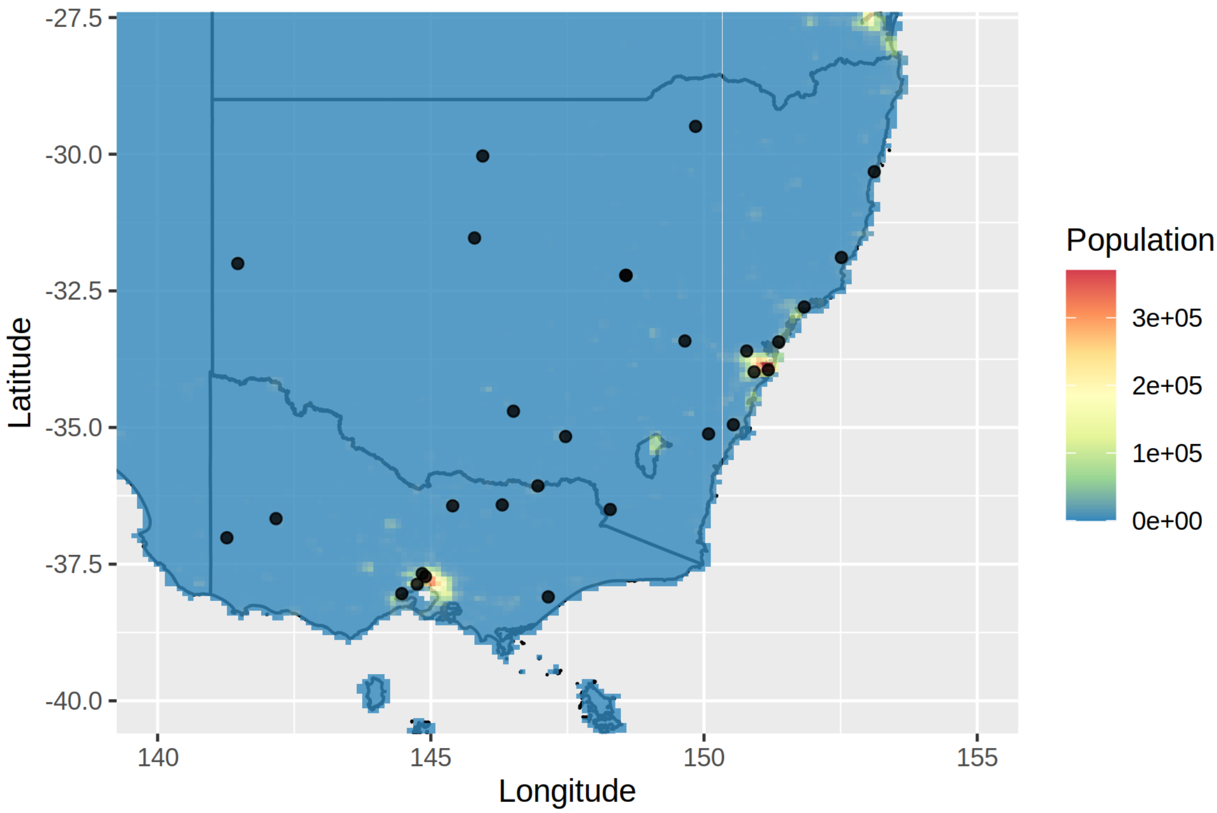

Figure 1: Black dots are the location of weather stations used for building models predicting electricity usage.

3 Model building

We used `xgboost` with the `tidymodels` framework in R to predict electricity usage. XGBoost is a gradient boosting technique for building ensembles of decision trees, and has been applied quite succesfully in a number of machine learning competitions. We considered also testing other methods, as well as hyperparameter tuning, but ran out of time in this project.

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, noon 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 the addition of temperature. 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. It also highlights the model is not perfect, and performs much better on some days than others. 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.

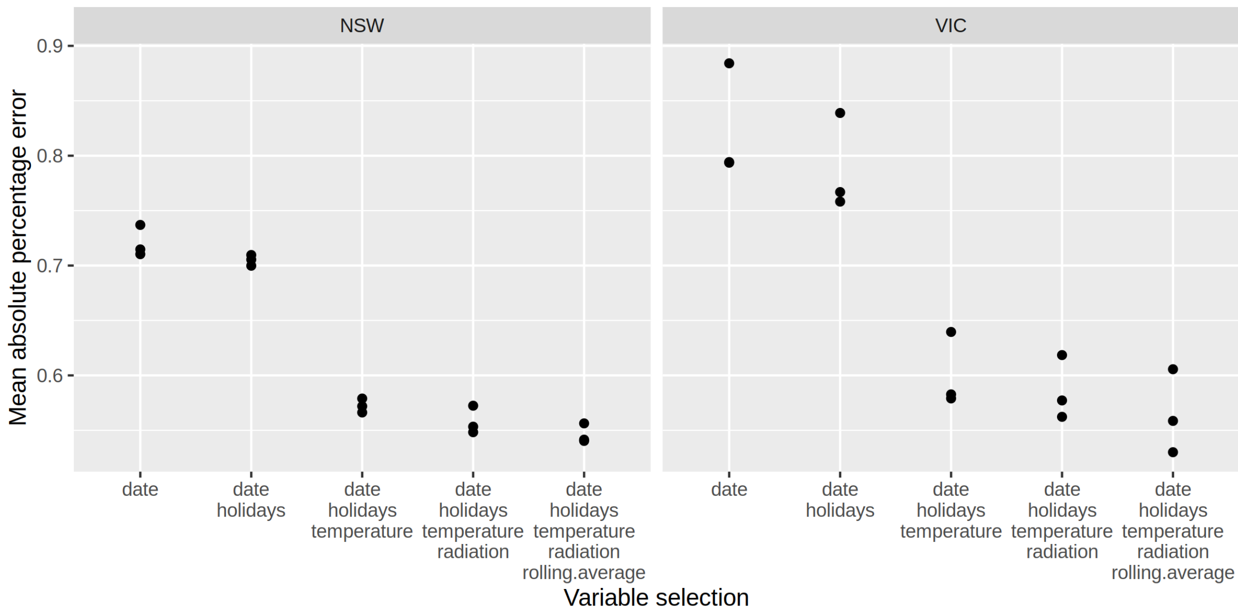

Figure 2: Model performance for VIC and NSW using different variable combinations. The three points for each set of variables are models tested on 2017, 2018 and 2019 data, but trained on all the preceding years.

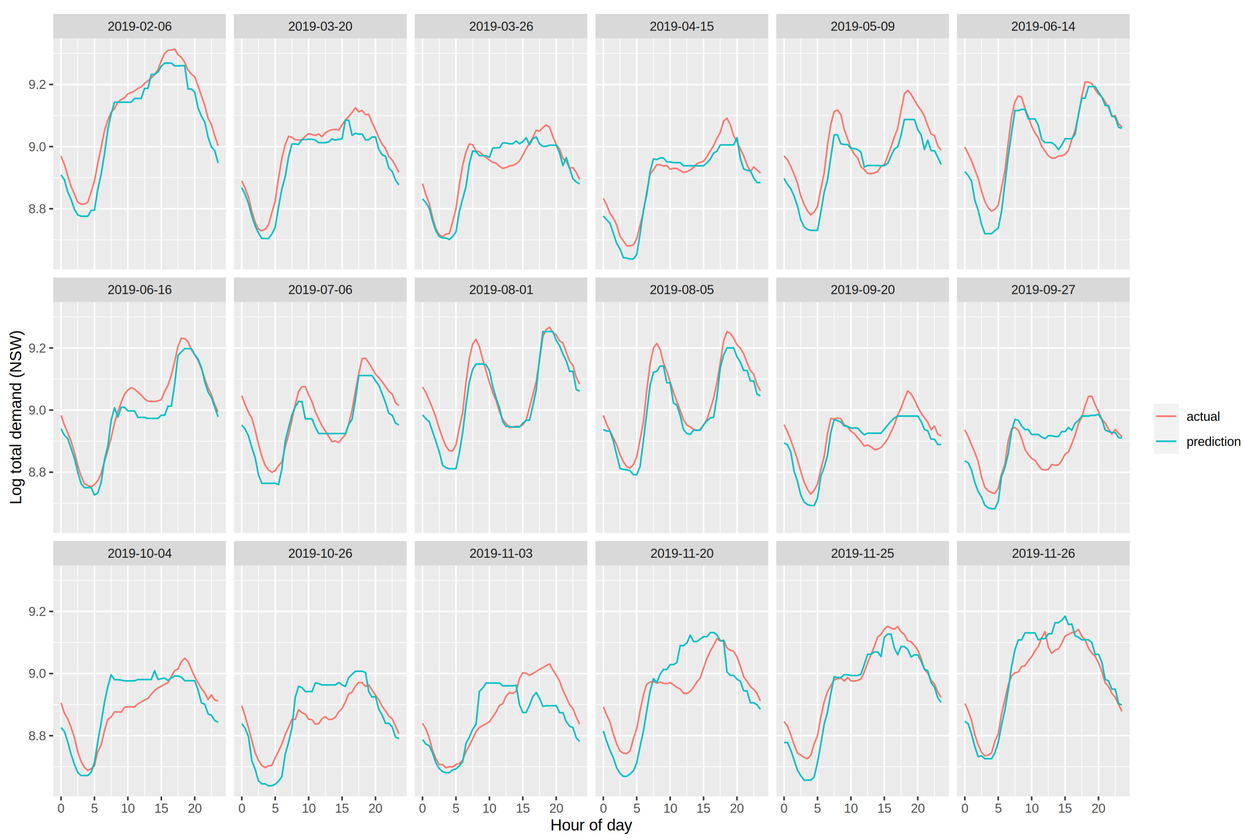

Figure 3: Raw results of model trained on all predictors for NSW over randomly selected days of the year in 2019.

4 Results

Working with eletricity usage data at the state level means we can only consider trends in overall usage. This is shown for 2019 and 2020 in Figure 4, alongside what our model predicts for 2020. While there are some large differences in usage between years, the difference between our model and the actual usage for 2020 appears to be quite minor. This is similar as to what others report, that in Victoria total usage fell only about 2% overall, but this masked a massive shift in where the energy was used. Namely, business usage fell, and residential usage increased. Unfortunately, we were not able to get access to any data for Australia that would have let us

If we have a look at the residuals, or the difference between actual usage and predicted for both 2019 and 2020, then we can start to see some of the smaller trends, as shown in Figure 4. In 2019 the model appeared to under-estimate electricity usage, except in December and January where it over-estimated usage. This suggests the model might need an additionaal variable to describe the holidays people generally take over the Christmas period. In 2020, the residuals for VIC look quite similar to the year before, but there is a striking difference in NSW during the first major lockdown, where it looks like energy usage slightly decreased over what was expected by the model. This lockdown was nationwide and ran from around the middle of March to the middle of May.

It is difficult to draw conclusions as to whether this difference is signficant enough to say that NSW was hit harder economically by this lockdown compared to Victoria, although finer scale data might be helpful in ascertaining this. In the quarter ending in June, state final demand fell by similar amounts (8.6% for NSW, 8.5% for Victoria). Unemployment also changed similarly, as shown in Figure 4. While both states both entered lockdown with similar levels of underemployment, NSW had a lower unemployment rate. By the end of that first lockdown, unemployment levels were similar between the states, but Victoria's underemployment was much higher. This suggests both states were impacted similarly, despite energy usage in NSW falling more.

Since the introduction of stage 4 restrictions in Melbourne, underemployment has begun to rise again while in NSW it has continued to decrease and come to a plateau in the last two months. However, the unemployment rate in both states has stayed stationary. In Victoria, energy usage also appears to have slightly decreased over what was expected since stage 4 restrictions were introduced, and in NSW there is a similar trend in the last month. It is uncertain if this decreasing trend in recent months will hold, but these numbers highlight the risk of a "w" shaped recession, particularly if there is a third coronavirus outbreak in either NSW or VIC, and that the economy is still much weaker compared to before the pandemic.

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.

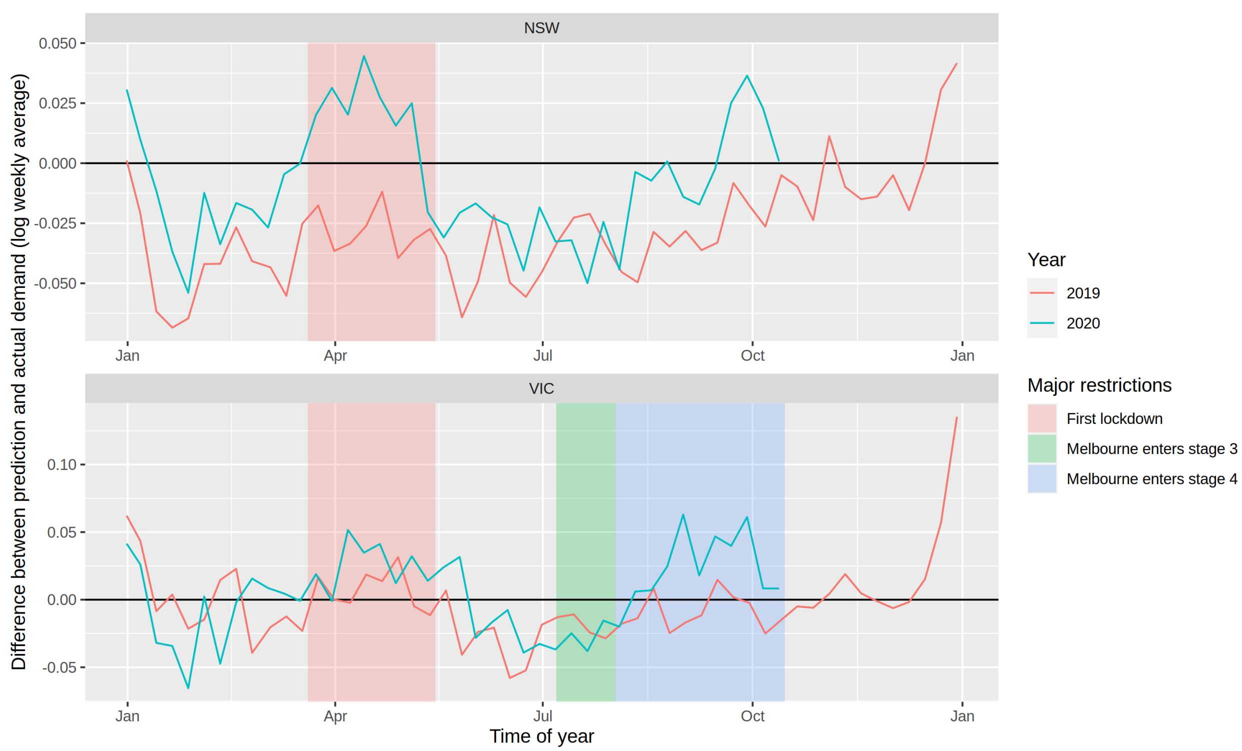

Figure 5: Difference between predictions and actual demand for 2019 and 2020, overlayed with approximate timing of major restrictions.

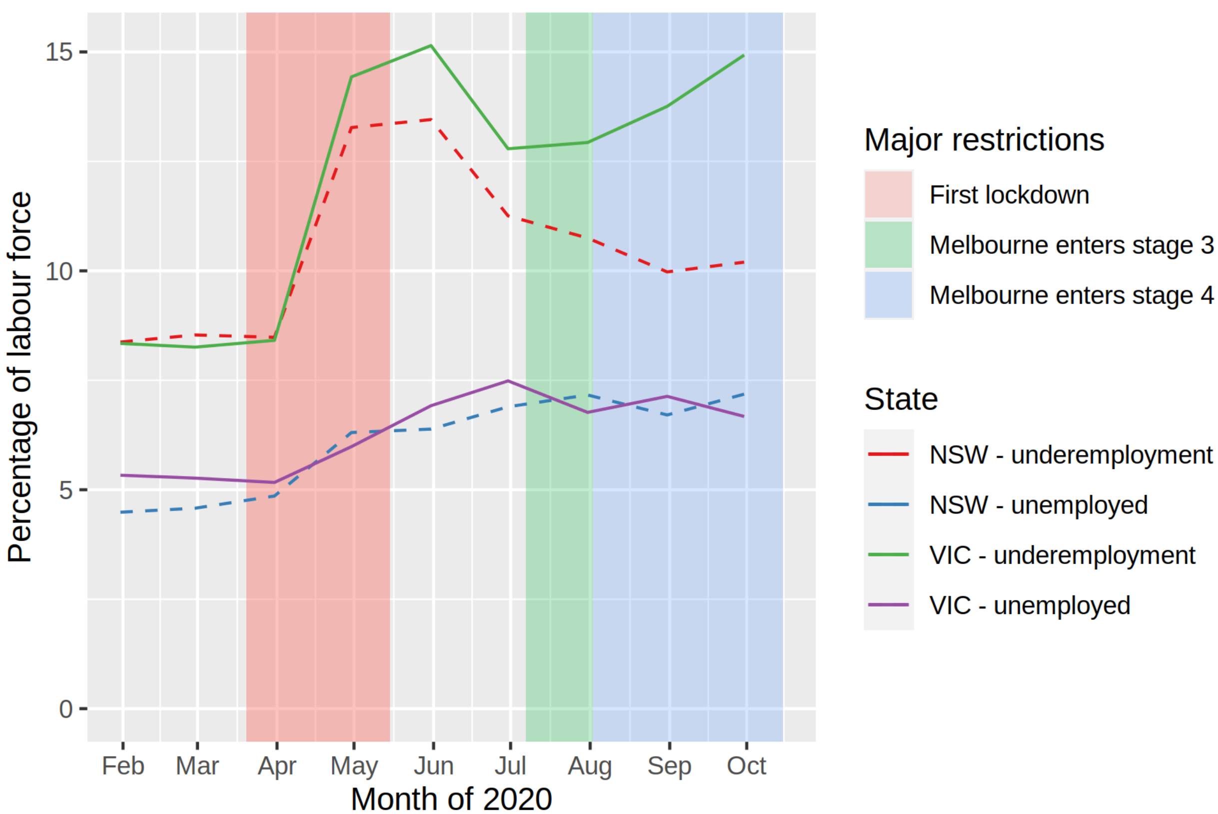

Figure 6: Unemployment data from the Australian Bureau of Statistics overlayed with the timing of major restrictions. Dashed lines are for NSW, and solid are for Victoria.