Sunlight Synchronisation: Exploring the Influence of Daylight Saving Time on CO2 Emissions and Electricity Consumption in Australia's Electricity Grid

Applied Econometrics Research Paper

Department of Economics

Toulouse School of Economics

Submitted to:

Jean Francois Fournel, PhD

Submitted by:

Davis, Koehler, Postler & Stegmaier

Student IDs: 22307993, 22308759, 22307997, 22308009

Degree Programme: M1 (Applied) Economics International Track

Toulouse, 19.03.2024

Contents

Acronyms

AEMO the Australian Energy Market Operator

DST Daylight saving time

DiD Difference-in-differences

DDD Difference-in-differences-in-differences

NEM National Electricity Market

USA United States of America

List of Figures

List of Tables

1 Introduction

Daylight saving time (DST) describes the practice of advancing clocks by one hour during the summer months of the year. For people whose daily schedule is based on clock time, this means that the sun appears to rise and set one hour earlier and they experience an additional hour of sunlight each day, as shown in Figure 1. Today approximately 70 countries have adopted DST as a policy (Prerau, 2006). Since DST was first introduced, the electricity industry has changed drastically. Solar generation is becoming more and more commonplace and inefficient compact fluorescent bulbs have been replaced with far more efficient LEDs. Today, lighting makes up only 15% of the overall electrical load worldwide (Zissis and Bertoldi, 2018). Despite being widely adopted over the last century, there is only moderate evidence to support the original energy-saving motivation for DST (Prerau, 2006). Most literature suggests that DST merely shifts load from the evening to the morning, causing a negligible net reduction in electricity consumption, or even a slight increase (Aries and Newsham, 2008; Guven et al., 2021; Kellogg and Wolff, 2008). In terms of emissions (and cost reduction), shifting *when* consumers consume electricity can be as impactful as reducing the volume consumed (Holland and Mansur, 2008). We propose that DST can act as a synchroniser between electricity demand and emission-free solar power generation. Our study adds to the existing literature by analysing the extent to which DST can change daily CO_2 emissions by aligning demand peaks with the availability of solar power. We base our analysis on recent data from Australia, whose grid has substantial solar capacity (34%, according to the Australian Energy Regulator (2023)). We make use of the fact that Australian regions had a heterogeneous take up of DST and that the observed demand shift only affects the morning and evening peaks but does not structurally change consumption behaviour around midday. We control for unobserved state-specific shocks by applying a Difference-in-differences-in-differences (DDD) framework. This strategy allows for the necessary identifying assumptions and is more fitting than an Difference-in-differences (DiD) model. Our results show that DST has a positive yet insignificant effect on greenhouse gas emissions, as well as on energy consumption, which is in line with existing literature (Kellogg and Wolff, 2008). The result is slightly larger, albeit still insignificant, when looking at the last five years, a period in which the overall renewable energy production increased. This might be because of the increased uptake of solar over these years.

2 Literature Review

There has been extensive research into the social and individual side effects of DST, such as increases in heart attacks, increases in criminal sentencing severity, decreases in crime and decreases in traffic accidents (Bünnings and Schiele, 2021; Cho et al., 2017; Doleac and Sanders, 2014; Jin and Ziebarth, 2020). However, fewer authors have researched the effect DST has on CO_2 emissions and electricity consumption.

Aries and Newsham (2008) give a comprehensive overview of the literature concerning the effect of DST on energy usage for lighting, an argument often used for the adoption of this policy. They find divided literature on this topic, presenting evidence for and against a decrease in energy usage, concluding that the research in this debate is "limited, incomplete, or contradictory". Nonetheless, they identify key variables which might affect the results of studies, namely geographical and climatological factors, as these play an important role in the energy consumption patterns of households.[1] Moreover, they also uncover evidence of a shift in demand within the day, particularly from the morning to the evening hours. This shift occurs because consumers can significantly decrease their lighting usage during the morning, given that the sun is already shining, but are more likely to increase it earlier in the evening, as the sun is setting earlier as well. Kotchen and Grant (2011) examine inter-day residential energy consumption patterns in Indiana (USA), where they find that DST correlates with an overall energy demand increase. They attribute this to a decrease in energy use for lighting, which is offset by heightened demand for heating and cooling purposes. The study does not dissect the consumption variations throughout the day but assumes a constant fuel mix and exclusively concentrates on household energy consumption. Consequently, it is hard to infer general results from their findings. Kellogg and Wolff (2008) analyse the effect an extension of DST for the Sydney 2000 Olympics had on electricity consumption. The authors also detect an intra-day shift in demand and apply a DDD framework, in which they assume that the midday load is not affected by DST as the third difference, to control for unobservables and shocks. The authors find that the intra-day shift does not affect the overall load in Australia, but only its timing. The study did consider the effect this shift could have on emissions, hypothesising that the emissions of generation units serving the peak loads can slightly affect the overall daily

[1] e.g. in Australia, cooling has a bigger impact on electricity demand than heating

emissions. Similarly, Bircan and Wirsching (2023) found a negligible effect on greenhouse gas emissions from a DiD analysis of Turkish electricity consumption and demonstrated that an intra-day change in load shape due to DST caused a decrease in emissions as a result.[2]
Our research will add a unique contribution to this literature by making use of the effect of DST on an intra-day demand shift, identified in prior research, to examine emissions within the energy sector. Using more recent and detailed data on the Australian intra-day fuel mix allows us to additionally analyse the potential impact an extensive use of renewable energy sources has on the alignment of electricity demand/supply[3] with sunshine and therefore, emissions.

3 Institutional Context and Data

3.1 Institutional Context

DST was first introduced country-wide in the German Empire and Austria in 1916 in order to support the war effort by minimising the use of fuel for artificial lighting (*Reichsgesetzblatt* 1916). Due to its energy-saving nature, the policy gained popularity among numerous countries, most notably almost all European countries, the United States of America (USA) and Russia. DST has seen many revisions and adjustments throughout the years (Prerau, 2006). It has been active nationwide in the USA since 1966 (*Uniform Time Act* 1966) and in most parts of Europe since the 1970s energy crisis (Pearce, 2017a). However, with the widespread adoption of highly efficient LED light bulbs nowadays, the original energy-saving argument of DST no longer applies (Aries and Newsham, 2008). Still, the policy remains active in approximately 70 countries (Prerau, 2006), excluding those near the equator where the minimal variation in sunrise and sunset times renders its implementation unnecessary. As of today, policymakers in many of those regions using DST are debating the abandonment of the policy. The European Union, Mexico, and the USA have considered discontinuing DST due to suspected minimal energy savings or non-energy-related side effects (Guven et al., 2021; *Ley del Sistema de Horario* 2022; *Sunshine Protection Act* 2022). However, uncertainty about the general effects of DST has hindered progress in implementing any changes, e.g in the EU since 2018 (*Directive 2000/84/EC*

[2] Although in a grid with a high prevalence of hydropower and negligible solar power
[3] Electricity demand must precisely match supply at all times, making the energy quantity solely determined by the inelastic demand. It is common for sunshine to be wasted.

2018).

Our focus on Australia is due to Australia's heterogeneous adoption of DST between states, which independently decide on the policy. After the short-term introduction of DST during both world wars in all states, the policy was reintroduced in Tasmania in 1967 because of a severe drought that led to a power shortage (*Daylight Saving Act* 1967).[4] In 1971, New South Wales, the Australian Capital Territory, Victoria and Southern Australia also followed DST.[5,6] Since then, DST has been active in the south-eastern states, with clocks shifting forward by one hour on the first Sunday of October and backwards on the first Sunday of April. In the Northern Territory DST has not been used since World War II.[7] In Western Australia, DST was active in different periods, but always quickly discontinued due to strong opposition from different parts of the population, most notably rural inhabitants (Pearce, 2017a). The main arguments against DST are Western Australia's closer location to the equator, which makes the policy less suitable, and its very hot climate - especially in the north-eastern parts - leading to negative effects of DST on the rural population in these areas. In Queensland, DST was active in several short periods, but always heavily discussed. On the one hand, the urban coastal regions on the (south) eastern side of Queensland are mostly in favour of DST and advances to re-introduce the policy are made every other year (Pearce, 2017b). On the other hand, most other areas - which have more extreme temperatures - are opposed to DST, arguing that one more hour of sun during the day would lead to negative effects on their inhabitants (Westcott, 2010). Furthermore, Queensland is located closer to the equator, making the state less suitable for the policy. The last referendum on DST in Queensland was decided with 55 % against the policy (*Daylight Saving Referendum Statistical Returns* 1992), which also illustrates how divided opinions on DST are.

With respect to total energy generation, Australia's energy grid has a substantial solar capacity (34%, according to the Australian Energy Regulator (2023)). The south-eastern states (New South Wales, Victoria, Southern Australia, Tasmania and Queensland) are linked into a single electricity grid, operated as the National Electricity Market (NEM) by the Australian Energy

[4]The majority of Tasmania's electricity generation comes from hydropower (the Australian Energy Market Commission, 2019)

[5]see *Daylight Saving Act* 1971a; *Daylight Saving Act* 1971b; Hasluck, 1972; *Standard Time Act* 1971

[6]Notice that due to its size, the ACT grid is electrically and commercially integrated as part of the NSW grid. Thus, when mentioning NSW this will include the ACT in our study

[7]The Northern Territory has a sparse population of roughly 250000 inhabitants (Northern Territory Government - Department of Treasury and Finance, 2024) making DST less relevant

Market Operator (AEMO). The breakdown of energy in the NEM over the last 12 months was 8% solar, 63% coal, 15% wind, 8% hydro. Comparatively, Queensland with 11% solar, 74% coal and 5% from wind energy (Australian Energy Market Operator, 2024) is thus relatively similar in its electricity grid.

Table 1 shows the energy generation and related CO_2 emissions throughout a typical day. A higher emissions intensity is observed shortly before sunrise and shortly after sunset compared to shortly after sunrise and before sunset, with a larger difference in the morning. Thus a shift of load from one time of day to another may not necessarily lead to a zero-sum change in emissions.

3.2 Data

Energy and carbon data from 2009 to 2023 was obtained from AEMO. The dataset does not include Western Australia nor the Northern Territory due to the structure of the NEM. To calculate the emissions per region over time, we take static emissions intensity (tCO2e/MWh) per generator, and join them with energy output (MW) per generator per 5 minutes. We sum to a region level, downsample to half-hourly data, and adjust for inter-region energy flows.[8] To obtain per capita values, quarterly population data per region was downloaded from the Australian Bureau of Statistics and linearly interpolated to a daily level. Weather data was obtained from the Bureau of Meteorology and Willy Weather. Temperature data was taken from capital cities, since most thermal load (building heating and cooling) occurs in capital cities. Wind and sunlight data was taken from region midpoints, as representative values for generation which is typically dispersed across the region. Our final dataset includes information on CO_2, kWh, temperature, solar exposure, wind, and population, as well as holiday and weekend dates. Table 2 presents summary statistics for all variables, showing that Queensland and the treated regions are relatively similar.

4 Methodology and Model

We analyse CO_2 emissions from the electricity grid before and after the exogenous DST change using a Difference-in-differences (DiD) framework. To compare our results to the literature

[8] available only with 30-minute granularity. The calculations are explained in detail in Appendix 6.

we additionally run an equivalent specification looking at electricity consumption in kWh. As shown in Figure 2, we specify Queensland, which does not use DST, as our control and the other states in the NEM as the treatment. To get plausible coefficients from the DiD regression, we first verify the common prior trend assumption, using an event-study graph. For this, we use the DST time shift in the respective states as the treatment, with emissions and electricity consumption as the outcomes, as shown in Figure 8. We include further controls and entity and time-fixed effects for every state and date to isolate the effect of DST from possible confounders. Further controls include average sunlight irradiance, average wind speed, maximum daily temperature, weekends and public holidays. DST coincides generally with longer daylight hours, and thus greater solar power generation, which displaces fossil fuel generation and can potentially reduce emissions. To control for this, average daily sunlight irradiance (adjusting for cloud cover and the Earth's tilt) is used as a control. Similarly, average wind speed is used as a control for wind generation. Maximum daily temperature is used as a control to explain heating and cooling loads.[9] Weather data is aggregated at a daily level because intra-day data (especially sun intensity) would be a collider. Controls were added for weekend and public holidays as they tend to reduce electricity demand significantly.

To correct for a potentially missing common prior trend and improved interpretability we follow Kellogg and Wolff (2008) and additionally implement a DDD design, first performed by Gruber (1994), and the associated event study. The DiD framework allows us to identify the average differences between the treatment and control states. However, it does not control for state-specific demand shifts. To do so, we make use of the fact that electricity demand during the midday period does not see a shift from DST, compared to morning and evening peaks.[10]

4.1 Difference-in-differences (DiD) Estimation

Following Callaway and Sant'Anna (2021) and Goodman-Bacon (2021), Equation 1 shows the DiD regression we implement.

$$\left(\frac{CO_2}{Population}\right)_{r,t} = \beta_0+\beta_1*Treatment_r+\beta_2 Post_t+\beta_3(Treatment\times Post)_{r,t}+\beta_4 Controls_{r,t}+\epsilon_{r,t} \tag{1}$$

[9]Due to thermal inertia, instantaneous temperature would be a less effective control
[10]compare Kellogg and Wolff (2008)

Treatment is a binary variable, which is 1 if a region has in general implemented a policy for DST (i.e. all studied regions except Queensland). *Post* is 1 during each period in which DST is active (October to March). Errors are clustered by region to mitigate the impact of serial correlation. Furthermore, the data is weighted by population. We also run the same DiD specification with kWh per capita as an alternative outcome. To examine the common prior trend assumption for both emissions and electricity consumption, equation 2 shows the event study performed using the Stata package provided by Clarke and Tapia-Schythe (2021).

$$y_{r,t} = \sum_{j=2}^{J} \beta_{-j} \times (Lead_j)_{r,t} + \sum_{k=0}^{K} \beta_k \times (Lag_k)_{r,t} + \mu_r + \lambda_t + X'_{r,t} + \epsilon_{r,t} \tag{2}$$

where $y_{r,t}$ is CO_2 or electricity consumption per capita for region r and time period t respectively. The β coefficients represent our event study estimates for the lead and lag effects of DST respectively, based on the time to treatment t. In our case, this is the number of days into DST, being negative when DST is not active, 0 on the days when both forward and backward DST transitions occur, and positive when DST is active. μ_r and λ_t are entity and time-fixed effects by region and day respectively. $X_{r,t}$ represent a vector of controls per region and day and their respective coefficients.

4.2 Difference-in-difference-in-difference (DDD) Estimation

The DDD design allows us to correct for unobserved factors in the DiD-framework affecting the control and treatment groups differently. To establish the DDD, we normalise by emissions and electricity demand during 12:00-14:30. Midday emissions and electricity demand respectively will be the largely unaffected by the time shift from DST, because the sun is at its highest point.

$$\begin{aligned}\left(\frac{CO_2}{Population}\right)_{r,t,m} &= \beta_0 + \beta_1 Treatment_r + \beta_2 Post_t + \beta_3 NotMidday_m \\ &+ \beta_4 (Treatment \times Post)_{r,t} + \beta_5 (Treatment \times NotMidday)_{r,m} \\ &+ \beta_6 (Post \times NotMidday)_{t,m} + \beta_7 (Treatment \times Post \times NotMidday)_{r,t,m} \\ &+ \beta_8 Controls_{r,t} + \epsilon_{r,t,m}\end{aligned} \tag{3}$$

Adding this third difference, we run our DDD regressions (Equation 3) with the same data, the same additional outcome of $kWh\ p.c$, and clustering by region and weighting by population. The variable of the third difference ($NotMidday$) is 0 for the half hours between 12:00 and 14:30 (local time), and 1 otherwise, with the subscript m specifying whether the observation takes place during the midday period. To check our prior common trends assumption, we again apply the event study design shown in Equation 2, adjusting our outcome variables by the midday values for CO_2 and electricity production respectively. This is equivalent to taking ratios of the outcome variable, and allow us to create a quasi-equivalent event study design for the DDD regression we perform, as specified by Olden and Møen (2022). The main coefficient of interest is β_7 that captures the effect of not being in the period between 12 and 14:30, in a treatment region, while DST is active.

5 Results

5.1 Difference-in-Differences (DiD) Results

The event study plot for the DiD design is shown in Figure 3. As anticipated, the common prior trend does not hold for this specification, confirming our reasoning in running the DDD model. Table 3 shows our estimation results but no causal interpretation can occur due to the common prior trend assumption not holding. Looking at the results for emissions, we observe that adding our control variables changes the significance of the DiD-coefficient to the point that it is no longer significant at the 5% level whilst the sign of the coefficient remains negative. We additionally see that the absolute size of the coefficient and therefore the estimated effect of DST on emissions decreases as we add our controls. We observe that only our controls for weekends and public holidays are significant, whilst temperature, solar exposure and wind are insignificant. This can be explained by the fact that the variation between working and non-working days is bigger than possible variations in the other confounders. The results for electricity volume generally do not differ greatly in their interpretation compared to emissions.

5.2 Difference-in-Difference-in-Differences (DDD) Results

After controlling for the respective midday values with an additional difference, the event study in Figure 4 shows that the common prior trends assumption now holds for both electricity consumption and emissions. No clear effect of DST on the respective outcomes can be seen from our event study graphs as our lag coefficients remain centred around 0.

Tables 4 and 5 show the results for the DDD regression. Compared to our DiD design, the sign of our DDD coefficient is reversed and positive for both outcomes (0.0163 Kg CO_2 and 0.0185 kWh respectively), albeit insignificant. Including our controls does not change this result. Weekend and public holidays remain significant, whilst our other controls remain insignificant. A possible reason for this is that by controlling for the midday values, we control for most of the between-day and region variations that would be explained by these controls.

5.3 Robustness Checks and Limitations

We perform multiple robustness checks. Tasmania is a region which differs greatly from the control region. It has far lower emissions, mainly hydropower, no solar, colder temperatures and a larger weather difference between summer and winter. However, given that Tasmania's population is notably small, excluding Tasmania from the analysis had minimal impact on the results (7), due to the weighting of the results according to population already carried out. After dropping some zero-emission periods (that are only present for Tasmania), the regressions can be performed using the natural logarithm of the dependent variables. As shown in Table 6, there are no major differences in significance, whilst our main coefficients of interest now represent a 1.86% and 1.53% increase in emissions and electricity production respectively. Solely taking the last 5 years of data as a robustness check increased the size of our coefficients of interest, whilst still being insignificant to the 5% level (see Figure 5 and Table 7). The increase is in line with our hypothesis that the larger percentage of solar energy in Australia's electricity grid in recent years has an emissions-reducing effect on our results. Additionally, we separated our regressions and event studies by DST transition direction, as shown in Figure 6. Whilst different coefficients are observed between shifting the time forwards and shifting backwards in the event study graphs and the DDD results, our results remain insignificant. Finally, the 12:00-14:30 midday control period was initially calculated using local time, as per Kellogg and Wolff, 2008.

Using standard Queensland time for all regions instead did not meaningfully change the result. Although our results hold to multiple robustness checks, there are still limitations to our study. Firstly, although we do cluster our standard errors on the region level to account for serial-correlation, this approach is somewhat flawed since the number of clusters used is very small. Secondly, our data is for the overall aggregated grid, lacking more granular household-level demand data that might give clearer insights into the causal effects of DST. An interesting future study would be to do a regression discontinuity design, comparing households in treated and untreated regions. Moreover, the closer location of Queensland to the equator directly affects the effectiveness of DST and consequently its non-reintroduction. Although we control for multiple factors, we can not be certain to isolate the true causal effect of DST on emissions.

6 Conclusion

Given the environmental imperative to reduce greenhouse gas emissions in all aspects of life, we take a closer look at the policy of daylight saving time (DST) through which people's clock-based behaviour is more aligned with sunshine. In an electricity network with a significant percentage of energy produced by solar, a synchronisation between this production and the demand might result in a reduction in overall greenhouse gas emissions, when controlling for all possible confounders. We analyse the effect of DST on CO_2 emissions and electricity in Australia, taking advantage of the heterogeneous adoption of DST in the nation for a time span of fourteen years from 2009 until 2023 with two DST-transitions per year. Using half-hourly panel data on electricity and CO_2 production per region, we apply a DDD treatment effect model. We use the framework to possibly identify significant differences during DST between the treatment and control regions, and using midday loads, which are not affected by the demand shift, to control for inter-day shocks. Our results do not suggest that a significant effect of DST exists on CO_2 equivalent emissions nor on electricity production. The interpretation of our results remains mostly unaffected by multiple robustness checks. Compared to previous literature, our results align with the findings of Kellogg and Wolff (2008), who also do not find a significant aggregate effect but instead an intra-day shift. Thus, our results could also bring a new impetus to the ongoing policy-discussions of DST, since no strong evidence is found for the original energy-saving argument of the policy.

References

Aries, M. B. and Newsham, G. R. (2008), “Effect of daylight saving time on lighting energy use: A literature review”, *Energy Policy* 36. (6), 1858–1866.

the Australian Bureau of Statistics (2023), *National, state and territory population*, URL: https://www.abs.gov.au/statistics/people/population/national-state-and-territory-population/jun-2023#data-downloads (visited on 11/28/2023).

the Australian Energy Market Commission (2019), *Changing Generation Mix: Tasmania*, URL: https://www.aemc.gov.au/energy-system/electricity/changing-generation-mix/tasmania (visited on 03/18/2024).

Australian Energy Market Operator (2024), *Data Dashboard - Fuel Mix*, URL: https://aemo.com.au/energy-systems/electricity/national-electricity-market-nem/data-nem/data-dashboard-nem (visited on 03/18/2024).

– (2023a), *NEMWeb MMSDM*, URL: https://www.nemweb.com.au/Data_Archive/Wholesale_Electricity/MMSDM/ (visited on 11/25/2023).

– (2023b), *NEMWeb Reports*, URL: https://www.nemweb.com.au/REPORTS/ (visited on 11/25/2023).

Bircan, Ç. and Wirsching, E. (2023), “Daylight saving all year round? Evidence from a national experiment”, *Energy Economics* 127, 107093.

Bünnings, C. and Schiele, V. (2021), “Spring Forward, Don’t Fall Back: The Effect of Daylight Saving Time on Road Safety”, *The Review of Economics and Statistics* 103. (1), 165–176.

the Bureau of Meteorology (2023), *Climate Data Online*, URL: https://reg.bom.gov.au/climate/data/index.shtml (visited on 11/28/2023).

Callaway, B. and Sant’Anna, P. H. (2021), “Difference-in-Differences with multiple time periods”, *Journal of Econometrics* 225. (2), 200–230.

Cho, K., Barnes, C. M., and Guanara, C. L. (2017), “Sleepy Punishers Are Harsh Punishers: Daylight Saving Time and Legal Sentences”, *Psychological Science* 28. (2), 242–247.

Clarke, D. and Tapia-Schythe, K. (2021), “Implementing the panel event study”, *The Stata Journal: Promoting communications on statistics and Stata* 21. (4), 853–884.

Daylight Saving Act (1967), Parliament of Tasmania, URL: http://classic.austlii.edu.au/cgi-bin/download.cgi/cgi-bin/download.cgi/download/au/legis/tas/num_act/dsa196733o1967211.pdf (visited on 11/03/2024).

Daylight Saving Act (1971a), Parliament of Victoria, URL: http://classic.austlii.edu.au/cgi-bin/download.cgi/cgi-bin/download.cgi/download/au/legis/vic/hist_act/dsa1971165.pdf (visited on 11/03/2024).

Daylight Saving Act (1971b), Parliament of Southern Australia, URL: https://www.legislation.sa.gov.au/__legislation/lz/c/a/daylight%20saving%20act%201971/current/1971.54.auth.pdf (visited on 11/03/2024).

Daylight Saving Referendum Statistical Returns (1992), URL: https://documents.parliament.qld.gov.au/tableoffice/tabledpapers/1992/4692t2333.pdf.

Doleac, J. L. and Sanders, N. J. (2014), "Under the Cover of Darkness: How Ambient Light Influences Criminal Activity", *Review of Economics and Statistics* 97, 1093–1103.

Goodman-Bacon, A. (2021), "Difference-in-differences with variation in treatment timing", *Journal of Econometrics* 225. (2), 254–277.

Gruber, J. (1994), "The Incidence of Mandated Maternity Benefits", *The American Economic Review* 84. (3), 622–641.

Guven, C. et al. (2021), "When does daylight saving time save electricity? Weather and air-conditioning", *Energy Economics* 98, 105216.

Hasluck, P. (1972), *Standard Time and Summer Time Ordinance*, Governor General of the Commonwealth of Australia, URL: https://classic.austlii.edu.au/cgi-bin/download.cgi/cgi-bin/download.cgi/download/au/legis/act/num_ord/stasto1972332.pdf (visited on 12/03/2024).

Holland, S. P. and Mansur, E. T. (2008), "Is Real-Time Pricing Green? The Environmental Impacts of Electricity Demand Variance", *The Review of Economics and Statistics* 90. (3), 550–561.

Jin, L. and Ziebarth, N. R. (2020), "Sleep, health, and human capital: Evidence from daylight saving time", *Journal of Economic Behavior & Organization* 170, 174–192.

Kellogg, R. and Wolff, H. (2008), "Daylight time and energy: Evidence from an Australian experiment", *Journal of Environmental Economics and Management* 56. (3), 207–220.

Kotchen, M. J. and Grant, L. E. (2011), “Does Daylight Saving Time Save Energy? Evidence From a Natural Experiment in Indiana”, *The Review of Economics and Statistics* 93. (4), 1172–1185.

Ley del Sistema de Horario (2022), Senado De La Republica (Senate of the Republic of Mexico), URL: https://comunicacionsocial.senado.gob.mx/informacion/comunicados/4104-pleno-del-senado-aprueba-dictamen-para-expedir-ley-de-los-husos-horarios-en-mexico (visited on 11/28/2023).

Northern Territory Government - Department of Treasury and Finance (2024), *Northern Territory Economy - Population*, URL: https://nteconomy.nt.gov.au/population (visited on 03/18/2024).

Olden, A. and Møen, J. (2022), “The triple difference estimator”, *The Econometrics Journal* 25. (3), 531–553.

Pearce, C. (2017a), *The great daylight saving time controversy*, 1st edition 2017, electronic, OCLC: 973478648, Acacia Ridge, Qld.: Australian eBook Publisher.

– (2017b), “History of daylight saving time in Queensland”, *Queensland History Journal* 23. (6), 389–403.

Prerau, D. (2006), *Seize the Daylight: The Curious and Contentious Story of Daylight Saving Time*, Basic Books.

Directive 2000/84/EC (2018), *Proposal for a Directive of the European Parliament and of the Council discontinuing seasonal changes of time and repealing Directive 2000/84/EC*, Council of the European Union, URL: https://www.consilium.europa.eu/media/39190/seasonal-time-changes_progress-report_3-december-2018.pdf (visited on 03/16/2024).

Reichsgesetzblatt (1916), *Reichsgesetzblatt Jahrgang 1916 Nr 67*, URL: https://www.reichsgesetzblatt.de/D/RGBl-D/1916/index.htm (visited on 11/03/2024).

Standard Time Act (1971), Parliament of New South Wales, URL: https://classic.austlii.edu.au/cgi-bin/download.cgi/cgi-bin/download.cgi/download/au/legis/nsw/num_act/sta1971n38164.pdf (visited on 11/03/2024).

State of the Energy Market (2023), Organization: Australian Energy Regulator, URL: https://www.aer.gov.au/system/files/2023-10/State%20of%20the%20energy%20market%202023%20-%20Full%20report_1.pdf.

Sunshine Protection Act (2022), *Sunshine Protection Act of 2021*, Congress of the United States of America, URL: https://www.congress.gov/bill/117th-congress/senate-bill/623 (visited on 11/28/2023).

Uniform Time Act (1966), *Uniform Time Act of 1966*, Senate and House of Representatives of the United States of America in Congress assembled, URL: https://www.govinfo.gov/content/pkg/STATUTE-80/pdf/STATUTE-80-Pg107.pdf (visited on 03/15/2024).

Westcott, M. (2010), *Daylight saving in Queensland: Daylight Saving for South East Queensland Referendum Bill 2010 (Qld)*, OCLC: 672278234, Brisbane: Queensland Parliamentary Library.

Willy Weather (2024), *Weather Stations*, URL: https://www.willyweather.com.au/climate/weather-stations/nsw/central-west/dubbo-airport.html?superGraph=grain:daily,graphRange:5days&climateRecords=period:all-time&longTermGraph=plots:temperature,period:all-time,month:all&windRose=period:1-year,month:all-months (visited on 02/01/2024).

Zissis, G. and Bertoldi, P. (2018), *Status of LED-Lighting world market in 2017*, European Comission.

Appendix

Appendix A: Figures and Tables

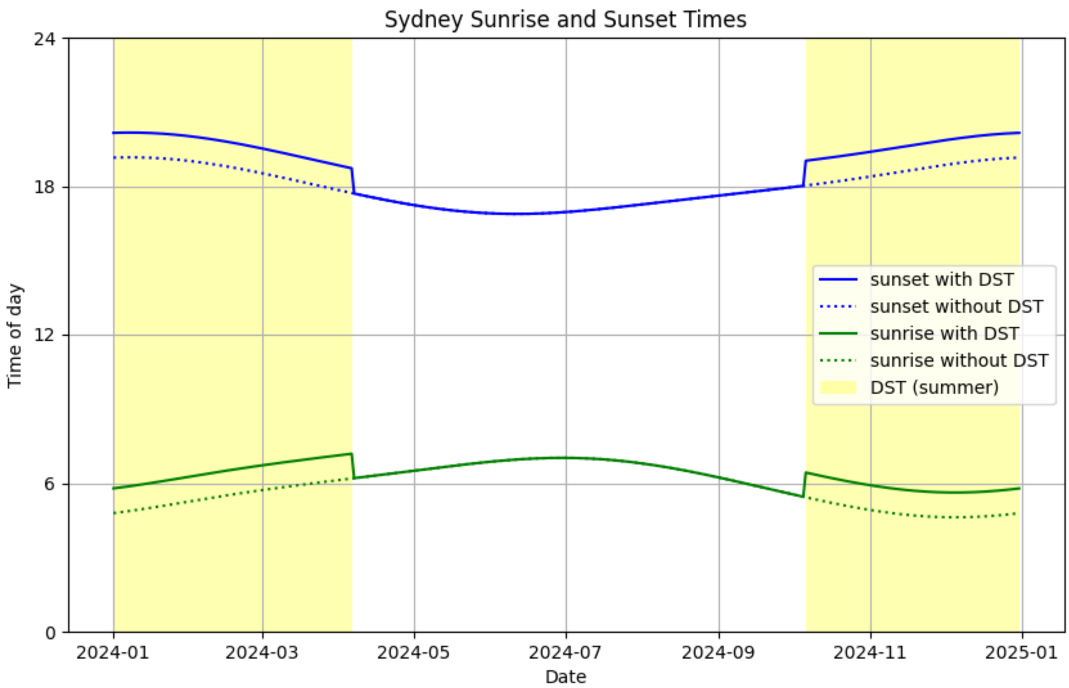


Figure 1: Effect of DST on apparent sunrise and sunset times in Sydney. Note that since Australia is in the southern hemisphere, summer and DST occur at the start and end of the calendar year.

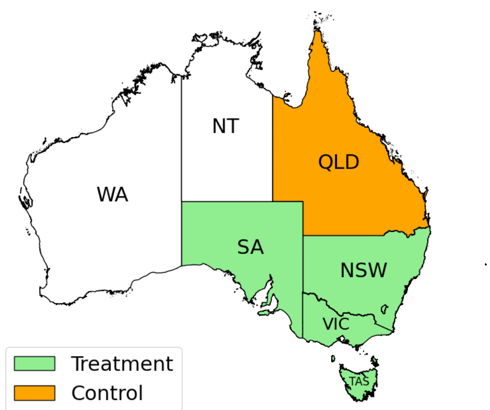
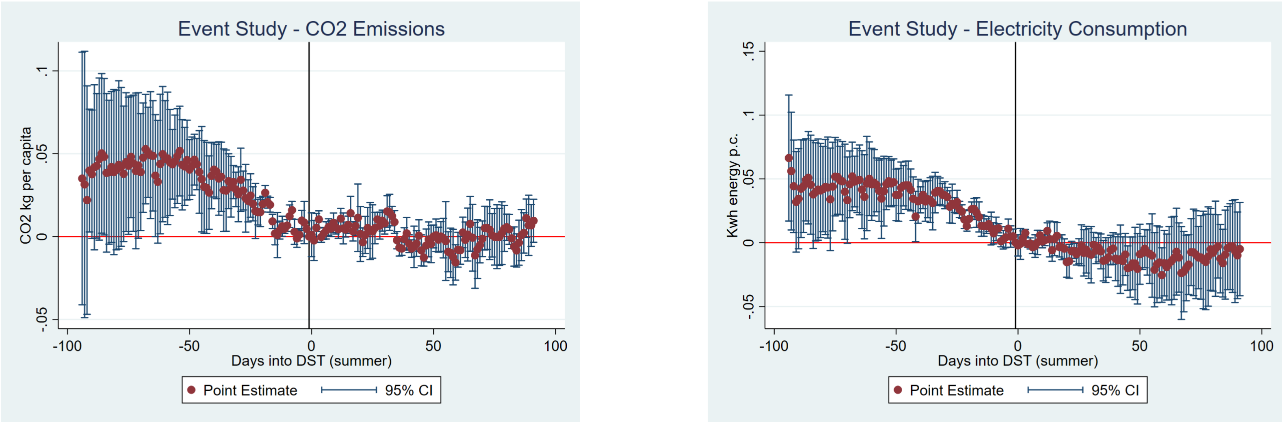


Figure 2: Map of Australia showing treatment and control regions

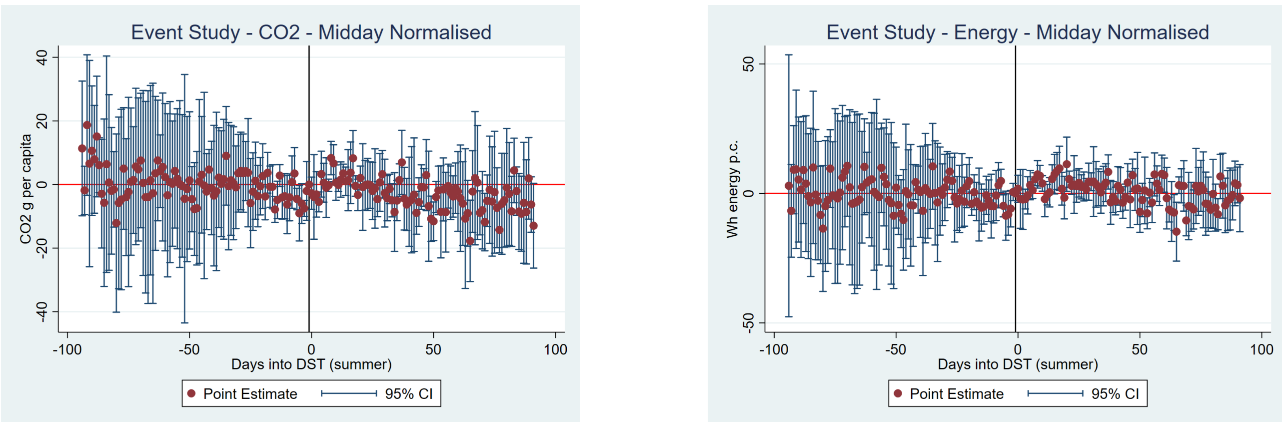


(a) DiD Event study plot for CO2 Emissions (b) DiD Event study plot for CO2 Emissions

Figure 3: Event study plots for DiD. Both forward and backward clock changes are represented on the same plot, with the latter flipped accordingly. Summer is on the right of each plot, winter on the left. A common prior trend does not exist, because the points on the left half of each plot are jointly significant. The right half of each plot is close to zero, suggesting some sort of common post trend. However this is the treatment period, so such a result cannot be used to justify a DiD to the control region.

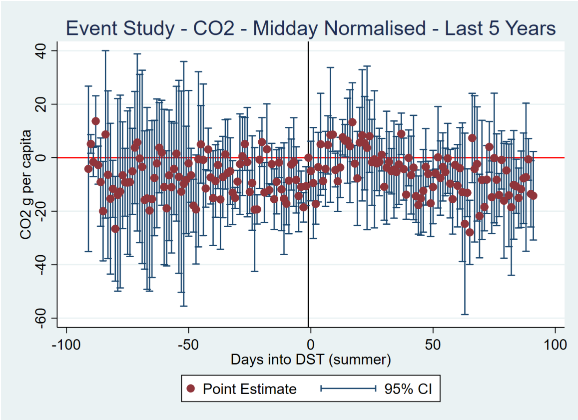
Table 1: Energy and emissions near sunrise and sunset. Emissions intensity is lower when the sun is up, but the difference is not the same.

period of day	power (W per capita)	CO2 (g/h per capita)	CO2 Intensity (g / kWh)
the hour before sunrise	477	453	982
the hour after sunrise	513	476	959
the hour before sunset	577	520	929
the hour after sunset	592	530	923
remainder of day	506	468	952

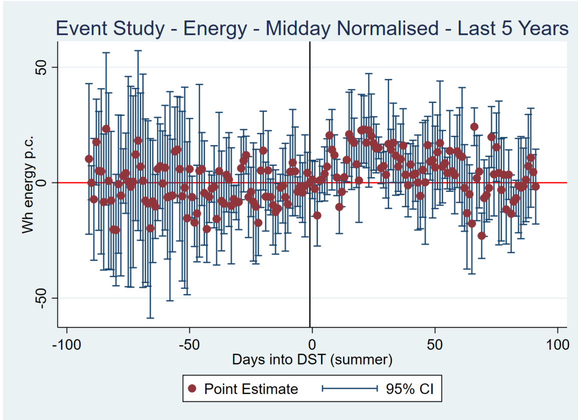


(a) DDD Event study plot for CO2 Emissions (b) DDD Event study plot for CO2 Emissions

Figure 4: Event study plots for DiD. For each day, for each region, the average value of emission or energy during 12:00-14:30 was calculated, and subtracted from the value for all intervals. Then the same procedure was applied as for Figure 3. The left half of each plot is approximately zero, so there is a common prior trend. The right half is approximately zero, so there is a null result.

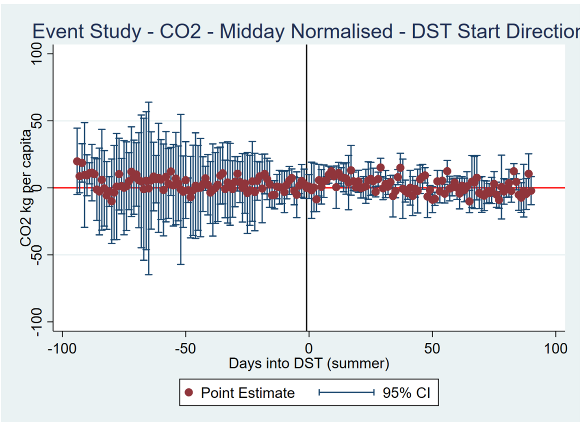


(a) DDD Event study plot for CO2 Emissions - Last 5 years

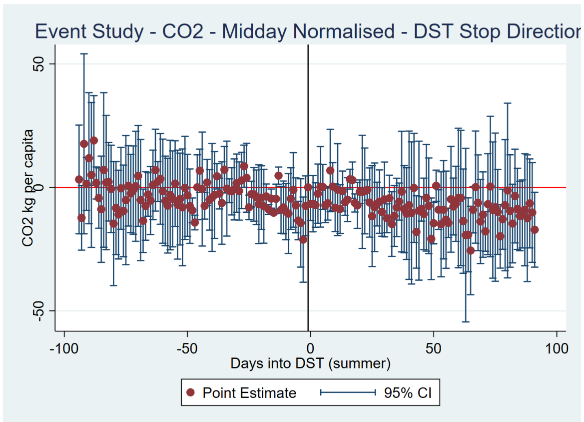


(b) DDD Event study plot for Electricity consumption - Last 5 years

Figure 5: Event study plots for DDD. Using only the last 5 years of our data we see the size of our coefficients increase relative to the errors. However, our DDD regression results still remain insignificant at the 5% level. 7

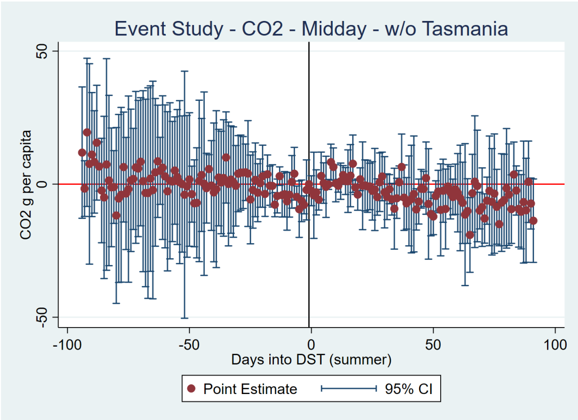


(a) DDD Event study plot for CO2 Emissions only for the DST start (clock forward)

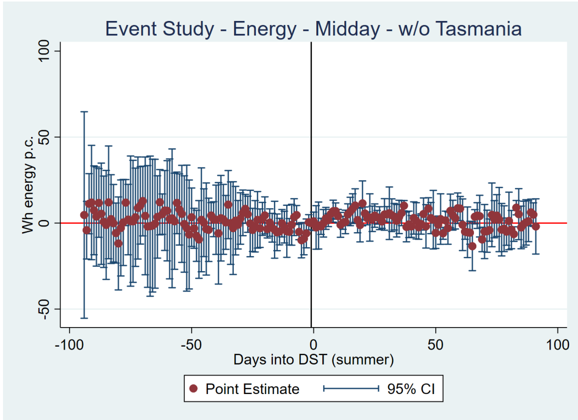


(b) DDD Event study plot for CO2 Emissions only for the DST stop (clock backward). The plot chronologically reversed. The data starts in the middle of summer, during DST on the right of the plot, and time progresses forwards towards the left. The clocks are moved back on the morning of day -1. Then once the treatment is unapplied, the "Days into DST" continues to become more negative, with mid-winter on the far left.

Figure 6: Event study plots for DiD, split by clock direction. This is the same as Figure 4a, but split into clock forward and clock back transitions, as a robustness check. The common prior trend and null result still hold.

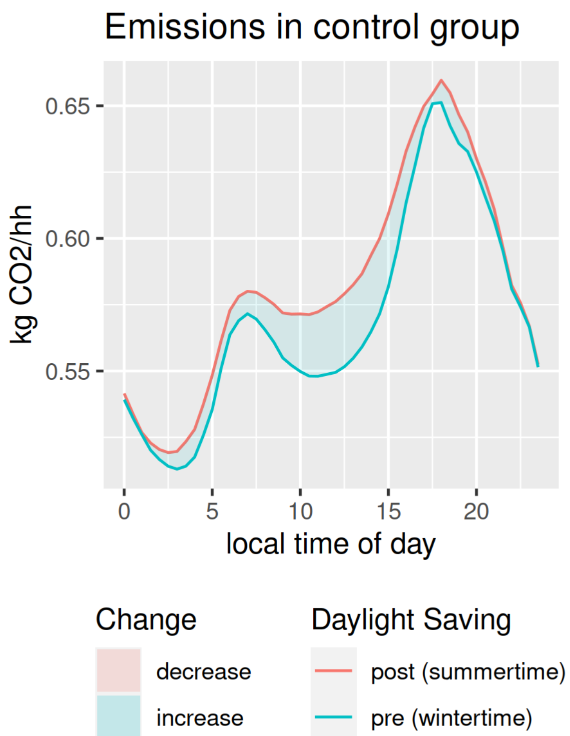


(a) DDD Event study plot for CO2 Emissions when normalising for midday values and dropping Tasmania

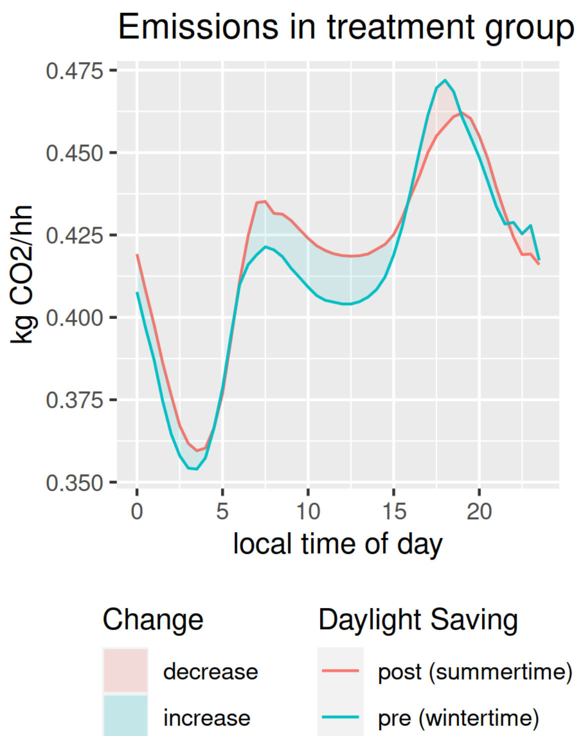


(b) DDD Event study plot for Electricity consumption when normalising for midday values and dropping Tasmania

Figure 7: Event study plots for DDD, dropping Tasmania as a robustness check. The common prior trend and null result still hold.

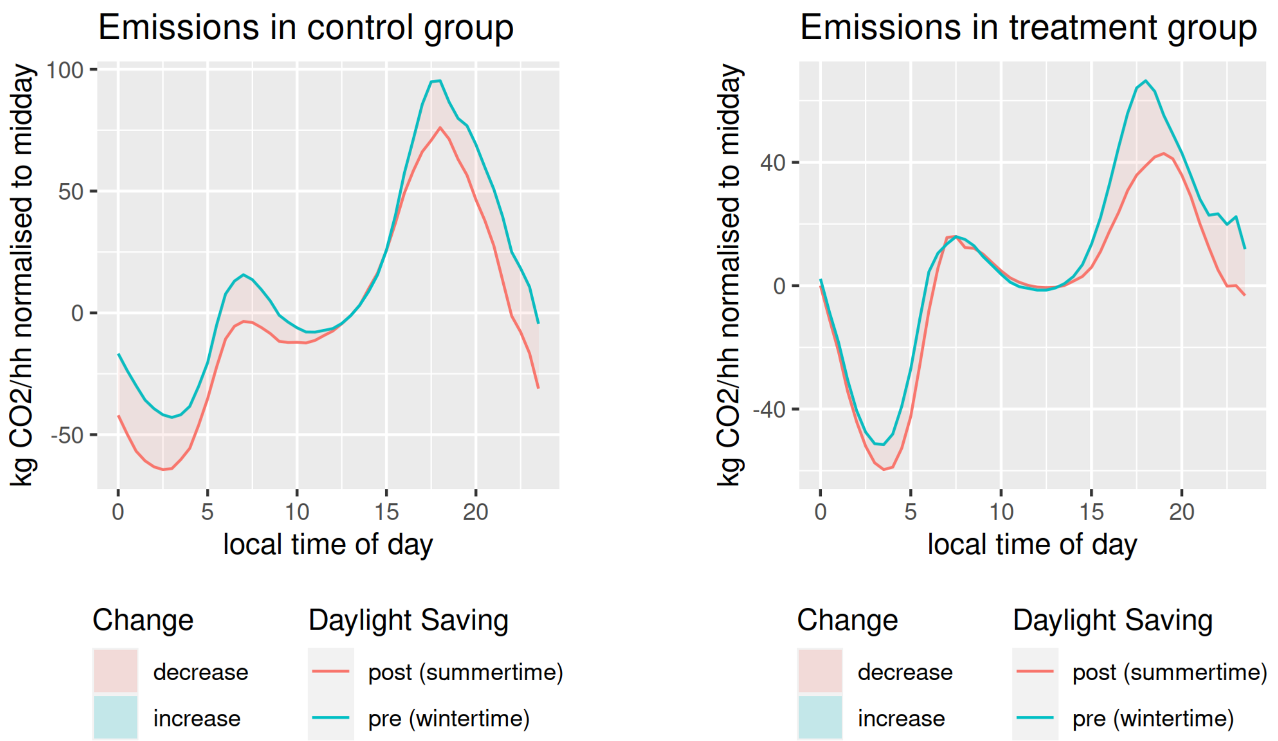


(a) Average intraday emissions in the control region



(b) Average intraday emissions in the treated regions

Figure 8: Average intraday emissions, with and without DST, for treated and control. The fact that there is more of the red (decrease) area and less of the green (increase) for the treated region than the control suggests that DST does reduce emissions. However this effect disappears once accounting for controls.



(a) Average intraday emissions in the control region

(b) Average intraday emissions in the treated regions

Figure 9: Average intraday emissions, normalising for midday. This is the same as Figure 8, except the average of the emissions for that day for that region was calculated for 12:00-14:30, and that was subtracted from all intervals for that day. This graphically represents the third differencing step. The fact that the red area (decrease) is larger for the control group suggests that DST increases emissions. However once controls are accounted for in the regression, this difference changes sign and becomes statistically insignificant.

Table 2: Summary statistics for all variables, for treatment and control states

	Variable	[unit]	Mean	Minimum Value	Maximum Value
Treatment	Population	$[inhabitans]$	4,127,718	505,468	8,806,160
	Greenhouse gas emissions per capita	$[kg/halfhour])$	0.315	0.000	0.935
	Energy production per capita	$[kWh/halfhour]$	0.533	0.001	41.719
	Temperature	$[°C]$	20.97	7.20	46.40
	Solar exposure	$[kWh/m^2]$	4.75	0.11	9.80
	Wind	$[km/h]$	16.09	0.00	51.50
	Holiday	$[\%ofdays]$	3.41	3.15	3.57
	Weekend	$[\%ofdays]$	0.29		
Control	Population	$[inhabitans]$	4,870,395	4,350,135	5,459,413
	Greenhouse gas emissions per capita	$[kg/halfhour]$	0.591	0.277	0.911
	Energy production per capita	$[kWh/halfhour]$	0.656	0.375	1.027
	Temperature	$[°C]$	26.64	12.60	41.20
	Solar exposure	$[kWh/m^2]$	6.13	0.50	9.27
	Wind	$[km/h]$	16.37	3.60	41.00
	Holiday	$[\%ofdays]$	3.27	3.27	3.27
	Weekend	$[\%ofdays]$	0.29		

Table 3: CO_2 and Electricity Consumption Results - DiD

	Dependent variable:			
	kg CO2 p.c.	kWh energy p.c.	Kg CO2 p.c.	kWh energy p.c.
	(1)	(2)	(3)	(4)
Treatment	-0.133*	-0.180**	-0.144**	-0.215***
	(0.0391)	(0.0284)	(0.0281)	(0.0198)
Post	0.0243***	0.0317***	0.0263	0.0315
	(1.50e-14)	(6.03e-15)	(0.0174)	(0.0174)
Treatment × Post	-0.0330**	-0.0568**	-0.0181	-0.0235
	(0.00663)	(0.00920)	(0.0222)	(0.0161)
Weekend			-0.0338***	-0.0462***
			(0.00132)	(0.00181)
Public Holiday			-0.0378**	-0.0498***
			(0.00556)	(0.00387)
Temperature			-0.0109	-0.0250
			(0.0142)	(0.0168)
Temperature^2			0.000272	0.000521
			(0.000235)	(0.000289)
Solar Exposure			-0.00422	-0.00815
			(0.00631)	(0.00600)
Wind[3]			-0.0594	0.00417
			(0.0249)	(0.0119)
Constant	0.576***	0.656***	0.736*	1.008*
	(1.41e-14)	(6.97e-15)	(0.215)	(0.265)
r2	0.152	0.326	0.213	0.378
r2_a	0.152	0.326	0.213	0.378

Note: Errors clustered by region, weighted by population.
Dependent variables are measured per half hour.
Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Results For CO2 and Electricity Consumption DDD With Controls

	Dependent variable:	
	kg CO2 p.c.	kWh energy consumption p.c.
	(1)	(2)
Treatment	-0.134^{*}	-0.216^{***}
	(0.0306)	(0.0216)
Post	0.0605^{*}	0.0788^{*}
	(0.0174)	(0.0174)
Treatment × Post	-0.0303	-0.0401
	(0.0186)	(0.0196)
Not Midday	0.0272^{***}	0.0143^{***}
	$(8.78e-13)$	$(1.00e-12)$
Treatment × Not Midday	-0.0115	0.000875
	(0.00874)	(0.00608)
Post × Not Midday	-0.0382^{***}	-0.0529^{***}
	$(2.00e-12)$	$(2.33e-12)$
Treatment × Post × Not Midday	0.0136	0.0185
	(0.0129)	(0.0110)
Weekend	-0.0338^{***}	-0.0462^{***}
	(0.00132)	(0.00181)
Public Holiday	-0.0378^{**}	-0.0498^{***}
	(0.00556)	(0.00387)
Temperature	-0.0109	-0.0250
	(0.0142)	(0.0168)
Temperature^2	0.000272	0.000521
	(0.000235)	(0.000289)
Solar Exposure	-0.00422	-0.00815
	(0.00631)	(0.00600)
Wind[3]	-0.0594	0.00417
	(0.0249)	(0.0119)
Constant	0.712^{*}	0.995^{*}
	(0.215)	(0.265)
r2	0.214	0.380
r2_a	0.214	0.380

Note: Errors clustered by region, weighted by population.
Dependent variables are measured per half hour.
*p<0.1; **p<0.05; ***p<0.01

Table 5: Results For CO2 and Electricity Consumption Base DDD Without Controls

	Dependent variable:	
	kg CO2 p.c.	kWh energy consumption p.c.
	(1)	(2)
Treatment	-0.123^{*}	-0.180^{**}
	(0.0435)	(0.0256)
Post	0.0585^{***}	0.0790^{***}
	$(1.26e-13)$	$(2.01e-13)$
Treatment × Post	-0.0451^{***}	-0.0734^{**}
	(0.00506)	(0.0142)
Not Midday	0.0272^{***}	0.0143^{***}
	$(1.28e-13)$	$(1.27e-13)$
Treatment × Not Midday	-0.0115	0.000858
	(0.00875)	(0.00608)
Post × Not Midday	-0.0382^{***}	-0.0529^{***}
	$(1.67e-13)$	$(2.29e-13)$
Treatment × Post × Not Midday	0.0136	0.0185
	(0.0129)	(0.0110)
Constant	0.552^{***}	0.643^{***}
	$(1.28e-13)$	$(1.34e-13)$
r2	0.153	0.328
r2_a	0.153	0.328

Standard errors in parentheses.

Dependent variables are measured per half hour.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6: Results For ln(CO2) and ln(Electricity Consumption) DDD With Controls

	Dependent variable:	
	ln(kg CO2 p.c.)	ln(kWh energy consumption p.c.)
	(1)	(2)
Treatment	-0.351^{*}	-0.409^{***}
	(0.124)	(0.0464)
Post	0.130	0.108^{*}
	(0.0509)	(0.0280)
Treatment × Post	−0.0490	−0.0566
	(0.0524)	(0.0337)
Not Midday	0.0569^{***}	0.0224^{***}
	$(3.57e-12)$	$(1.43e-12)$
Treatment × Not Midday	0.00157	0.0116
	(0.0291)	(0.0131)
Post × Not Midday	-0.0671^{***}	-0.0764^{***}
	$(6.78e-12)$	$(3.25e-12)$
Treatment × Post × Not Midday	0.0186	0.0153
	(0.0311)	(0.0242)
Weekend	-0.0774^{**}	-0.0947^{***}
	(0.0124)	(0.0109)
Public Holiday	-0.0900^{*}	-0.110^{**}
	(0.0226)	(0.0155)
Temperature	−0.0178	−0.0402
	(0.0352)	(0.0225)
Temperature^2	0.000501	0.000872
	(0.000578)	(0.000380)
Solar Exposure	−0.0171	−0.0148
	(0.0204)	(0.00913)
Wind^3	−0.205	0.00503
	(0.102)	(0.0206)
Constant	−0.278	0.111
	(0.525)	(0.366)
r2	0.171	0.400
r2_a	0.171	0.400

Note: Errors clustered by region, weighted by population.
Dependent variables are measured per half hour.
*p<0.1; **p<0.05; ***p<0.01

Table 7: Results For CO2 and Electricity Consumption DDD With Controls - Last 5 Years

	Dependent variable:	
	kg CO2 p.c.	kWh energy consumption p.c.
	(1)	(2)
Treatment	-0.137^{*}	-0.209^{***}
	(0.0324)	(0.0237)
Post	0.0614^{*}	0.0747^{*}
	(0.0170)	(0.0176)
Treatment × Post	-0.0609^{**}	-0.0680^{*}
	(0.0128)	(0.0175)
Not Midday	0.0792^{***}	0.0472^{***}
	$(3.65e-13)$	$(5.65e-13)$
Treatment × Not Midday	-0.0311	0.00107
	(0.0168)	(0.00890)
Post × Not Midday	-0.0355^{***}	-0.0496^{***}
	$(5.61e-13)$	$(8.43e-13)$
Treatment × Post × Not Midday	0.0230	0.0322^{*}
	(0.0127)	(0.0115)
Weekend	-0.0263^{***}	-0.0368^{***}
	(0.00125)	(0.00167)
Public holiday	-0.0288^{*}	-0.0407^{**}
	(0.00777)	(0.00594)
Temperature	-0.00144	-0.0216
	(0.0111)	(0.0168)
Temperature^2	0.000112	0.000471
	(0.000184)	(0.000293)
Solar Exposure	-0.00724	-0.00978
	(0.00641)	(0.00595)
Wind[3]	-0.0471	0.0152
	(0.0243)	(0.0128)
Constant	0.477^{*}	0.869^{*}
	(0.167)	(0.264)
r2	0.424	0.445
r2_a	0.424	0.445

Note: Errors clustered by region, weighted by population.
Dependent variables are measured per half hour.
*p<0.1; **p<0.05; ***p<0.01

Appendix B: Data column explanation

This section describes the dataset used, after all the joins, transformations and enrichment are performed.

For the independent variable, (y), there is:

`co2_kg_per_capita` kilograms of CO_2 emitted, per capita in this region (accounting for population growth over time), within this time interval. (e.g. within half hour for the half-hour file, or within the day for the daily file.)

`energy_kwh_per_capita` kilowatt hours of energy consumer, per capita in this region (accounting for population growth over time). Note that rooftop solar is counted as negative load by AEMO. So 10kWh of load plus 3kWh of solar appears here is 7kWh.

For the dependent variables (x) there is:

`regionid` the geographical state, as per AEMO convention. (AEMO always ends region id with a `1`) This is a string enum/factor. Options are:

- `QLD1` Queensland (our control region)
- `NSW1` New South Wales. (This includes the Australian Capital Territory (ACT))
- `VIC1` Victoria
- `SA1` South Australia
- `TAS1` Tasmania

`dst_now_anywhere` "post" - dummy variable - is there daylight saving in this time interval. Even in the control region this is true.

`dst_here_anytime` "treatment" - dummy variable - is this a region which has daylight saving. True even if there is not daylight saving in this time interval. Note that this value changes at midnight, even though in theory DST transitions happen at 2am or 3am. In practice everyone changes their clocks before going to bed, so we don't expect these 5 hours per year to introduce much error. It simplifies the code and graphs to think of the 'post' as applying to a whole date.

`dst_now_here` treatment x post - dummy variable. True if there is daylight saving in this region, on this day

`midday_control` a dummy - true if this half hour falls within 12:00-14:30. This is used for the third difference in our DDD. 12:00-14:30 was chosen to match Kellogg and Wolff, 2008. 12:00-14:30 is Queensland time (no DST) not local time.

`midday_control_local` Same as `midday_control`, but calculated based on local time in this region

`days_into_dst` How far are we into the daylight saving period?

- On the day when the clocks are moved forward, this is 0.
- The day after the clocks moved forward, it is 1.
- In the middle of summer it is around 90.
- The day before clocks move back (the last day with daylight saving) this is 0
- the day the clocks move back, this is -1
- the day after the clocks move back, this is -2
- in the middle of winter, this approximately -90
- the day before the clocks move forward in spring, this is -1

Our other time variables are:

`Date` the date of the observation. (First letter capitalised to avoid a namespace clash with R's `date` function)

`public_holiday` dummy variable for if this date is a public holiday in this region.

`hh_end` The datetime of the end of this half hour (when we have one row per half hour). The timezone is Queensland time (UTC+10, Australia/Brisbane, no daylight saving) even if this row is for a different region.

`hh_start` the start of this half hour period `dst_date` The date of the nearest daylight saving transition (which may be in the future or the past). Note that all treatment regions move their clocks on the same day. So the value is the same for all regions on a given day. Even for the control region (Queensland) this value is populated.

`dst_direction` A string factor/enum about the direction of the clock change at `dst_date`. Either `start` (move clocks forward, in October, spring) or `stop` (move clocks back, in Autumn).

`dst_transition_id` A unique string to represent each clock transition. e.g. `2009-start`, `2009-stop`. This is a string identifier for `dst_date`.

`days_before_transition` The number of days before the nearest clock change. If the nearest clock change is in the past, this is a negative number.

`days_after_transition` The number of days since the nearest clock change. If the nearest clock change is in the future, this is a negative number.

`dst_start` a dummy variable, for if `dst_direction` == `start`

`after_transition` a dummy variable. True if the most recent clock change is closer to the current date than the upcoming clock change

`days_into_dst_extreme_outlier` dummy variable - clock changes always happen on a Sunday morning. It's not on the same calendar day each year. Thus there are slight variations in the number of days between clock changes. There is one year which has one more day between the clock changes than other years. For that day only, this column is true. This is just to reflect the fact that for this value of `days_into_dst`, we only have one day of observations. We use this column to exclude this outlier from some graphs. But we do not exclude it from the actual regressions.

`days_into_dst_outlier` Similar to the previous variable, except this one is true for a few days across the time period. True if this value of `days_into_dst` is so large that it does not occur in some years. Once again, we may use this to exclude outliers from graphs, but not for the regression itself.

`day_of_week` integer - 1=Sunday, 0=Monday, ... 7=Saturday (Because that's what `lubridate::wday` does)

`weekend` dummy variable

`dst_transition_id_and_region` a concatenation of `dst_transition_id` and `regionid`. e.g. `2009-start-NSW1`. Useful when playing around with error clustering, fixed effects etc.

`hr` a float/decimal number representing the hour. e.g. 1:30pm-2pm is `13.5`

`hh_end_local` a datetime for the end of this half hour, in the local timezone of each region.

`hh_start_local` a datetime for the start of this half hour, in the local timezone of each region.

`date_local` same as `Date`, but calculated based on the local time in this region. (i.e. date changes one hour sooner in treatment regions during daylight saving)

Our controls are:

`rooftop_solar_energy_mwh` AEMO tends to report rooftop solar generation as negative load, mixed in with actual load. (Because they can't actually measure it.) For some years we are able to separately obtain it from AEMO's estimates. However this is only from 2016 onwards, so this column was not used for the main analysis. Units are megawatt hours.

`population` number of people in this region. This varies over time. The data source uses 3 month data, which we linearly interpolate. These might be a fraction of a person just due to the arithmetic of interpolation. Whilst population growth tends to be exponential, over a 3 month period linear is a sufficient approximation. This comes from the public website of the Australian Bureau of Statistics.

`temperature` maximum temperature each day, in each region, in degrees C. (We use maximum not average, because that tends to be a more representative driver of air conditioner load in summer.) For each region, we choose a weather station approximately in the biggest metropolitan area of the region, as this is the point where the largest demand for heating/air conditioning exists. All Data from the public website of the Bureau of Meteorology. In Detail:

SA1 : weather station 23034, Adelaide

QLD1 : weather station 40913, Brisbane

TAS1 weather station 94029, Hobart

VIC1 weather station 86038, Melbourne

NSW1 weather station 66037, Sydney

`solar_exposure` Amount of sun irradiance, measured in kWh/m^2, in this region for this day. (Not for this particular half hour.) For each region, we choose a weather station approximately in the middle of the region, as (solar energy) production is likely to be in less densely inhabited places. This data is from from The Bureau of Meteorology:

VIC1 : weather station 81123, Bendigo

SA1 weather station 16007, Cooberpedy

NSW1 weather station 65070, Dubbo

TAS1 weather station 94193, Hobart

QLD1 weather station 30045, Richmond

`wind_km_per_h` average wind speed, measured in km/h. For each region, we choose a weather station approximately in the middle of the regions, as (wind energy) production is likely to be in less densely inhabited places. Relevant for estimating potential wind turbine power generation. Standard physics theory (and personal experience) tells us that wind farm output is proportional to wind speed cubed. This data was obtained from Willy Weather, 2024. Some specific weather stations differ to those used for solar. This is because of differences in historical measurement availability.

VIC1 : weather station 411, Bendigo

SA1 weather station 133, Cooberpedy

NSW1 weather station 340, Dubbo

TAS1 weather station 501, Hobart

QLD1 weather station 236, Longreach

`total_renewables_today_mwh` Megawatt hours - "non-scheduled generation" (i.e. wind and solar) forecast, from AEMO, table `DISPATCHREGIONSUM` column `TOTALINTERMITTENTGENERATION`.

`total_renewables_today_mwh_uigf` Megawatt hours - another forecast of "non-scheduled generation" (i.e. wind and solar) from AEMO, table `DISPATCHREGIONSUM` column `UIGF`.

For the raw AEMO data, the meaning of each column is documented in the the MMS Data Model Report.

Appendix C: Data Wrangling Explanation

The dataset we downloaded is 300GB of compressed CSV files, totalling 1.4TB when uncompressed. Handling datasets larger than the size of your hard drive, with individual files larger than memory, is quite a technical challenge. As an example of the challenge, note that when processing this data with a new big-data library (Polars, 'the new Pandas'), the tool threw segmentation fault errors, because the dataset is too large for this big-data tool. This is noteworthy because whilst Polars is used from Python, it is written in Rust, and therefore segmentation faults are theoretically impossible. (The bug has since been fixed, after we reported it on GitHub.)

The main dataset comes from AEMO. Their target audience are electricity industry participants (e.g. coal generators), who generally want to know everything about the market. So the data is somewhat mixed together. e.g. individual files telling you total energy (MWh) for a region also tell you the total price, and the marginal cost of electrical transmission constraints, and FCAS ancillary service charges and so on. We were able to identify about one third of files as being definitely not needed (e.g. ones about gas) prior to downloading them. Of the remainder we need to download them, unzip them, "split" them (e.g. split rows of energy data from price data within the same file), and only then can we figure out which of the final 300 "tables" they belong to. At that point we can discard most data.

Market participants use proprietary software from AEMO to download the data, process it and write it into an Oracle database. This proprietary software is no longer publicly available.[11] The infrastructure cost of running such a large database is in the region of €10,000 per year.

[11] The public binary was taken down when the log4j vulnerability (`CVE-2021-44832`) became known. The subsequent version from AEMO which was patched was not released publicly.

Such transactional row-based databases are optimised for operational queries, not analytical queries. For these reasons a new pipeline was written for this project from scratch, to process the CSV files into parquet files.

AEMO publishes the files publicly on their website. Downloading the files is surprisingly complicated due to throttling, corrupt files and other issues.

The raw files are zips of zips of CSV files. However the CSV files are not standard tabular files. They are a concatenation of tables into one file. This bespoke format is unique to AEMO. The documentation is no longer on their website, but can be found in the Internet Archive. After decompressing and filtering, what remains is a large number of small files. Due to the schema changing over time, files which are part of the same dataset do not have the same columns. Note that the python library Pandas cannot handle empty null values for integer cells. However more sophisticated tools such as Arrow cannot implicitly merge integers and floats into a single datatype. Metadata about the dataset (the names and data types of each column, in each "table") was webscraped from AEMO's online documentation, to allow for explicit schema handling across the hundreds of tables. Two errors in the schema documentation were identified. AEMO has not responded to our reports about the errors.

For each of the hundreds of "tables" in the AEMO dataset, a single parquet file is produced. Parquet is an alternative format to CSV. The main motivation for using it was as a technical solution to keep the dataset small and fast. (e.g. it allows us to use predicate pushdown in subsequent scripts.) However the largest of the files relevant to us is a 5GB parquet file. When loaded into memory (e.g. in R) this would take up about 20GB. None of our laptops have that much memory. The processing we want to do is to take 5-minute data per generator, multiply it by the constant CO2 emissions factor, sum within each region, aggregate to half hour intervals. Then it's small enough to join with some other data, e.g. to account for inter-region import-export. One challenge is that AEMO's files contain duplicate data. (e.g. they have 5-minute files, and daily summaries of those files, and monthly summaries of those, etc.) Deduplicating data generally requires loading the whole thing in memory. So this is a really hard big data task. What we do is use Apache Arrow to lazy-load the parquet files, such that we can use filter and predicate pushdowns into the storage layer, along with physical repartitioning, to end up with something small. From then onwards normal R joins can be used to add additional datasets.

For data about DST transitions itself, we need to know what days the clocks move. We also want some enriched data about this. e.g. for each calendar day, is the nearest clock change in the future, or past? How many days away? We did not download this data from anywhere. Python itself has a copy inside it, which it uses for timezone conversions of datetimes. We use that instead of downloading, because it's easier and less likely to have mistakes than manually downloading and combining many CSVs.

Appendix D: Inter-region flow emission adjustment calculation

In the data section, we explained that we might have inter-region flows between Queensland and the treated regions. We adjust for this by calculating adjusted emissions taking into account between-region import and export data.

A substantial fraction (7%) of energy in the NEM is exported across region borders. For some interconnectors, the flows do not average to zero. As one of the data preparation steps mentioned in Section 3.2, these flows are adjusted for. For example if Queensland generates 3 GW of power and 3000 tonnes of CO2e, and exports 1 GW to New-South-Wales, we subtract 1000 tonnes of CO2 from Queensland's emissions and add it to New South Wales' emissions. That is, we do not use a marginal approach, but an average one. Calculating the marginal impact of energy export on emissions is extremely challenging, and requires expensive proprietary solvers to re-run AEMO's dispatch optimisation.

Note that whilst the generation data is available with 5 minute granularity, interconnector data is only available with a 30 minute granularity (for most of the time period studied). That is one reason why we downsample to 30 minute frequency. A 30 minute frequency also aligns with the literature (Kellogg and Wolff, 2008), make the dataset set manageable, reduces the impact of serial correlation, and align to market trading periods.