

## What is the relationship between household wealth and rooftop solar in Australia?

VEPC Working Paper 21-11  
November 2021

Bruce Mountain, Victoria University\*  
Kelly Burns, Victoria University  
Ben Willey, Victoria University

### Abstract

Around one in three Australian homes have installed rooftop solar photovoltaics. The relationship between rooftop solar and household income or wealth has been a focus of research in Australia and in other countries. Studies in Australia using survey data tend to the conclusion that rooftop solar uptake is positively associated with wealth, although not necessarily income. We critique those studies and suggest the conclusion on the relationship between wealth and solar installation is not robust. In fact, using data from customers' electricity bills we find a negative relationship between wealth and rooftop solar. However, we do not think our models are sufficiently robust to reach a firm conclusion. A richer dataset is needed. Continued effort in this area will be valuable considering the possibility that distributed (behind-the-meter) storage will increasingly be paired with rooftop solar. This will potentially have a large effect on the recovery of shared network costs. If it is the case that richer households are able to opt-out of increasingly expensive grid-supplied electricity, regressive impacts may become significant.

### Highlights

- The relationship between household wealth and rooftop solar is not well understood because it is difficult to model robustly using available data.
- Claims that wealth is positively related to solar uptake are not robust when data is segmented to take into account the differences in the barriers faced by owners and renters, and homes with and without shared roof space.
- Analysing data from customers' bills we find that for owner-occupied detached houses wealth and solar uptake is negatively related, although this analysis is not sufficiently robust to reach a firm conclusion.

**Keywords:** Rooftop solar, wealth, fairness, equity, just transition, residential

**JEL:** D10, C53, Q40, Q42

**DOI:** <https://doi.org/10.26196/qvgk-dx81>

### Acknowledgements

We thank Gasseem without whose data science, website scraping and PDF parsing skills this research would not have been possible. We also acknowledge customer group **CHOICE** for the provision of processed electricity bill data.

### Declarations

Bruce Mountain and Amine Gasseem developed price comparison software that was used by customer group **CHOICE** to extract data from electricity bills and to price those bills.

### Funding

This work was supported by a multi-year funding grant from the Government of Victoria for the establishment of the Victoria Energy Policy Centre.

\* The corresponding author's contact is: [bruce.mountain@vu.edu.au](mailto:bruce.mountain@vu.edu.au)

## 1. Introduction

Across the south and eastern states of Australia, about one in three houses and small businesses (2.6m) have now installed rooftop photovoltaics. In total these systems are capable of providing around 12 GW of power at their peak on a sunny day.<sup>1</sup> These solar homes produce around 19 TWh of electricity each year (about 10% of end-use electrical demand). In some regions, solar panels are installed in more than one in two eligible roofs. As far as we know this level of distributed (behind the meter) solar installation is higher than in other large developed countries.

While rooftop solar was initially heavily subsidised by tax payers and electricity consumers, these subsidies declined quickly when solar costs declined and grid-supplied electricity prices rose (Mountain & Szuster, 2015). In some regions, rooftop solar continues to receive significant public subsidy, albeit means tested.<sup>2</sup>

While rooftop solar retains political and community support, energy regulators and some community groups have increasingly drawn attention to their equity concerns.<sup>3</sup> Their arguments have focused, mainly, on the claim that solar homes may in future impose costs on electricity networks that those solar homes, rather than all consumers should pay for. However, some welfare advocates also contend that there is a positive relationship between wealth and rooftop solar.<sup>4</sup> Rooftop solar, and its continued policy support, has therefore become topical in the context of the broader debate on a “just transition” (i.e. socially progressive decarbonization policy). These issues are topical elsewhere too. For example, in the United States (Welton & Eisen, 2019) suggest that the acceleration of equitable rooftop PV has an important role to play in clean energy justice. Similar arguments can be found with respect to the Global North (Carley & Konisky, 2020) and California (Lukanov & Krieger, 2019).

With the highest country-level residential rooftop PV uptake that we are aware of, the relationship between wealth and PV uptake is topical. We replicate and critique the existing literature, which use Australian Bureau of Statistics (ABS) survey data. The ABS survey data provide a rich source of information on the households’ socio-economic characteristics (such as net wealth and income) but limited information on household electricity consumption, solar production or prices. We also analyse

---

<sup>1</sup> <http://www.cleanenergyregulator.gov.au/RET/Forms-and-resources/Postcode-data-for-small-scale-installations#Smallscale-installations-by-installation-year>

<sup>2</sup> In Victoria for example see: <https://www.solar.vic.gov.au/solar-panel-rebate>

<sup>3</sup> See for example: <https://www.aemc.gov.au/rule-changes/allowing-dnsp-charge-exports-network>

<sup>4</sup> For example: [https://www.aemc.gov.au/sites/default/files/documents/consultation\\_paper\\_-\\_der\\_integration\\_-\\_updating\\_regulatory\\_arrangements\\_1.pdf](https://www.aemc.gov.au/sites/default/files/documents/consultation_paper_-_der_integration_-_updating_regulatory_arrangements_1.pdf)

a sample of electricity bills from 6,067 houses across the south and eastern states of Australia. The bills provides a rich source of information on the prices households actually pay for grid purchases and receive for solar exported to the grid, and the amount of electricity consumed (grid consumption and self-consumption). Unlike the ABS survey data, our electricity bill data can adjust for heterogeneity in electricity prices and end-use consumption, in assessing the relationship between wealth and solar uptake.

Unlike the existing studies, our analysis of bill data supports a conclusion that wealth and solar uptake are negatively associated. However, rigorous testing of our analysis as well as the literature that relies on ABS survey data leads us to the views that strong conclusions on the relationship between wealth and solar uptake are not possible. The existing studies and ours suffer from omitted variables. Our contribution to the literature is in clarifying that the relationship between wealth and rooftop solar uptake remains uncertain in Australia, and in identifying the sort of data that is likely to support a confident conclusion.

Section 2 provides background covering relevant literature and descriptive data. Section 3 and 4 present the methodology and results, respectively. Section 5 discusses the findings. Section 6 presents the main conclusions and policy implications.

## **2. Literature Review**

There is an extensive literature on the relationship between income or wealth and solar uptake, in the United States in particular. Low-to-moderate income households are less likely to adopt rooftop PV than high-income households (Carley and Konisky, 2020). However while solar adopters generally skew towards higher incomes, that trend continues to diminish over time and solar adopter incomes' vary considerably and encompass many low-to-moderate income households (Barbose et al., 2021). Demand-side explanations for adoption inequity include access to capital, renting rather than owning, building form, language barriers, race and ethnicity (Sunter et al., 2019). Supply-side explanations include income-targeted marketing (O'Shaughnessy et al., 2021). PV leasing and property-assessed financing have increased the diffusion of PV adoption among low and middle income households in existing markets and have driven more installations into previously underserved low-income communities (Barbose et al., 2021).

The Australian literature on the relationship between wealth/income and solar is more limited and focussed mainly on establishing whether there is a statistically significant relationship between wealth/income and solar uptake, using survey and census data. Best et al. (Best et al., 2019b) apply logit models to ABS data from the Survey of Income and Housing (SIH) survey 2015-16 and Household Energy Consumption (HEC) survey 2012 to identify economic, social and environmental factors that affect solar uptake (and the intention to install solar). They estimate solar uptake and the intention to install solar with different combinations of the various factors including wealth. They conclude that higher net wealth is generally associated with higher solar uptake (but not the intention to install solar); that income does not influence solar uptake, but that income does affect the intention to install solar.

Best et al. (Best et al., 2021) use household level data from the ABS SIH 2015-16 and 2017-18 to estimate a probit model for solar uptake as a function of solar subsidises, (log) wealth and (log) disposable income. They conclude that subsidies and wealth drive solar uptake. Consistent with Best et al. (Best et al., 2019b) they find income is not significant using the 2015-16 SIH, but is significant (at 10% level of significance) for the 2017-18 SIH. They recognise that their results may suffer from multicollinearity based on the “substantial correlations” between net wealth and other variables. Further critique and analysis of the Best et al. (Best et al., 2019b) and Best et al. (Best et al., 2021) follows in Sections 3 and 4.

(Best et al., 2019a) use a cross-sectional model to estimate, for each postcode in Australia, the proportion of households with solar installed and solar PV capacity. Using postcode level data, they match data on small-scale PV installations (Australian PV Institute, 2019) to ABS census data on number of dwellings, rented dwellings, flats and apartments, and household income (Australian Bureau of Statistics, 2018), as well as solar exposure data (Bureau of Meteorology, 2018) and additional individual taxation income data (Australian Taxation Office, 2018). They find subsidies, location, whether the home is rented and whether the dwelling is a flat or apartment are significant drivers of solar uptake at the postcode level. Household and personal income are also a significant drivers of solar uptake, except for homes with very low income (less than A\$20,799 p/a). However, Best et al. (2019a) do not include net wealth as an explanatory variable but do include superannuation as a proxy for accumulated capital. They find the proxy measure for accumulated capital is not a significant driver of solar uptake, but is a significant driver of solar PV capacity.

### **3. Data and methodology**

### **3.1 Revisiting Best et al. (Best et al., 2019b) and Best et al. (Best et al., 2021)**

Best et al. (Best et al., 2021) draw on the ABS SIH 2015-16 data to examine how solar uptake varies with household wealth across Australia. The Survey of Income and Housing (SIH) is conducted bi-annually with around 15,000 households across Australia and collects a rich source of socio-economic data. Best et al. (Best et al., 2021) do not segment the data by any cohort. Rather, they examine the proportion of all households that have solar across the ten net wealth percentiles. However, as we will show, wealth is strongly correlated to building form and ownership. We therefore segment the ABS SIH data into different cohorts according to ownership and building form, and re-examine how solar uptake varies with wealth.

Best et al. (Best et al., 2019b) similarly do not separate the data into owned homes and rented homes but instead account for ownership by including a flag for rental properties as a dependent variable in their econometric (logit) model. However, as we will show, rent is strongly correlated with wealth: rented homes dominate the lowest wealth deciles, conversely for owned homes. Consequently the model suffers from multicollinearity. To address this, we replicated the logit model in (Best et al., 2021) using the ABS SIH 2015-16 (Australian Bureau of Statistics, 2017) for all dwellings and then apply this model to segmented datasets based on ownership and building form. We compare the significance, sign and magnitude of the estimated coefficients for each model (as applied to all dwellings, owner-occupied, rentals, other dwellings and owner-occupied houses) to assess whether the findings presented by Best et al. (2021) are robust when data are appropriately segmented.

### **3.2 Our Data**

Australian customer group CHOICE<sup>5</sup> consented to our use of processed electricity bill data obtained from 10,050 unique households. These data were extracted from original PDF-format monthly or quarterly electricity bills that household electricity consumers – many of whom were members of CHOICE - had uploaded to CHOICE's website between May and November 2018, in order to make use of an online price comparison service "CHOICE TRANSFORMER". CHOICE TRANSFORMER offered repeated electricity price comparison and customer switching but required the payment of a subscription fee if available savings (by switching to cheaper offers) of more than \$100 per year were found. Customers were recruited mainly through email advertising to CHOICE's members and

---

<sup>5</sup> [www.choice.com.au](http://www.choice.com.au)

supporters. All households were in the contestable retail electricity markets of New South Wales (NSW), Victoria (VIC), South-East Queensland (QLD) or South Australia (SA).

The bill data includes information on billing address, postcode, retailer, network service provider, state, tariff type, grid consumption in the billing period (kWh), estimated annual bill amount (\$), volume of solar exported to the grid (kWh) and the solar feed-in price (cents per kWh).

The estimated annual bill calculated by CHOICE TRANSFORMER is based on the electricity prices and estimated annual consumption. We obtained data on the type of dwelling (fully detached home, semi-detached house, townhouse, apartment, unit, flat, etc) is obtained by scraping this information from the website [www.domain.com.au](http://www.domain.com.au).

Of the 10,050 household bills, 8,529 (85%) are on flat (single) rate tariffs, and 1,521 (15%) are charged on tariffs that vary by time of day and day of week. Time of use customers are excluded from this analysis.<sup>6</sup> Of the 8,529 records for homes on flat rate tariffs, we further exclude records where the dwelling structure could not be determined (1,064 records). Furthermore, since the focus of this research is on the relationship between wealth and solar in households that unambiguously have control over access to their own roofs (i.e. fully detached houses) we exclude a further 1,398 records that are likely to be semi-detached houses, townhouses, apartments, units and flats (noting these dwellings are more likely to be rented than owned, as discussed earlier). Our final sample comprises 6,067 fully-detached houses on flat rate tariffs that are very likely to be owner-occupied.

Households with solar are remunerated for their solar exports to the grid (at their feed-in rate) and also obtain value from their solar systems by using their solar-produced electricity to substitute for grid supply. To correctly estimate energy consumption, it is therefore necessary to estimate the total annual solar generation (this is not measured) and subtract the solar exports to find the residual amount of solar generation that is self-consumed. This is then added to the grid purchases to estimate the annual electricity consumed in solar homes. The methodology used to estimate gross solar production is explained in (Mountain et al., 2020).

---

<sup>6</sup> As the CHOICE dataset does not provide the price paid for grid purchases or the amount of electricity consumers (or solar exported to the grid for homes with solar) in the peak and off-peak period, the necessary data is not available to analyse homes charged a time-varying tariff. Evidence suggests the (vast) majority of Australian households are on flat rate tariffs we do not think this approach biases our analysis.

The daily fixed charge levied by the household electricity supplier is not available in the CHOICE household bill dataset. We estimate the annual fixed (supply) charges as follows: NSW: \$365, VIC: \$402, QLD: \$0.90 and SA: \$0.8, based on the approximate average of the Standing Offers in 2018.

To estimate the variable price (cents per kWh) price paid for grid-supplied electricity we use the estimated annual bill, add back the solar income (the product of the solar exports and feed-in rate), deduct the supply charge and divide the resulting numerator by the estimated volume of grid purchases.

Our data allows us to account for the effect of electricity bills on households' likelihood of installing solar. To do this it is necessary to estimate the adjusted annual electricity bill (in other words, what the bill would be if the household did not have solar). This is done by multiplying the estimated annual consumption (after grossing up for the volume of self-consumed solar) by the variable price and adding back the supply charge.

The households' postcode is matched to the 2016 postcode-specific decile Index of Relative Socio Economic Advantage and Disadvantage (IRSAD). The IRSAD ranking is used as a proxy for household wealth.

### **3.3 Preliminary data analysis**

Descriptive statistics on the sample of households used for this analysis are reported in Table 1. The distribution of our data by state is similar to the true population. Within our sample, 39% of fully detached houses are in NSW, 31% in Victoria, 21% in QLD and 10% in SA. The ABS Census of Population of Housing 2016 reports a similar distribution for fully detached houses across these four state (NSW 34%, Vic 31%, QLD 25%, SA 10%) (Australian Bureau of Statistics, 2018).

The proportion of fully detached homes within our sample with solar installed (19%) is also similar to that reported in the ABS SIH 2017-18 (Australian Bureau of Statistics, 2019). Specifically, of the 14,060 responses in the 2017-18 ABS SIH, 6,998 were identified as fully detached houses in the states included in our sample (NSW, Vic, QLD and SA) (Australian Bureau of Statistics, 2019). Of these, 22% (1,551) had solar installed. As an aside we note that the distribution of ABS SIH fully detached houses across these four states does not approximate the population (NSW 24%, Vic 28%, QLD 22%, SA 26%). As expected, adjusted annual bill is positively correlated to consumption (0.93) and price (0.22). However, price and consumption show almost no association (-0.05).

Table 1. Data summary

		All households	Non-solar households	Solar households
Number		6,067	4,918	1,149
Consumption (kWh p/a)	Median	6,608	6,300	7,949
	Average	7,858	7,504	9,373
	Min	50	50	638
	Max	48,993	38,064	48,993
Price (\$/kWh)	Median	0.23	0.23	0.24
	Average	0.24	0.24	0.26
	Min	0.00	0.00	0.09
	Max	1.62	1.62	1.23
Adjusted Annual Bill (\$ p/a)	Median	1,909	1,824	2,281
	Average	2,246	2,142	2,689
	Min	379	379	420
	Max	13,022	10,330	13,022
NSW		2,355	1,966	389
VIC		1,872	1,730	142
QLD		1,259	839	420
SA		581	383	198
Wealth (decile)	1	182	138	44
	2	219	178	41
	3	266	197	69
	4	377	297	80
	5	425	319	106
	6	520	381	139
	7	552	428	124
	8	819	650	169
	9	1,153	952	201
	10	1,554	1,378	176

### 3.4 Econometric model

Our goal is to quantify the relationship between the probability that a house has solar and the individual characteristics of the house. Our binary dependent variable ( $y_i$ ) is 1 (house has solar) or 0 (house does not have solar). Logit and probit models are suitable when attempting to model a dichotomous dependent variable as these methods fit a nonlinear function to that data to better enable a prediction the dependent variable equals 0 or 1 (by fitting an S-shaped curve, rather than a straight line, between outcomes 0 and 1). The logit model assumes there is a logistic distribution of errors, and the probit model assumes there is a normal distribution of errors. Logit and probit models yield similar results, however, we prefer the probit model because this enables marginal analysis of changes in the probability a household has solar (as opposed to changes in the log of the odds ratio the household has solar, as is the case for a logit model).



We model the probability a house has solar as  $Pr(y_i = 1|x_i, \beta) = 1 - \phi(-x_i'\beta)$ , where  $\phi$  is a continuous, strictly increasing function that takes a real value and returns a value ranging from 0 to 1. It follows that  $Pr(y_i = 0|x_i, \beta) = \phi(-x_i'\beta)$ . In a probit model, the value of  $\phi$  is the cumulative distribution function of the standard normal distribution.

We estimate the parameters for  $Pr(y_i = 1|x_i, \beta) = 1 - \phi(-x_i'\beta) = \phi(x_i'\beta)$  using the method of maximum likelihood. Using the CHOICE data set, the probit model to estimate the probability a house has solar installed is as follows:

$$Pr(y_i) = \alpha + \beta_1 AdjBill_i + \beta_2 QLD_i + \beta_3 SA_i + \beta_4 VIC_i + \beta_5 Wealth_i + \varepsilon_i \quad [1]$$

where:  $Pr(y_i)$  is the probability the house has solar.

$AdjBill_i$  is the adjusted annual total bill (but for the installation of solar and including fixed annual charges), as described above.

$QLD_i$  is a dummy variable for homes in QLD (takes values of 1 if dwelling is in QLD, 0 otherwise).

$SA_i$  is a dummy variable for homes in SA (takes values of 1 if dwelling is in SA, 0 otherwise).

$VIC_i$  is a dummy variable for homes in NSW (takes values of 1 if dwelling is in VIC, 0 otherwise).

$Wealth_i$  refers to the ABS IRSAD (takes values from 1 to 10).

We assess the robustness of model to fit our sample data and provide reliable predictions on the relationship between wealth and solar uptake using a variety of methods. First, standard t-testing determines if the estimated coefficients are statistically significant. A redundant variable test on the *Wealth* variable using the log likelihood ratio (LR) test statistic has a null hypothesis that wealth is a redundant variable, and the test statistic follows the  $\chi^2$  distribution with 1 degree of freedom. We also consider McFadden's (1974) "pseudo"  $R^2$ , a likelihood based statistic, to assess the adequacy of the probit model fit of the data. The pseudo  $R^2$  is calculated based on a likelihood ratio that calculates the improvement in the model over an intercept only model. Similar to the traditional  $R^2$ , a higher value indicates a better fit. However, pseudo  $R^2$  values are lower than the traditional  $R^2$  (as the model is fitting a binary rather than continuous dependent variable) and values between 0.2 to 0.4 are considered an excellent fit (McFadden, 1979).

We then evaluate the robustness of the probit model to predict (in-sample) whether a household has solar, using three established techniques. First, we generate the expectation-prediction (classification) table to estimate the correct and incorrect classifications based on a probability success

cutoff = 0.5. The classification table provide a contingency table of the predicted responses of the model classified against the observed dependent variable. The classification table also reports the results of the predicted responses of a constant probability model (only intercept term included). “Correct” classifications are obtained when the predicted probability is less than or equal to 0.5 and the observed  $y=0$ , or when the predicted probability is greater than 0.5 and the observed  $y=1$ . The weighted average of the percentage of times the model correctly predicts  $y=1$  and  $y=0$  is used to measure the overall proportion of times the model correctly predicts whether a house has solar installed. The difference in the expectation-prediction results for the two models is then compared to calculate the percentage increase in prediction accuracy between the informed (our estimated probit model) and uninformed (constant probability) models. However, an important limitation to the expectation-prediction accuracy is that it relates to the ability of the model to predict *all* observations in the overall sample (where  $y=1$  and  $y=0$ , based on the weighted average of correctly predicted outcomes) and does not separately measure the ability of the model to predict whether a home *has solar* or *does not have solar*. As a result, in samples (such as ours) where the data is heavily skewed towards homes that don’t have solar, the measure of overall accuracy is misleading. We therefore provide the overall weighted measure of the prediction-evaluation tests, but the accuracy of the model to predict the subsets of homes with and without solar that is most relevant.

We also perform two goodness-of-fit tests of the model predictions using the Andrews (1988) and Hosmer-Lemeshow (1989) tests. These tests assess how well observed events match the expected events within subgroups of the model population. First, the data is grouped into deciles based on the Hosmer-Lemeshow grouping (subgroups are based on probabilities in which the first group consists of observations with the lowest 10% of predicted probabilities, the second group consists of the observations with the lowest 20% of predicted probabilities, and so on). The tests compare the fitted expected values (where  $y=1$ ) to the actual values in each decile (for example, if observations in the first group have a predicted probability of 0.15 then a correctly specified model would estimate the likelihood solar is installed to be approximately 15%). If the difference in the actual and predicted probability of solar uptake across each decile is large, the model provides an insufficient fit of the data and should be rejected (the null hypothesis is that the model is correctly specified, the test statistic follows a  $\chi^2$  distribution with  $g-2$  degrees of freedom, and ( $g$ ) is the number of groupings (10)). The Andrews test performs a similar test to assess goodness of fit of the model predictions (using a similar data grouping structure). We report the p-values for both tests (a p-value at or below 0.100, 0.05 and 0.001 indicates the null hypothesis (the model is correctly specified) should be rejected at the 10, 5 and 1% level of significance, respectively).

### 3.5 Marginal wealth effects

Marginal wealth effects refer to the incremental change in the probability a home has solar when the household's wealth decile ranking changes by one. Marginal wealth effects cannot be estimated using the estimated coefficients of the probit model because the probability the home has solar is conditional upon the values of *all* other explanatory variables (unlike an ordinary least squares model) (IHS Markit, 2017). To examine the marginal effect of a change in one explanatory variable on the probability the home has solar, the model must be estimated by choosing values for all other explanatory variables. Common practice is to examine marginal effects of one explanatory variable by holding all other explanatory variables at the mean or median (IHS Markit, 2017). We examine the impact of marginal changes in wealth on the probability a house has solar by analysing marginal effects at the median adjusted bill (\$1,909) and median wealth decile (8) (Table 1). For robustness, we also examine marginal effects at the 20<sup>th</sup> and 80<sup>th</sup> percentile adjusted bill (\$1,238 and \$3,884, respectively).

### 3.6 Out-of-sample predictive accuracy

We investigate whether our probit model is effective at predicting solar uptake on unseen data. We also examine how the performance of our model compares to a logit model, as well as non-linear, tree-based models (Rahman & Fazle, 2011). Both the ABS SIH 2015-16 and **CHOICE** data sets are randomly partitioned into a *training set* consisting of 75% of the data and a *test set* consisting of the remaining 25%. Our models are trained on the training set, used to predict solar uptake for the test set, and scored with reference to the known test set.

Our comparison models are a decision tree, a random forest and a gradient boosted tree; all of which are commonly used in machine learning applications. A decision tree model (Quinlan, 1986) outputs a binary tree allowing an unseen data point to be classified by following a path from the root vertex of the tree to a leaf (a vertex of order one). A random forest (Ho, 1995) uses a collection of separately trained decision trees and classifies an unseen data point by choosing the class selected by most trees. Gradient tree boosting (Schapire, 2003) is another way of combining several trees to form a single classifier. Random forests are able to avoid the problem of overfitting that a single decision tree sometimes suffers from. In turn, gradient tree boosting (through the use of *lightgbm*) typically produces a higher accuracy model than a random forest when trained on the same data. More information on these techniques can be found in the Appendix.

We fit a decision tree, a random forest, and a logit model to the training sets using *scikit-learn*, an open source machine learning library available for the Python programming language. We fit a gradient boosted model to the training sets using *lightgbm*, a gradient boosting framework available

for Python that uses tree-based learning algorithms (Ke et al., 2017). We fit a probit model to the training sets using *statsmodels*, again an open source machine learning library available for Python.

For each model, we calculate *Accuracy* (the percentage of test cases the model classifies correctly) and *Balanced Accuracy* (the arithmetic mean of the class specific accuracies) (Liu et al., 2014). If the model performs equally well on either class, then the accuracy and the balanced accuracy will be equal. We then define a *robust* classification model as being one that achieves a balanced accuracy of 70% or more.

Since around one in five households in our sample have rooftop solar installed and since our training set and test set was selected randomly, this imbalance in the number of data points in each class is also reflected there too. A training set consisting of different numbers of representatives from each class may result in a model that is biased towards the majority class. As an extreme example, our model might assign every test case to the majority class and thereby achieve an accuracy equal to the percentage of test cases belonging to the majority class. In other words: an imbalanced training set may lead to inflated accuracy scores. To remedy this, we employ two techniques:

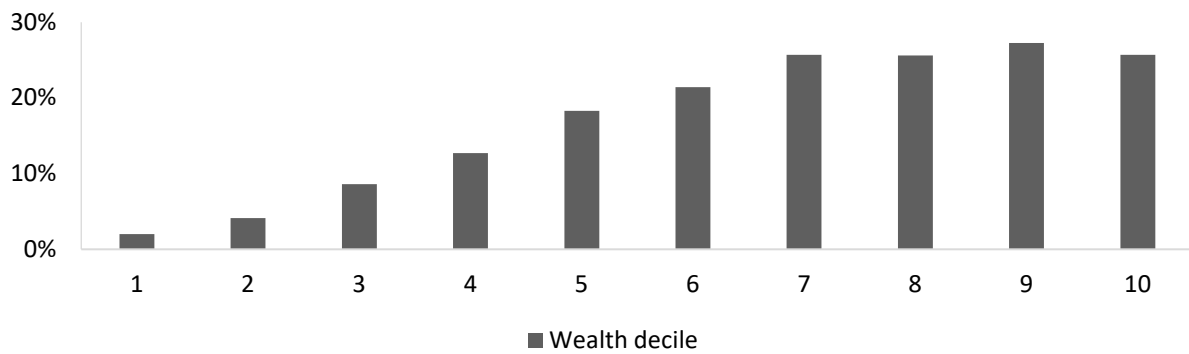
1. Oversampling the minority class in the training set using *SMOTE* (Synthetic Minority Oversampling Technique).
2. Choosing the balanced accuracy rather than the accuracy as our performance metric. This ensures that the performance metric values accuracy at predicting the majority and the minority classes equally (*Classification - Training a Decision Tree against Unbalanced Data - Cross Validated*, n.d.).

## 4. Results

### 4.1 Revisiting Best et al. (Best et al., 2019b) and Best et al. (Best et al., 2021)

In Figure 1, we use the ABS SIH 2017-18 to replicate the results of Best et al. (Best et al., 2021). Consistent with the results of Best et al. (Best et al., 2021), wealth appears to be positive associated with solar uptake. However, examining the distribution of solar uptake for all dwellings across wealth decile does not allow us to isolate the impact on wealth on solar uptake holding the interrelationship between ownership, building form and wealth constant. We now proceed to segment the SIH data and examine how solar uptake, building form and ownership vary with wealth.

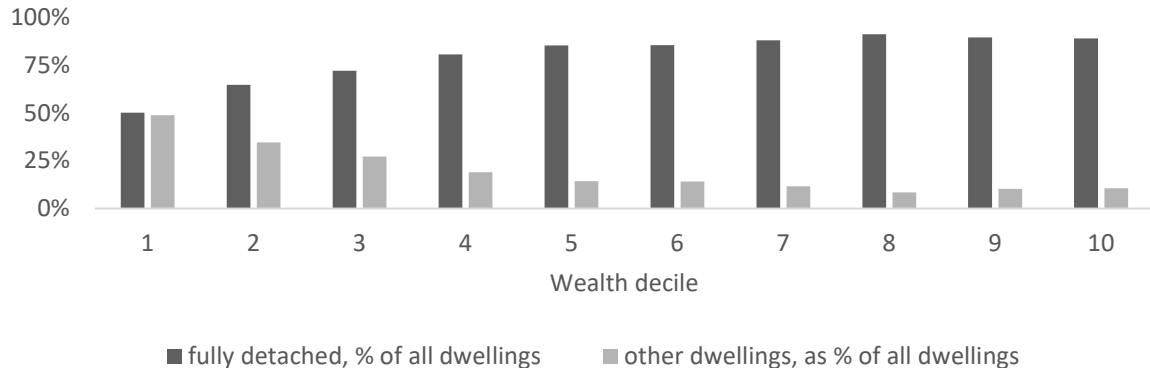
Figure 1. Proportion of homes with solar by wealth decile



Source: (Australian Bureau of Statistics, 2019)

Of the 14,060 respondents in the ABS SIH 2017-18 (Australian Bureau of Statistics, 2019), 11,207 resided in fully detached houses (80%) and the remainder split between semi-detached dwellings and dwellings with shared roof space (11% and 9%, respectively). As Figure 2 shows, building form and net wealth are skewed: fully detached houses skew towards upper wealth deciles while other dwellings with shared roof space (i.e. flats, units, apartments and semi-detached dwellings) skew towards lower wealth deciles.

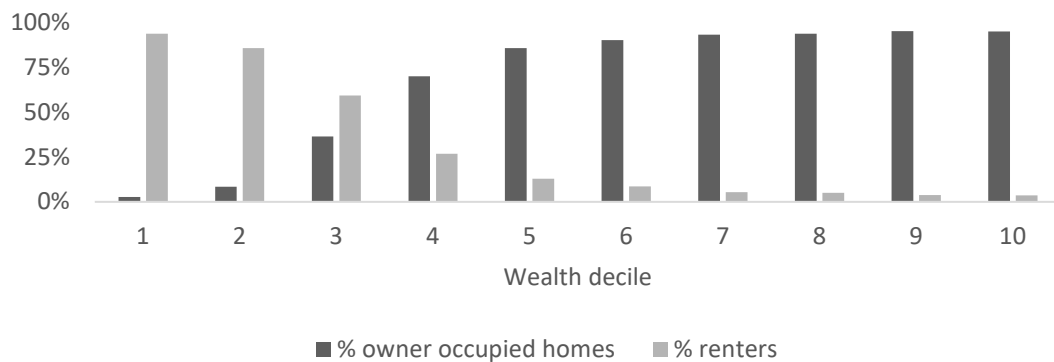
Figure 2. Proportion of fully detached and other dwellings across each wealth decile



Source: (Australian Bureau of Statistics, 2019)

The ABS SIH reports houses are more likely to be owner occupied (75% of fully detached houses are owner occupied, compared to 35% of other dwellings) and dwellings with less/shared roof space are more likely to be rented (62% of other dwellings are rented, compared to 23% of fully detached houses) (Australian Bureau of Statistics, 2019). Figure 2 shows that owner occupied homes are heavily skewed to the higher wealth deciles and renters are heavily skewed to the lower wealth deciles.

Figure 3. Proportion of owner occupied homes and renters across each wealth decile

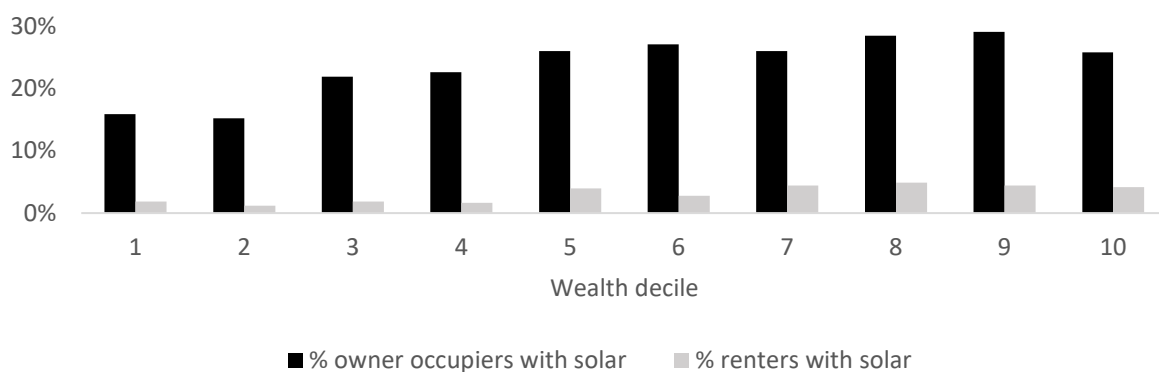


Source: (Australian Bureau of Statistics, 2019)

The ABS SIH 2017-18 data shows that the proportion of owned homes with solar is much higher than rented homes with solar (24% of owned homes have solar compared to 3% of rented homes (Australian Bureau of Statistics, 2019)). Figure 4 shows that when ownership is taken into account, the proportion of homes with solar does not vary markedly across wealth deciles (excluding the least wealthy homes in the lowest two deciles).

This difference in solar uptake among renters and owners is likely to be explained by transaction costs (high density rented properties are likely to require special arrangements to allocate solar production on a shared roof to individual renters); property form (rented properties are typically higher density and so less roof space in relation to floor space) and property rights (solar installation requires landlord approval and system ownership may be assigned to the landlord). Low solar adoption amongst renters may also result from “split incentives” (i.e. the proposition that landlords can not recover the cost of the solar system in higher rents). The notion of split incentives is however contested: (Wood et al., 2012) found no evidence of the split incentives, while Zander (2020) found the contrary.

Figure 4. Proportion of owner occupied homes and rented homes with solar



Source: (Australian Bureau of Statistics, 2019)

If we segment the ABS SIH 2017-18 data by building type as shown in Figure 3, we see that detached homes dominate the dataset, although semi-detached, row or terrace house or townhouse is not too far behind. A small fraction of the homes are flats, units or apartments. Solar penetration is clearly lower in the lower four wealth deciles for detached, semi-detached, row, terrace or townhouse. However there is no obvious trend of proportions of solar households for flat, unit or apartments across wealth deciles.

Figure 5. Proportion of dwellings with rooftop solar, by dwelling type and wealth decile



Source: (Australian Bureau of Statistics, 2019)

If we further segment the ABS SIH 2017-18 data into fully detached and owner occupied houses (Figure 6), we clearly see that wealth is strongly and positively associated with the likelihood of being an owner occupier of a fully detached house. Importantly, when we segment the data by ownership and building form, there is no observable relation between wealth and solar.

Figure 6. Proportion of owner occupied fully detached houses with solar by wealth decile



Source: (Australian Bureau of Statistics, 2019)

In the first column of Table 2, we replicate Best et al. (Best et al., 2019b) for all dwellings (17,429).<sup>7</sup> Columns 2 and 3 present the results for owner-occupied dwellings (12,417) and rented dwellings (5,012). As the sample of rented dwellings is biased towards fully detached houses (60%), we also replicate Best et al. (Best et al., 2019b) for “other” dwellings (i.e. dwellings other than fully detached houses) and this cohort is dominated by renters (61%). The results for “other” dwellings are presented in column 4. In the final column, we replicate Best et al. (Best et al., 2019b) for owner-occupied houses (10,792).

The model applied to all dwellings (first column in Table 2 ) produces the second highest pseudo  $R^2$  (0.14) and correctly estimates 83% of observations overall (comprising a 99% correct prediction of the 14,256 homes without solar and 5% correct prediction of the 3,039 homes with solar). The model also satisfies the Andrews and Hosmer-Lemeshow tests (p-values are above 0.100) indicating the model is correctly specified.

The estimated coefficients for solar uptake by owner-occupied dwellings are almost identical to those for all dwellings (second column in Table 2). However, the model does not perform as well when applied to owner-occupied dwellings only: the pseudo  $R^2$  falls by half (down from 0.14 to 0.07), the Andrews and Hosmer-Lemeshow tests find the model is not correctly specified, and the overall proportion of correctly predicted outcomes falls to 77% (comprising 99% of correctly predicted homes without solar and 6% correctly predicted homes with solar).

The estimated coefficients for rented dwellings only (third column in Table 2) are similar for wealth, income, building form, and number of bedrooms. However, occupancy (persons, employed persons, number of dependent children), pension income and number of credit cards are no longer found to be statistically significant. As was the case with owner occupied dwellings, the model when applied to rented dwellings does not perform as well: the pseudo  $R^2$  falls to 0.08, the model satisfies the Hosmer-Lemeshow test but fails the Andrews test (meaning we cannot be certain the model is correctly specified) and even though the expectation-prediction increases to 97% overall, the model does not accurately predict any rental home with solar within the sample (all of the 4,849 houses without solar are accurately predicted, whereas none of the 163 houses with solar are correctly predicted).

The model applied to other dwellings that are predominantly rented (fourth column in Table 2) produces substantively different results. Importantly, the estimated co-efficient on the wealth variable is not statistically significant. Although the model produces a markedly higher the pseudo  $R^2$

---

<sup>77</sup> We replicate model (5) from Table 2 (Best et al., 2019b) as this model is less likely to suffer from multi-collinearity (as there is no net wealth (log) squared term) and provides the second highest pseudo  $R^2$ . 2015-16 ABS SIH microdata are filtered in line with the methodology described in (Best et al., 2019b)(Best et al., 2021) .



(0.21) and is considered a good fit of the data, the model fails the Andrews test (suggesting the model is not correctly specified).

The model applied to owner-occupied houses (fifth column) indicates wealth is a positive and statistically significant driver of solar uptake. However, the magnitude of the effect is estimated to be around 25% lower than the wealth effect estimated by the original model (first column). Furthermore, the model is an extremely poor fit of the data (pseudo  $R^2$  0.06) and Andrews and Hosmer-Lemeshow tests both indicate the model is not correctly specified.

In summary, when the data are segmented to take into account ownership and/or building form, the Best et al. (Best et al., 2019b) model produces different results and the positive wealth effect does not necessarily hold. Furthermore, there is strong evidence to suggest the model does not provide a reasonable fit of the data and is not correctly specified. On this basis, we suggest the data and model used by Best et al. (Best et al., 2019b) do not produce sufficiently robust results to conclude there is a positive wealth effect.

Table 2. Best et al. (Best et al., 2019b) replication results for all dwellings and other dwelling cohorts

		All dwellings n=17,429	Owner occupied (all dwellings) n=12,417	Rented (all dwelling) n=5,012	Other dwellings (owner occupied and rented) n=3,243	Owner-occupied houses n=10,792
Net wealth, log	Co-efficient	0.19 ***	0.19 ***	0.18 **	0.17	0.14 ***
	Std. error	0.027	0.030	0.072	0.109	0.264
Income, log	Co-efficient	-0.01	-0.01	0.01	0.01	0.01
	Std. error	0.027	0.028	0.097	0.127	0.029
Private pension div. by income	Co-efficient	0.59 ***	0.58 ***	-0.07	0.45	0.59 ***
	Std. error	0.112	0.133	1.306	0.438	0.119
Long property tenure	Co-efficient	-0.22 ***	-0.23 ***		-0.14	-0.30 ***
	Std. error	0.059	0.059		0.397	0.061
Rent	Co-efficient	-1.60 ***			-1.90 ***	
	Std. error	0.107			0.376	
Apartment	Co-efficient	-1.82 ***	-1.86 ***	-1.47 ***		
	Std. error	0.246	0.279	0.524		
Bedrooms	Co-efficient	0.22 ***	0.20 ***	0.51 ***	0.59 ***	0.15 ***
	Std. error	0.030	0.031	0.109	0.145	0.033
Persons	Co-efficient	0.19 ***	0.21 ***	-0.07	0.58 ***	0.17 ***
	Std. error	0.033	0.034	0.125	0.152	0.036
Employed persons	Co-efficient	-0.13 ***	-0.15 ***	0.06	-0.42 ***	-0.11 ***
	Std. error	0.032	0.033	0.153	0.153	0.034
Dependent children	Co-efficient	-0.14 ***	-0.18 ***	0.15	-0.42 *	-0.14 ***
	Std. error	0.041	0.043	0.140	0.218	0.044
Mortgage	Co-efficient	-0.02	0.00		0.09	-0.11 *
	Std. error	0.057	0.058		0.267	0.062
Credit cards	Co-efficient	0.08 ***	0.09 ***	-0.02	-0.17	0.10 ***
	Std. error	0.023	0.024	0.091	0.105	0.024
Region fixed effects		Yes	Yes	Yes	Yes	Yes
McFadden's R-squared		0.14	0.07	0.08	0.21	0.06
Hosmer-Lemeshow test		0.170	0.010***	0.184	0.364	0.130***
Andrews test		0.153	0.004***	0.100*	0.000***	0.137***
Expectation-Prediction evaluation test (%)		82.66	77.04	96.75	95.81	75.17

Notes: An \*\*\*, \*\* and \* indicates statistical significance at the 1, 5 and 10 per cent level, respectively. P-values are reported for the Hosmer-Lemeshow and Andrews tests.

#### 4.2 Our estimation results

The estimation results and model diagnostics for equation [1] using the CHOICE data are reported in *Table 3*. Adjusted bill is statistically significant and positively related to the likelihood solar is installed.

Wealth is also statistically significant however the model estimates a negative relationship between wealth and solar uptake. All state dummy variables are also statistically significant.

Table 3. Equation [1] estimation results

Variable	Coefficient	Std. Error
C	- 0.88***	0.068
ADJBILL	0.00***	0.000
FLAGSA	0.48***	0.063
FLAGQLD	0.55***	0.048
FLAGVIC	- 0.36***	0.054
WEALTH	- 0.05***	0.007
McFadden's R-squared	0.09	
Redundant variable (LR) test	0.000***	
Hosmer-Lemeshow test	0.583	
Andrews test	0.499	
Expectation-Prediction evaluation test (%)	81.18	

Notes: An \*\*\* indicates statistical significance at the 1 per cent level. P-values are reported for the redundant variable, Hosmer-Lemeshow and Andrews tests.

McFadden's R-squared is low suggesting the model does not provide a good fit of the data. This score is comparable to the results produced by (Best et al., 2021) using a similar number of explanatory variables. The model satisfies the Andrews and Hosmer-Lemeshow tests, indicating the model is correctly specified. The classification table of the expectation-predication evaluation tests (where the probability success cutoff = 0.5) shows that our models correctly estimate 81% of the observations overall (comprising 99% of the 4,893 homes without solar and 3% of the 1,117 homes with solar). We reject the null hypothesis that wealth is a redundant variable using the LR test, confirming the relation between wealth as a driver of solar uptake.

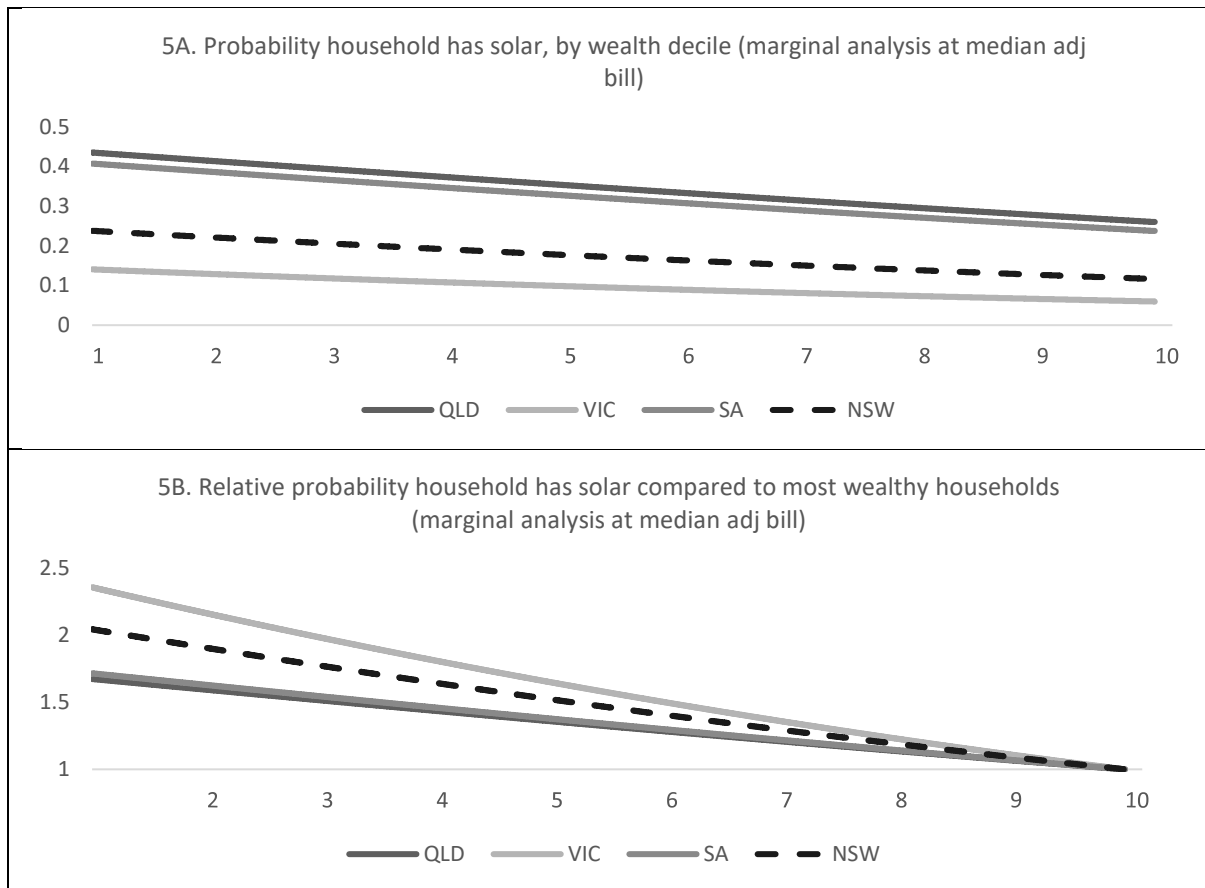
#### 4.3 Marginal wealth effects

On the basis of the estimated coefficients from equation [1], we estimate the probability a house has solar (assuming the house pays the 20<sup>th</sup>, 50<sup>th</sup> and 80<sup>th</sup> percentile adjusted bill, and separately for each state by setting the relevant state flags to one). As Figure 7 shows, there is a strong negative relationship between household wealth and solar uptake for houses that pay the median adjusted bill (and this relationship holds for the 20<sup>th</sup> and 80<sup>th</sup> percentile adjusted bills also).<sup>8</sup> Despite differences in the probability of solar uptake across states (solar uptake is highest in QLD and lowest in VIC), the impact of wealth on solar uptake (as measured by the gradient of the line) is similar across all states.

<sup>8</sup> These results are available from the authors upon request.

Further, wealth and solar uptake are almost perfectly negatively correlated for all states (simple correlation coefficient between wealth decile and probability house has solar is -1.000, -0.997, -1.000 and -0.998 for QLD, Vic, SA and NSW, respectively).

Figure 7. Summary marginal effects at median adjusted bill, by wealth decile



#### 4.4 Out-of-sample predictive accuracy

Table 4. compares out-of-sample classification performance for the decision tree, random forest and gradient boosted tree with the performance of a logit and probit model using the CHOICE data set for both the original training data and the oversampled training data.

Table 4. Classification performance for decision tree, random forest and boosted tree using CHOICE data

	<b>Model</b>	<b>Accuracy</b>	<b>Balanced Accuracy</b>
Probit & Logit models	probit model	80%	54%
	probit model, oversampling	75%	63%
	logit model	80%	51%
	logit model, oversampling	63%	62%
Decision tree models	decision tree	73%	56%
	decision tree, oversampling	69%	55%
	random forest	80%	55%
	random forest, oversampling	76%	59%
	boosted tree	80%	54%
	boosted tree, oversampling	75%	59%

Table 5. compares out-of-sample classification performance for the decision tree, random forest and gradient boosted tree with the performance of a logit and probit model using the ABS SIH data set, filtered to include only owner occupied homes, for both the original training data and the oversampled training data.

Table 5. Classification performance for decision tree, random forest and boosted tree using ABS SIH data (owner occupied homes only).

	<b>Model</b>	<b>Accuracy</b>	<b>Balanced Accuracy</b>
Probit & Logit models	probit model	77%	50%
	probit model, oversampling	47%	54%
	logit model	77%	50%
	logit model, oversampling	23%	50%
Decision tree models	decision tree	66%	53%
	decision tree, oversampling	58%	51%
	random forest	73%	52%
	random forest, oversampling	63%	54%
	boosted tree	76%	50%
	boosted tree, oversampling	61%	55%

When used to train classification models, both the CHOICE and the ABS SIH data set used by Best et al. (Best et al., 2019b) lead to balanced accuracy scores that do not indicate a robust classifier. We reiterate that these results do not contradict our assessment of the goodness-of-fit of specific

generalized linear models; a model can fit the available data well but be a poor classifier of unseen data.

For the ABS SIH data set, the situation in the case of the logit model is stark: both standard and oversampled models perform the same as a naïve classification model would (one classifying all unseen data as not having solar panels). Clearly, therefore, we should be reluctant to draw any firm conclusions on variable association from such a model.

We note that, in the case of the CHOICE data set, the (oversampled) probit model classifies unseen data best. The balanced accuracy score (63%) while falling short of the robust score (above 70%) is higher than the 50% achieved from the ABS SIH data. Nevertheless we conclude that drawing categorical conclusions from the generalized linear models fitted to the CHOICE data may not be appropriate.

## **5. Discussion**

What is the relationship between household wealth and rooftop solar in Australia? Our assessment is that the existing literature has not adequately accounted for the relationship between property ownership, building form, wealth and solar uptake. When breaking the data into “owned”, “rented”, “other dwellings” and “owner occupied houses” cohorts, we observe that solar uptake is not associated with household wealth. When the modelling is applied to the segmented data, we do not find wealth to be a significant driver of solar uptake for “other dwellings”. For the remaining cohorts, there is a statistically significant positive relationship between wealth and solar uptake. But the model provides a poor fit of the data, fails classification tests and is likely mis-specified. On this basis, we suggest that the model’s results can not be relied upon.

Our own analysis relies on a much more parsimonious model based on data obtained from a sample of customers’ bills from which we know or can estimate location (by state), end-use consumption volumes, electricity prices and wealth (proxied by IRSAD decile). This model finds a statistically significant inverse relationship between wealth and solar uptake. The model provides a comparable fit of the data and satisfies in-sample prediction tests. However, our model fails the out-of-sample classification tests, albeit only marginally. Our approach can be criticized also for not actually measuring wealth (wealth is proxied by IRSAD score of the postcode that the house is located in) and by using group (post-code) level data. While we concede our “wealth” measure is proximate, such group level criticism is only likely to undermine the plausibility of the relationship we find, if individual

household wealth is systematically different to the “wealth” (as measured by IRSAD) of the post codes in which the house is located. We have no reason to suppose that such systematic difference exists.

Nevertheless the relatively poor predictive accuracy of our model suggests that a richer dataset is likely to be needed in order to confidently conclude that wealth and solar installation is negatively associated. This would include information on real-world factors that are likely to impact solar installation decisions. Such factors might include cluster/peer effects (see for example (Palm, 2017)) solar installation customer acquisition strategies that target specific areas or customer cohorts (see for example (O’Shaughnessy et al., 2021) of this in the U.S.); or local regulations that may promote or undermine solar installation (such as heritage protections or building codes). More generally, we suggest that research that is able to include some of the individual, social and information predictors – see (Alipour et al., 2020) with the consumption, production, irradiance and wealth/income is likely to be able to provide more confident estimation of the probability that households install rooftop solar. Finally, while public subsidy for rooftop solar is generally now much smaller than in the early years, untangling the effect of past subsidies will be valuable in concluding with greater certainty on the relationship between wealth and solar uptake.

## **6. Conclusion and policy implications**

The relationship between wealth and solar uptake by Australian households is not well understood. Our critique of the existing studies suggests claims that solar uptake is positively related to household wealth is not robust when data are appropriately segmented by ownership and/or building form. Our analysis of bills finds that wealth is negatively related to solar uptake in detached owner-occupied houses. However we do not consider the modelling results presented here, or in previous studies, to be adequately robust at predicting solar uptake as a function of household wealth.

Confident estimation of the relationship between solar uptake and wealth is valuable in assessing the case for means-tested policy support. This is likely to become even more important if behind-the-meter batteries paired with solar become popular. This is because greater self-consumption of rooftop solar will have a significant impact on the recovery of shared grid costs. More generally, broadening the research agenda to understand better the range of demand-side and supply-side factors affecting distributed production and storage uptake will be valuable.

## Appendix

Below we provide additional details on the three tree-based models used in the paper. In each case we use features (either the CHOICE data set or the ABS data set) to predict the value of a response variable (solar uptake) in a process known as supervised learning. More details on all three methods can be found in (Hastie et al., 2021), or (for the mathematically inclined) (Hastie et al., n.d.).

The set of possible values for our response variable is  $\{0,1\}$  where 1 represents a household with solar and 0 a household without solar. Our features are one of three types: real valued (such as wealth), categorical (such as number of persons in household) and binary (such as whether the household is in Victoria or not). The set of all possible values for the features is known as the feature space ( $F$ ) and the aim of a supervised learning method is to use the training data to construct a function  $\hat{f}: F \rightarrow \{0,1\}$  such that the value  $\hat{f}$  takes on an element of the test set matches the real value as often as possible.

### Decision Tree

A decision tree is produced by partitioning the feature space into hyper-rectangles (a higher dimensional analogue to the two dimensional rectangle) and then constructing a function  $\hat{f}$  that is constant on each hyper-rectangle. The resultant function can be visualized as a binary tree with each leaf representing one of the hyper-rectangles; this aids in interpreting the model and is one reason for its popularity.

The algorithm we have used to produce the decision tree (implemented in the scikit-learn Python package) constructs the tree iteratively. At each stage in the construction of the tree, it examines all possible ways to further partition the feature space and chooses the one with the lowest gini impurity. Decision trees trained in this way can often be overly complex and over-fit the training data. For this reason we have also used two other tree-based methods to train a classifier.

### Random Forest

A random forest consists of many decision trees and is an example of an ensemble learning method. For an element of the test set, a decision tree  $\hat{f}_i$  outputs a predicted response, and the random forest  $\hat{f}$  is then defined as the response selected most often by the collection of trees.

To train the decision trees, we use a technique known as bootstrap aggregating. For example, to produce a random forest model consisting of  $B$  trees using a training set of size  $n$ , we produce  $B$  synthetic training sets of size  $n$  by randomly sampling the original training set with replacement, and



then use these synthetic training sets to produce the collection of trees. This typically leads to better model performance on unseen data by decreasing the sensitivity of the model to the training set and thus decreasing the probability of over-fitting the training set.

### **Gradient Tree Boosting**

Gradient tree boosting is another ensemble learning method built from decision trees. It builds up a model  $\hat{f}$  iteratively. At the first step of the iteration, we fit a decision tree  $\hat{f}_1$  to the training set. Given the features  $x_j$  and response  $y_j$  of an element of the training set we can calculate the difference

$$h_1(x_j) = \hat{f}_1(x_j) - y_j$$

The values  $h_1(x_j)$  can then be considered new response variables for the training set, and so we can fit a decision tree  $\hat{f}_2$  to this adjusted training set. We terminate this process after the  $k$ th stage to produce our final model  $\hat{f}_k$ .

Gradient tree boosting algorithms typically outperform both individual decision trees and random forests in classification problems, but come at the expense of lower interpretability of the model.

## References

- Alipour, M., Salim, H., Stewart, R. A., & Sahin, O. (2020). Predictors, taxonomy of predictors, and correlations of predictors with the decision behaviour of residential solar photovoltaics adoption: A review. In *Renewable and Sustainable Energy Reviews* (Vol. 123). Elsevier Ltd. <https://doi.org/10.1016/j.rser.2020.109749>
- Andrews, D. (1988). Chi-Square Diagnostic Tests for Econometric Models : Theory. *Econo*, 56(6), 1419–1453.
- Australian Bureau of Statistics. (2016). *Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), Australia, 2016 cat. 2033.0.55.001*.
- Australian Bureau of Statistics. (2017). *6540.0 - Microdata: Household Expenditure, Income and Housing, 2015-16*.
- Australian Bureau of Statistics. (2018). *Census of Population and Housing 2016, Table Builder*.
- Australian Bureau of Statistics. (2019). *6541.0.30.001 - Microdata: Income and Housing, Australia, 2017-18*.
- Australian PV Institute. (2019). *PV Postcode Data. Solar Map*.
- Australian Taxation Office. (2018). *Table 28 by postcode, for the 2015–16 income year*.
- Barbose, G., Forrester, S., Shaughnessy, E. O., & Darghouth, N. (2021). *Residential Solar-Adopter Income and Demographic Trends : 2021 Update. 34158*. [https://eta-publications.lbl.gov/sites/default/files/solar-adopter\\_income\\_trends\\_final.pdf](https://eta-publications.lbl.gov/sites/default/files/solar-adopter_income_trends_final.pdf)
- Best, R., Burke, P. J., & Nishitatenno, S. (2019a). Evaluating the effectiveness of Australia’s Small-scale Renewable Energy Scheme for rooftop solar. *Energy Economics*, 84, 104475. <https://doi.org/10.1016/j.eneco.2019.104475>
- Best, R., Burke, P. J., & Nishitatenno, S. (2019b). Understanding the determinants of rooftop solar installation: evidence from household surveys in Australia. *Australian Journal of Agricultural and Resource Economics*, 63(4), 922–939. <https://doi.org/10.1111/1467-8489.12319>
- Best, R., Chareunsky, A., & Li, H. (2021). Equity and effectiveness of Australian small-scale solar schemes. *Ecological Economics*, 180. <https://doi.org/10.1016/j.ecolecon.2020.106890>
- Bureau of Meteorology. (2018). *Solar Exposure for Australia*.
- Carley, S., & Konisky, D. M. (2020). The justice and equity implications of the clean energy transition. *Nature Energy*, 5(8). <https://doi.org/10.1038/s41560-020-0641-6>
- classification - Training a decision tree against unbalanced data - Cross Validated*. (n.d.). Retrieved October 29, 2021, from <https://stats.stackexchange.com/questions/28029/training-a-decision-tree-against-unbalanced-data>
- Clean Energy Regulator. (2021). *Postcode data for small-scale installations*.

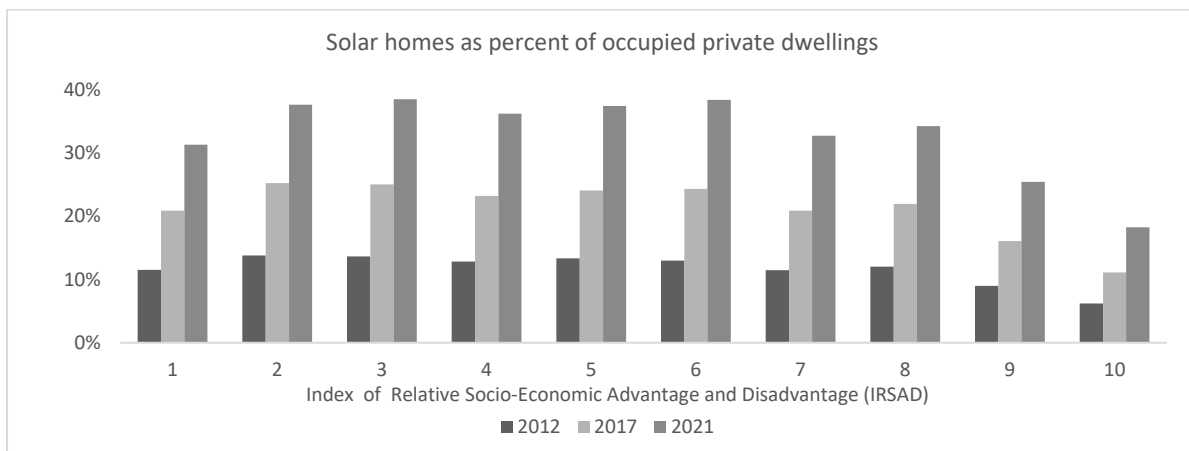
- Fadden, D. M. (1974). Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics* (Vol. 33, Issue 8).
- Hastie, T., Tibshirani, R., & Friedman, J. (n.d.). *Springer Series in Statistics The Elements of Statistical Learning Data Mining, Inference, and Prediction*.
- Hastie, T., Tibshirani, R., James, G., & Witten, D. (2021). An Introduction to Statistical Learning (2nd Edition). *Springer Texts, 102*, 618.
- Ho, T. K. (1995). Random decision forests. *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR, 1*, 278–282. <https://doi.org/10.1109/ICDAR.1995.598994>
- Hosmer, D. W., & Lemeshow, S. (1989). *Applied Logistic Regression Third Edition*.
- IHS Markit. (2017). *EvIEWS 10 User's Guide II* (2nd ed.). IHS Global Inc.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems, 2017-Decem(Nips)*, 3147–3155.
- Liu, Y., Zhou, Y., Wen, S., & Tang, C. (2014). A Strategy on Selecting Performance Metrics for Classifier Evaluation. *International Journal of Mobile Computing and Multimedia Communications, 6(4)*, 20–35. <https://doi.org/10.4018/IJMCMC.2014100102>
- Lukanov, B. R., & Krieger, E. M. (2019). Distributed solar and environmental justice: Exploring the demographic and socio-economic trends of residential PV adoption in California. *Energy Policy, 134*. <https://doi.org/10.1016/j.enpol.2019.110935>
- McFadden, D. (1979). Quantitative Methods for Analyzing Travel Behaviour on Individuals: Some Recent Developments. In D. Hensher & P. Stopher (Eds.), *Behavioural Travel Modelling* (1st ed.).
- Mountain, B., Gasseem, A., Burns, K., & Percy, S. (2020). *A model for the estimation of residential rooftop PV capacity* (Issue April). <https://doi.org/10.26196/5ebca99c43e1a>
- Mountain, B. R., & Szuster, P. (2015). Solar, Solar Everywhere: Opportunities and Challenges for Australia's Rooftop PV Systems. *IEEE Power and Energy Magazine, 13(4)*. <https://doi.org/10.1109/MPE.2015.2416113>
- O'Shaughnessy, E., Barbose, G., Wiser, R., & Forrester, S. (2021). Income-targeted marketing as a supply-side barrier to low-income solar adoption. *IScience, 24(10)*. <https://doi.org/10.1016/j.isci.2021.103137>
- O'Shaughnessy, E., Barbose, G., Wiser, R., Forrester, S., & Darghouth, N. (2021). The impact of policies and business models on income equity in rooftop solar adoption. *Nature Energy, 6(1)*. <https://doi.org/10.1038/s41560-020-00724-2>
- Palm, A. (2017). Peer effects in residential solar photovoltaics adoption—A mixed methods study of

- Swedish users. *Energy Research and Social Science*, 26, 1–10.  
<https://doi.org/10.1016/j.erss.2017.01.008>
- Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81–106.  
<https://doi.org/10.1007/bf00116251>
- Rahman, R. M., & Fazle, F. R. (2011). Using and comparing different decision tree classification techniques for mining ICDDR,B Hospital Surveillance data. *Expert Systems with Applications*, 38(9), 11421–11436. <https://doi.org/10.1016/j.eswa.2011.03.015>
- Schapire, R. E. (2003). The Boosting Approach to Machine Learning: An Overview BT - Nonlinear Estimation and Classification. *Nonlinear Estimation and Classification*, 171(Chapter 9), 149–171.
- Sunter, D. A., Castellanos, S., & Kammen, D. M. (2019). Disparities in rooftop photovoltaics deployment in the United States by race and ethnicity. *Nature Sustainability*, 2(1).  
<https://doi.org/10.1038/s41893-018-0204-z>
- Welton, S., & Eisen, J. (2019). Clean energy justice: Charting an emerging agenda. *Harvard Environmental Law Review*, 307–371.
- Wood, G., Ong, R., & McMurray, C. (2012). Housing Tenure, Energy Consumption and the Split-Incentive Issue in Australia. *International Journal of Housing Policy*, 12(4), 439–469.  
<https://doi.org/10.1080/14616718.2012.730218>
- Zander, K. K. (2020). Unrealised opportunities for residential solar panels in Australia. *Energy Policy*, 142(April), 111508. <https://doi.org/10.1016/j.enpol.2020.111508>

The ABS Census of Population and Housing (“the census”) is conducted every four to five years and provides another source of information to help explore the relationship between rooftop solar and household socio-economic characteristics including income, ownership and location (i.e. postcode) (Australian Bureau of Statistics, 2018). Based on the information from the census, the ABS also ranks each post-code to an Index of Relative Socio-economic Advantage and Disadvantage (IRSAD) that varies between 1 (lowest) to 10 (highest) (Australian Bureau of Statistics, 2016). Data on rooftop solar installations by postcode from 2001 to 2021 is available from (Clean Energy Regulator, 2021). Mapping the data on solar installations to occupied private dwellings by postcode using the ABS census data provides another avenue to explore the relation between solar and wealth.

Figure 4 shows the rooftop solar penetration in 2012, 2017 and 2021 ranked by IRSAD decile. The chart shows similar rooftop solar penetration except in the lowest socio-economic decile and the highest two deciles where the penetration declines relative to other deciles, increasingly so comparing 2012 to 2021.

Figure 8. Solar homes as percentage of occupied private dwellings by IRSAD decile



Source: (Australian Bureau of Statistics, 2016, 2018; Clean Energy Regulator, 2021)