Modeling photovoltaic diffusion: an analysis of geospatial datasets

Carolyn Davidson, Eesan Drury, Anthony Lopez, Ryan Elmore and Robert Margolis

National Renewable Energy Laboratory, 15013 Denver West Parkway Golden, CO 80401, USA

E-mail: carolyn.davidson@nrel.gov

Received 22 October 2013, revised 3 June 2014
Accepted for publication 5 June 2014
Published 14 July 2014

Abstract
This study combines address-level residential photovoltaic (PV) adoption trends in California with several types of geospatial information—population demographics, housing characteristics, foreclosure rates, solar irradiance, vehicle ownership preferences, and others—to identify which subsets of geospatial information are the best predictors of historical PV adoption. Number of rooms, heating source and house age were key variables that had not been previously explored in the literature, but are consistent with the expected profile of a PV adopter. The strong relationship provided by foreclosure indicators and mortgage status have less of an intuitive connection to PV adoption, but may be highly correlated with characteristics inherent in PV adopters. Next, we explore how these predictive factors and model performance varies between different Investor Owned Utility (IOU) regions in California, and at different spatial scales. Results suggest that models trained with small subsets of geospatial information (five to eight variables) may provide similar explanatory power as models using hundreds of geospatial variables. Further, the predictive performance of models generally decreases at higher resolution, i.e., below ZIP code level since several geospatial variables with coarse native resolution become less useful for representing high resolution variations in PV adoption trends. However, for California we find that model performance improves if parameters are trained at the regional IOU level rather than the state-wide level. We also find that models trained within one IOU region are generally representative for other IOU regions in CA, suggesting that a model trained with data from one state may be applicable in another state.

Keywords: solar photovoltaic, technology diffusion, distributed generation

1. Introduction

Adoption of photovoltaic (PV) systems by US households has witnessed a dramatic increase over the past decade with substantial continued growth anticipated. While industry forecasts provide a sense of aggregate market growth, technology diffusion models have the potential to provide a spatial description of market growth. These models can be used by various market participants, ranging from utilities and regulators planning for increased distributed generation in their services territories, to companies targeting market segments with higher propensities for adoption.

As a result, there is growing interest in developing PV diffusion models to characterize PV market demand to a range of factors; including future PV price trends, solar policies, access to financing, and others (Cai et al 2013, Darghouth et al 2014, Paidipati et al 2008, Denholm et al 2009, Drury et al 2012). Diffusion characteristics are frequently formulated using aggregate diffusion models such as Bass diffusion (Denholm et al 2009, Guidolin and Mortarino 2010, Zhang et al 2011), Fisher–Pry diffusion (Paidipati et al 2008), logistic regression models (Lobel and Perakis forthcoming), system dynamics frameworks (R W Beck 2009, EIA 2012), and agent-based models (Robinson 2013). Many PV diffusion models assume that diffusion patterns are...
largely a function of the estimated value of a PV system, but have relied on few empirical constraints to inform diffusion parameterizations and market segmentation. While more recent business models have enabled consumers immediate money savings upon installing a PV system, PV is often valued for its ‘green’ attributes. Segmenting and profiling green consumers often relies on socio-demographic variables, particular as a first cut, due to the ease of obtaining such data through publically-available source (Diamantopoulous et al 2003). Hines et al (1986) meta-analysis characterized individuals that engage in pro-environmental behavior as more likely to be young, affluent and well-educated.

Several recent studies relying on surveying or evaluating historic PV adoption trends have suggested that demographic variables may drive unique adoption trends in different market segments (Faiers et al 2007, Rai and Robinson 2013, Rai and Sigrin 2013, Drury et al 2012, Bollinger and Gillingham 2012, Kwan 2012). For example, Drury et al (2012) found that leasing PV systems may appeal to younger, less affluent demographics. Rai and Sigrin (2013), relying on survey data combined with electricity consumption data suggested that a households’ available cash flow critically drove market segmentation. Kwan (2012) relied on a geospatial dataset to predict PV adoption nationally and found that, in addition to several demographic variables, electricity costs, solar insolation and financial incentives were key drivers. Results from these studies suggest that geospatial population characteristics may be useful for predicting future PV market demand by defining market segments and constraining diffusion models accordingly.

In this analysis, we evaluate a more expansive set of variables than have been evaluated in previous studies. Further, we frame results in the context of implications for diffusion modeling; discussing which geospatial data are the most predictive of historical PV adoption trends as well as explore the implications of varying data resolution and regional coverage. We associate address-level PV adoption data (118,471 homes) from the California Solar Initiative (CSI) with several sources of geospatial data (demographics, housing characteristics, vehicle ownership, and others). Since this analysis is limited to California, we explore how these predictive factors vary regionally and at different spatial scales. By identifying which types of geospatial data are the most predictive of historical PV adoption trends, we can isolate subsets of population parameters that could be used to inform the structure of diffusion models and constrain parameterizations of diffusion dynamics.

2. Data

This study uses several data sources, including PV adoption data from the CSI incentive program, population characteristics from the US Census, vehicle ownership information from R L Polk, and other data sources. Table 1 lists each of the data sources used, along with a brief description of the data and its native spatial resolution. We describe each data source in further detail in the following subsections. The majority of our analysis relies on aggregating all data sources to the ZIP code level, though we compare regressions at the block group level in section 5.

We summarise all data at the target resolution by averaging the data where the geographies of the two datasets intersect. When areas contain more than one region, we rely on an area-weighted average.

2.1. PV adoption data

We use PV adoption data from the CSI, a solar incentive program that serves California’s three IOUs: Pacific Gas and Electric (PG&E), Southern California Edison (SCE) and San Diego Gas and Electric (SDG&E). The CSI, administered by the California Public Utilities Commission (CPUC), is the largest state solar incentive program in both installed PV capacity and funding. For this analysis, we use CSI data ranging from January 2007 through 29 June 2013. This data includes 139,886 residential systems, 4,750 commercial systems, 2,747 government systems and 991 non-profit systems.

1 Note that each data source represents a snapshot for a particular time, which varies depending on the dataset. As a result, the data is used as a static representation of each area over the entire time period 2007–13. Certain geographic measures were likely impacted by the US recession, such as foreclosure rates and various income measures, and may have changed significantly over this time period. Other measures, such as race, family status and housing characteristics were likely more static. The potential impact this could have on results is discussed in section 3.1.

2 This method introduces some error in that misaligned topology (due to data source differences) between data and target regions can cause coterminous edges to be overlapped. Further, simple overlays of large regions can over- or under-estimate data where populations are concentrated.
From these data, we excluded all commercial, government and non-profit systems, as well as residential systems that had been canceled, withdrawn, removed, suspended, or transferred. This left 118,471 residential PV systems in the data used for analysis.

In addition to the publicly available CSI data, we also received system addresses from the CPUC. System addresses were geolocated using Google’s geocoding service. This data enabled us to associate each residential PV installation to other data sources at various resolutions, from block-level information (US Census) to ZIP code level information (car ownership, foreclosure rates, etc).

2.2. Demographic data (American Community Survey (ACS) and Demographic Profile (DP1))

We use two types of demographic data in this analysis: (1) Census data from 2010 and (2) ACS data from 2007–11. In our study, we use the Census 2010 Summary File 1 DP1, which contain summary statistics of demographic questions asked of every household which includes information on occupant race, age, education, household size and composition (US Census 2011). The US Census Bureau provides pre-joined geographies for the DP1 data (http://census.gov/geo/maps-data/data/tiger-data.html), and we use Census data at Tract-level resolution in this study.

The ACS is a statistical survey conducted by the US Census Bureau that samples a small percentage of the US population every year in an effort to explain how people live, and is designed to provide communities with demographic, housing, social, and economic data (US Census 2008). ACS provides 1, 3 and 5 year rolling data, depending on regional population. The 5 year ACS data are based on significantly larger survey samples than the 1 or 3 year ACS data, which makes them more reliable and includes information on smaller populations.

2.3. Vehicle ownership data (Polk)

R L Polk (Polk) is an automotive consulting company that collects and manages vehicle ownership information, including vehicle registrations, sales, and titles for personal and commercial cars, light- and heavy-duty trucks, motorcycles, and RVs. In this analysis, we use data on the number of registered hybrid electric vehicles, diesel vehicles and electric vehicles in each CA ZIP code.

2.4. Foreclosure risk data

The Local Initiative Support Corporation (LISC), a community development support organization, developed a ‘foreclosure risk score’ indicator that combines the following indicators via a weighting/adjusting scheme: percentage of residential units with (a) first-lien mortgage, (b) subprime first-lien mortgages, (c) first-lien mortgages delinquent 30 or more days, and, (d) vacancies. This study relied on scores updated in March 2013. The highest risk ZIP code in a state is assigned a score of 100, and other ZIPs in a given state are assigned a score relative to the highest score (LISC).

2.5. Canopy density data

The Multi-Resolution Land Characteristics Consortium (MRLC) is a group of coordinated federal agencies that generate land-cover information (MRLC website). The canopy density dataset was created by MRLC based on empirical relationships between tree canopy density and Landsat satellite imagery through linear regression techniques (Huang et al. 2001).

2.6. Solar irradiance

The National Renewable Energy Laboratory (NREL) provides solar irradiance data through its National Solar Radiation Database (NSRDB, 2007). NREL’s gridded dataset was produced using geostationary satellite images to estimate global and direct irradiance at hourly intervals at a 10 × 10 km horizontal spatial resolution.

3. Methodology

3.1. Dependent variable

In this analysis, we define the installed base as the total number of residential PV systems in a given target region (ZIP code or Census block group) over the period 2007–13.

The number of residential systems installed in the 1218 California ZIP codes over the 2007–13 study period ranged from 1 to 819. We log-transformed the number of cumulative installations and use this as the model dependent variable in order to produce more normally distributed model residuals.

3.2. Explanatory variables

Table 2 lists an illustrative set of population variables used in this study that are derived from the datasets described in section 2. Additional geospatial data was used in this study (89 additional variables), and the list in table 2 represents the subset of variables that were retained by one or more model. The appendix provides further description of each independent variable used in this analysis. Here and elsewhere, variables are color-coded according to their parent dataset.

We retained most information contained in these datasets, but often collapsed or summarized the data into broader
categories. For example, rather than including a variable for age in increments of 5 years, we collapse this into 10 year age groups. In a few cases, we excluded a category of data—for example, number of workers by occupation. We aggregate all explanatory variables at the ZIP code level.

In order to enable intuitive comparison of regression coefficients across variables with large variations in units, we standardized the data by subtracting the mean from each observation and dividing by the standard error (Gujarati 2011). The resulting independent variables are unit-less, with mean zero and a standard deviation of one. Coefficients can then be interpreted as the resultant change in the standard deviation of the dependent variable resulting from a one standard deviation change in the independent variable.

### 3.3. Model selection

This analysis aimed to evaluate key variables and possible associations that could be used to inform PV adoption and diffusion parameters. We employed an ordinary least squares (OLS) specification due to the desirable properties of the OLS estimator as well as the computational facility of implementing and comparing several models. We assume the following form, with a logarithmic transformation of the dependent variable:

\[
Y = \alpha + \beta X + \epsilon
\]

where \(Y\) is the logarithm of the number of cumulative residential PV installations in ZIP code \(i\), \(X\) represents the vector of explanatory variables for each ZIP code \(i\) and \(\epsilon\) represents the random error term. We evaluate several models relying on data at different spatial scales and over specific geographic regions.

Partial correlation coefficients between explanatory variables illustrated high pairwise correlation between explanatory variables—which suggested that multiple variables communicate the effect of some common attribute (for example, disposable income). This flags concern for multicollinearity in a multiple regression model.

To identify a parsimonious model, we rely on a stepwise regression procedure (Kutner et al. 2004). Stepwise regression selects a subset of variables from a larger set by relying on an algorithm that tests the addition of each variable; after a new variable is added, the algorithm tests if variables can be deleted without significantly impacting the Akaike Information Criteria (AIC), finally selecting the set of variables that minimizes AIC. It is important to note that while this procedure reduces multicollinearity by dropping redundant variables used in this analysis are included in the appendix.

### Table 2. Illustrative sample of explanatory variables.

| Variable name | Variable description | Number of unseen variables | Pool |
|---------------|----------------------|---------------------------|------|
| total         | Total population     | 20                        |      |
| under20       | Total population under 20 | 20             |      |
| age20-30      | % total population age 20-30 | 20           |      |
| age30-40      | % total population age 30-40 | 20           |      |
| age40-50      | % total population age 40-50 | 20           |      |
| age50-60      | % total population age 50-60 | 20           |      |
| age60-70      | % total population age 60-70 | 20           |      |
| age70+        | % total population age 70+ | 20           |      |
| white         | % population white    | 20                        |      |
| black         | % population black    | 20                        |      |
| child         | % households with > 1 children | 20   |      |
| husbandwife   | Percent of husband-wife households | 20 |      |
| female        | Family size           | 20                        |      |
| ownerocc      | Housing units occupied by owner | 20 |      |
| lotownerocc   | # owner – occupied housing units | 20 |      |
| firefighters  | Households with one worker | 20               |      |
| fireworkers   | Households with two workers | 20            |      |
| geographic    | Households with three workers | 20 |      |
| samehouse     | Same house 1 year ago  | 20                        |      |
| car           | Car transportation to work | 20                 |      |
| school        | Work from home        | 20                        |      |
| traveltime    | Travel time to work   | 20                        |      |
| age90        | Aggregate earnings in past 12 months | 20 |      |
| age40        | Household income quintile upper limit | 20 |      |
| rooms         | Median number of rooms | 20                        |      |
| utilities     | House heating fuel supplied by utility gas | 20 |      |
| total         | Mean hours worked     | 20                        |      |

**Legend:**
- DPI
- ACS
- Canopy
- Pool
- Foreclosure
- Insulation
- CSI

---

6 In this case, there were 20 additional occupations. These were excluded for parsimony, but could be included in future analysis.

7 This analysis assumed spatially independent errors. Future analysis will focus on evaluating the potential impact of spatial autocorrelations on coefficient estimates.

8 In the case of multicollinearity, overall model prediction remains reliable, but the coefficients on individual predictors with respect to their impact on the dependent variable can be imprecise, and fluctuate significantly based on model specification and data.

9 The AIC is a commonly used measure of goodness-of-fit that rewards better fits but penalizes losses in degrees of freedom (Greene 2011). This procedure was implemented using the MASS package in R (http://cran.r-project.org/web/packages/MASS/index.html).
variables, it does not ensure that remaining variables are the most significant, nor that the model does not exclude a key variable. Selected variables may simply be a proxy for an adoption driver. As a result, it is important to interpret results as relevant to predicting adoption, rather than driving adoption.

In order to assess the sampling variability of our explanatory variables under study, we randomly sampled (with replacement) 100 training data sets containing 70% of the original data. We chose the model with the smallest mean squared error (MSE) over the test set as a metric for predictive performance. We then ran this model on the full set of data as the best-fit model.

4. Results

4.1. Full model

Table 3 presents results; columns 1 and 2 present the mean and standard deviation on the trained model on the full data set and columns 3 and 4 represent the mean and standard deviation across 100 training runs. To simplify presentation, we illustrate only the variables selected in 30% or more of the model runs. All identified variables were significant at the 5% level or less—with the exception of diesel. The mean adjusted $R^2$ for all samples was 0.55 and the mean MSE was 0.49.

Eight variables were selected for inclusion in every sample model: masters (+), rooms (+), foreclosure (+), hev (+), mortgage2orHE (+), insolation (+), car (−), percentTPO (−). Two variables, value500kto1mil (+) and value200kto300k (+), were selected for inclusion in all but one model. Based on their standardized coefficients, these variables were also found to be some of the largest predictors of cumulative adoption/non-adoption, in addition to child (+) and age60to70 (+) and bachelor (−). For example, in Model 1 an increase in one standard deviation in value500kto1mil (in this case a 10% increase in owner occupied houses valued between $500,000 and $1,000,000) results in a 0.31 standard deviation increase in the log of adoption (an additional 22 cumulative PV systems per zip code). Coefficient stability varied, in some cases substantially, depending on the variable and sample. This was likely driven by variation in the specific set of best-performing variables selected by the step-wise algorithm, and based on differences in the randomly drawn samples.
4.2. Variable importance

While the results in section 4.1 evaluated the contribution of a large set of variables, in this section, we explore the marginal increase in model performance gained by adding incremental variables. This helps to inform the number of population variables that could be used to parameterize PV diffusion. We undertook a variable collection procedure, leaps to identify the most efficient subset of independent variables, for subsets ranging from one to eight variables\(^{10}\). This procedure solves for the most predictive subset of variables using a branch-and-bound algorithm, relying on AIC as the selection criterion. All regressions relied on the full PV adoption dataset at ZIP code resolution.

Figure 1 presents the results of the leaps procedure, where the selected variables (rows) are shown for each subset size (columns), along with regression coefficients. Also shown is model performance (adjusted $R^2$-squared) for each subset of variables, as well as the regression coefficients from the best-fit model identified in table 2.

The variable mortgage2orHE was identified as the single strongest indicator of adoption when the model was limited to one variable, but was dropped in larger subsets; suggesting that this variable provided a blunt positive correlate for adoption. Rooms and hev were consistently included in models of all subset sizes, suggesting these variables are unique positive correlates to PV adoption. Other variables that were consistently selected in subsets included masters (+), wood (−), value.over1mil (−) and percentTPO(−). In addition, these variables displayed largely robust coefficients across subset sizes. The degree to which these coefficients were consistent with the coefficients produced by the full step-wise model varied. We can infer that multicollinearity was more problematic for unstable coefficients—essentially, some variables communicate an unobserved factor that was highly correlated with several variables.

This procedure provides further insight into developing a parsimonious model. For example, while the best-fit model identified in section 4.1 included 40–50 additional variables, these variables only marginally increase the predictive power of the model (from an adjusted $R^2$ of 0.49 to 0.55). Including a core set of six to eight variables could potentially provide a model with similar explanatory power.

The variables identified in the leaps procedure (figure 1), as well as the variables with the largest standardized impact on adoption (table 2) are consistent with the literature on PV adoption and green consumption. Particularly, the positive relationship between higher education and PV adoption suggested by masters (% of population with a master’s degree) is consistent with Drury et al (2012). Signs and significance of value500kto1mil and value200kto300k are consistent with higher adoption in middle- upper middle class neighborhoods—consistent with Drury et al (2012) and Kwan (2012). Higher adoption in areas with a higher white population and higher insolation is consistent with Kwan (2012). The mean signs on the different age variables vary substantially (positive coefficient for age20to30, age40to50, age60to70, and age over80, and negative coefficients for age50to60, and age70to80) are inconsistent with Drury et al (2012) though somewhat consistent with Kwan (2012). However, none of the age variables are selected in subsets with fewer than eight variables in the leaps procedure suggests that age may not be a particularly strong predictor given other available variables.

Several key variables identified in table 3 and/or figure 1 had not been explored in previous PV adoption literature. The positive and significant sign of rooms (average number of rooms in house) may reflect that larger houses consume more electricity from higher tiers in California, increasing the cost savings from PV. Significance of mortgage2orHE (% of

\(^{10}\) This procedure was implemented using the LEAPS package in R (http://cran.r-project.org/web/packages/leaps/leaps.pdf).
population with a 2nd mortgage and/or home equity loan) may reflect a segment that is willing to leverage their resources to invest in property assets (including PV). This aligns with the Rai and Sigrin (2013) finding that free cash flow is a strong determinant of PV adoption decisions. The strong relationship found between hybrid electric vehicle adoption and PV adoption suggests overlapping demographics for the two green products. Finally, the rationale behind the strong performance of the foreclosure variable (a calculated foreclosure-likelihood score) in the step-wise regression is unclear—but likely serves as an example of a constructed variable that performs well in describing a particular PV segment. Note that foreclosure is highly correlated (over 0.50) with the following variables: number of household members under 20, household size, female-headed households and hybrid electric vehicle ownership.

Overall, both the variables included in the full model and the model limited to a subset of eight variables provide substantial explanatory power—explaining 55% and 48% of the variation in the dependent variable, respectively. Further, both models have an $F$-statistic that indicates overall model significance (test statistic of 36 and 28, respectively). In the full, min MSE model, all but one variable, diesel, was significant at the 5% level or less. In the subset model, all variables were significant at the 5% level or less.

5. Comparison of spatial resolution

Spatial data—like US PV adoption data made available by several incentive administrators (such as the Open PV project)—is frequently aggregated to the ZIP code level. This may be perceived as a modeling limitation when more detailed spatial granularity is desired. To evaluate whether the inferences gained from ZIP code level regressions are similar to those gained from analyses relying on higher resolution data, we tested an additional specification summarizing all data at the Census block group level. Data not available at the block group level was assigned the smallest level of granularity available, based on the methods outlined in section 2. We replicated the model selection procedures from section 3.3 for correlations at block group spatial resolution.

Table 4 provides the coefficients and standard errors of the model with best predictive accuracy (as defined by lowest MSE), the mean coefficient for each variable across all 100 samples, the number of samples that selected each variable. Table 3 also lists the native resolution of each variable and the corresponding coefficient for the best fit model identified from the ZIP code level analysis.

The best-performing model at the block group level had an adjusted $R^2$ and MSE of 0.48. We find that the predictive performance of block group-level models (0.48 adjusted $R^2$, 0.48 MSE), is lower than that of ZIP code level models (0.58 adjusted $R^2$, 0.38 MSE). It is more difficult to accurately predict PV adoption trends at higher spatial resolution without higher resolution data. Several of the independent variables have tract level resolution, which is coarser than the resolution of the dependent variable for this model. As a result, while these variables can provide unique information across tracts, they will not be able to provide unique information within tracts (i.e. block group or blocks), and therefore will be less useful predictors relative to the ZIP code-level model. We also find that the best-performing block group model selected different subsets of key variables than the ZIP code-level model. Key similarities include the consistent inclusion of value500 kto1 mil, rooms, masters and toownterocc. However, the magnitude of several of these coefficients was noticeably smaller (e.g., masters has a coefficient of 0.05 in the block group analysis as opposed to 0.30 in the ZIP code level analysis). In addition there were a few contradictory results. Namely, hev is only selected for inclusion in seven models, foreclosure, while selected in all 100 sample models, has a negative coefficient, and mortgage2orHE, selected in 83 models, also has a negative coefficient. The hev and foreclosure results may be attributed to the inherent limitations of using coarser-resolution data (ZIP code) to inform block group—level adoption trends.

Figure 2 shows the results of the most predictive models trained on subsets of variables ranging from 1 to 8 factors for the block group-level data. The results in figure 2 suggest, similar to the ZIP code level analysis, that a parsimonious subset of 5 to 8 variables may be nearly as predictive of PV adoption as a much larger dataset. However, a somewhat different set of variables were identified. Contrary to the ZIP code level analysis, house age (built2000s, built50s) and house value (value500 kto1 mil, value1over1 mil) as well as avgvehicles and husbandwife appear to be key predictive variables in both the regression and the subset models.

Similar to several of the variables identified in the ZIP code level analysis, many of these variables have no precedence in the PV adoption literature, yet have intuitive appeal. More recently built houses are less likely to require roof replacements. While areas with a higher percentage of married couples (husbandwife) are more likely to adopt PV, the negative coefficient on family size may be indicative of a more restrictive cash flow situation that precludes PV adoption for larger families.

Similar to the ZIP code level analysis, rooms, masters, famsize and white surfaced as key predictors. We also find that the data that is only available at coarser native resolutions (i.e. Polk data and foreclosure data) is not useful for representing block group-level variations in adoption trends since this data is assigned an equal value across several block groups.

6. Regional testing

Constraining diffusion model parameters using historical PV adoption trends is limited by the fact that PV adoption has primarily occurred in locations with relatively high electricity rates and significant PV incentives. California has, by far, the largest residential PV market of all the states in the US, and US-focused PV diffusion models will likely rely on heavily on California data to inform or constrain diffusion parameters. However, diffusion trends in California may not be representative of national market trends.
Table 4. Step-wise regression results at block group resolution.

| Variable                  | Coefficient (Min MSE model) | Std Err. (Min MSE model) | Mean Coefficient (all samples) | Number of samples that identified variable | Native Resolution | Zip Code Coefficient |
|---------------------------|-----------------------------|--------------------------|--------------------------------|--------------------------------------------|-------------------|----------------------|
| value_500to1mlm           | 0.20                        | 0.02                     | 0.20                           | 100                                        | Block group       | 0.29                 |
| avgvehicles               | 0.20                        | 0.010                    | 0.20                           | 100                                        | Block group       | NA                   |
| rooms                     | 0.18                        | 0.010                    | 0.18                           | 100                                        | Block group       | 0.27                 |
| built2000s                | 0.16                        | 0.008                    | 0.16                           | 100                                        | 10 km grid        | 0.09                 |
| isolation                 | 0.12                        | 0.009                    | 0.12                           | 100                                        | Block group       | 0.17                 |
| value_over1mlm            | 0.12                        | 0.018                    | 0.12                           | 100                                        | Block group       | NA                   |
| stable                    | 0.11                        | 0.010                    | 0.11                           | 100                                        | Tract             | NA                   |
| diesel                    | 0.10                        | 0.008                    | 0.10                           | 100                                        | Zip               | 0.02                 |
| lownowrecoc               | 0.10                        | 0.020                    | 0.09                           | 100                                        | Tract             | 0.22                 |
| value_2000to300k           | 0.09                        | 0.012                    | 0.09                           | 100                                        | Block group       | 0.18                 |
| masters                   | 0.05                        | 0.010                    | 0.05                           | 100                                        | Block group       | 0.32                 |
| gnd                       | 0.04                        | 0.008                    | 0.05                           | 100                                        | Block group       | -0.09                |
| built50s                  | 0.04                        | 0.007                    | 0.03                           | 100                                        | Block group       | NA                   |
| mortgage_over2000perincome| -0.03                       | 0.008                    | -0.03                          | 100                                        | Block group       | -0.11                |
| foreclosure               | -0.04                       | 0.010                    | -0.04                          | 100                                        | Zip               | 0.26                 |
| nonhomomortgage2to10HE     | -0.05                       | 0.009                    | -0.05                          | 100                                        | Block group       | NA                   |
| transacted                | -0.05                       | 0.008                    | -0.05                          | 100                                        | Tract             | NA                   |
| age75to89                 | -0.07                       | 0.014                    | -0.06                          | 100                                        | Tract             | NA                   |
| built50s                  | -0.06                       | 0.008                    | -0.07                          | 100                                        | Block group       | NA                   |
| wood                      | -0.07                       | 0.011                    | -0.07                          | 100                                        | Tract             | NA                   |
| famsize                   | -0.07                       | 0.018                    | -0.08                          | 100                                        | Tract             | -0.15                |
| eligis                    | -0.10                       | 0.010                    | -0.10                          | 100                                        | Tract             | NA                   |
| built50s                  | -0.03                       | 0.008                    | -0.03                          | 99                                         | Block group       | 0.06                 |
| upperquart                | 0.05                        | 0.017                    | 0.06                           | 98                                         | Tract             | NA                   |
| built40s                  | -0.02                       | 0.007                    | -0.02                          | 98                                         | Block group       | NA                   |
| value_3000to50k           | 0.05                        | 0.016                    | 0.05                           | 97                                         | Block group       | NA                   |
| value_1000to2000k         | 0.04                        | 0.013                    | 0.04                           | 97                                         | Block group       | NA                   |
| mortgage_2000to500perincome| -0.02                      | 0.008                    | -0.02                          | 97                                         | Block group       | NA                   |
| age67to79                 | 0.05                        | 0.018                    | 0.06                           | 94                                         | Tract             | NA                   |
| husband/child             | 0.07                        | 0.019                    | 0.06                           | 93                                         | Tract             | -0.26                |
| mortgage_2000to500HE      | 0.09                        | 0.009                    | 0.03                           | 83                                         | Block group       | 0.19                 |
| mortgage                  | -0.02                       | 0.019                    | -0.03                          | 76                                         | Block group       | NA                   |
| black                     | 0.02                        | 0.010                    | 0.03                           | 75                                         | Tract             | -0.19                |
| worker                    | 0.02                        | 0.011                    | 0.03                           | 72                                         | Block group       | NA                   |
| freeworkers               | 0.02                        | 0.009                    | 0.03                           | 68                                         | Tract             | NA                   |
| age4to59                  | NA                          | NA                       | 0.03                           | 59                                         | Tract             | NA                   |
| child                     | -0.06                       | 0.021                    | -0.05                          | 52                                         | Tract             | 0.29                 |
| total                     | -0.03                       | 0.018                    | -0.04                          | 46                                         | Tract             | -0.16                |
| built50s                  | NA                          | NA                       | -0.02                          | 42                                         | Block group       | 0.07                 |
| mortgage_2000to500HE      | -0.01                       | 0.007                    | -0.01                          | 35                                         | Block group       | NA                   |
| occupied                  | NA                          | NA                       | -0.02                          | 34                                         | Tract             | NA                   |
| age50to59                 | NA                          | NA                       | -0.02                          | 34                                         | Tract             | 0.10                 |
| R-squared                 | 0.48                        | 0.50                     | 0.50                           | 0.55                                       |                  |                      |
| MSE                       | 0.48                        | 0.51                     | 0.51                           | 0.38                                       |                  |                      |

Legend: DPI, ACS, Canopy, Polk, Foreclosure, Insolation, CSI

Figure 2. Subsets of the most predictive variables, with regression coefficients and associated R-squares, calculated using all the PV adoption data at block group resolution.
To explore the general applicability of models constrained using data from one region to other regions, we developed a ‘baseline’ model trained using PV adoption data at the ZIP code level from each of the three California IOU territories, and then applied this to each of the two remaining utility regions. Table 5 presents the results, with regression coefficients from each baseline model shown in columns (all coefficients are significant at 5% or less). The coefficients coded in dark green, light green and yellow indicate whether the particular variable was identified as a variable in the best (leaps) subset of 1, 3 and 5 variables, respectively. The bottom rows identified the adjusted $R^2$-squared for each baseline model applied to both the regions it was trained on (in red), and to the two other IOU regions (in black). For comparison, the last column includes the variables selected for the model that included all of the CSI data, and the adjusted $R^2$-squared for that model.

Table 4 shows that model performance improved, in all regions, by training the model with regional data instead of data from all IOU regions. This suggests that relationships between population variables and PV adoption differ across regions, and increased regional specificity allows the models to estimate more efficient parameters in regional models relative to a pooled model. However, applying the models trained in different IOU regions did not generally provide a substantial decrease in model performance. For example, the PG&E model adjusted $R^2$-squared only decreased from 0.63 to 0.58 and 0.57 when relying on the SDG&E- and SCE-trained model parameters, respectively. In part, this was likely

| Variable | PGE | SDGE | SCE | All CSI |
|----------|-----|------|-----|---------|
| percentTPO | -0.21 | -0.24 |       |         |
| age659950 | 0.24 | 0.15 | 0.22 |         |
| age60to90 | -0.24 | -0.23 | -0.12 |         |
| age6070 | 0.35 | 0.15 | 0.11 |         |
| over60 | 0.11 | 0.53 | 0.24 |         |
| black | -0.13 | 0.26 | -0.19 |         |
| asian | -0.09 | 0.27 | 0.08 |         |
| notowned | -0.27 |         |         |         |
| occupied | 0.29 | 0.54 | 0.78 | 0.19 |
| mortgage2+HE | 0.10 | -0.30 | 0.27 | 0.19 |
| norentmortgage2+HE | -0.22 | -0.22 | 0.11 |         |
| mortgage_20to40pcincome | -0.11 |         |         |         |
| value_1000to20k | 0.16 | 0.18 | 0.16 |         |
| value_200to500k | 0.17 | 0.88 | 0.29 |         |
| value_500to1ml | 0.45 |         |         |         |
| traveltime | -0.11 | -0.16 |       |         |
| avgvehicles | -0.79 |         |         |         |
| daycare | 0.09 | 0.06 | 0.27 |         |
| rooms | 0.24 | 0.50 | 0.40 | 0.27 |
| built1000s | -0.54 |         |         |         |
| built500 | 0.12 | 0.06 | 0.07 |         |
| built30 | -0.36 |         |         |         |
| built40 | -0.10 | -0.25 |       |         |
| wood | -0.58 | -0.23 |       |         |
| dollars | -1.06 |         |         |         |
| foreclosure | 0.23 | 0.31 | 0.33 | 0.26 |
| hev | 0.30 | 0.23 | 0.15 | 0.25 |
| diesel | 0.30 | 0.12 | 0.02 |         |
| highschool | 0.51 |         |         |         |
| bachelor | -0.33 | -0.19 |       |         |
| masters | 0.31 | 0.58 | 0.32 |         |
| pcc | -0.09 | 0.41 | -0.09 |         |
| isolation | 0.17 | 0.40 | 0.25 | 0.17 |
| mortgage2+HE | -0.70 |         |         |         |
| mortgage2+HE | -0.08 |         |         |         |
| single | -0.08 |         |         |         |
| twor | 0.15 |         |         |         |
| three | 0.19 |         |         |         |
| tenure | 0.10 |       |       |         |
| Adj. R-squared - PGE | 0.56 | 0.59 | 0.61 |         |
| Adj. R-squared - SDGE | 0.61 | 0.80 | 0.60 |         |
| Adj. R-squared - SCE | 0.59 | 0.57 | 0.63 |         |
| Adj. R-squared - All CSI | NA | NA | NA | 0.55 |
| Number of zip codes | 550 | 68 | 369 | 1139 |

11 Variables identified in any size subset up to five variables are color-coded accordingly. As subset size increases, some variables are swapped out for other variables; as a result, more than five variables are ultimately color coded. Further, in some cases, a variable was identified to be included in a model with a limited subset, but not in the best-fit model.
attributed across several common key explanatory variables identified across all three models: mortgage2orHE, hev and rooms. SDG&E provided an exception—the adjusted \( R^2 \) squared of other models applied to the SDG&E area provided substantially less explanatory power relative to the model trained using SDG&E adoption data. Despite being a relatively large geographic region, SDG&E has far fewer ZIP codes (68, compared to 550 and 369 for PG&E and SCE, respectively). As a result, substantial data variability may be averaged out, making the model more sensitive to inclusion/exclusion of regionally explanatory variables.

### 7. Discussion and future work

We find three key takeaways from the California PV adoption trends. First, we find that relatively small subsets of geospatial data could be nearly as predictive of historical PV adoption trends as much larger subsets of geospatial data. Several parameters from the ACS data (home age, heating source, number of rooms, mortgage status and household education) and single fields from foreclosure data, vehicle registration and solar insolation data provided key PV adoption indicators. This suggests that model diffusion parameters may be best constrained using relatively small subsets of data rather than trying to include as many sources of geospatial information as possible. Further, several of the signs of the estimated parameters are consistent with the literature, while several other variables have no precedent in the literature. Namely, number of rooms, education, house age, solar insolation, hybrid car ownership and having a second mortgage or home equity loan are found to positively correlate with PV adoption. Areas with a high reliance on wood heating source are found to negatively correlate with PV adoption.

Second, we find that the subsets of data that are most predictive of PV adoption vary for models trained at different spatial resolutions. Geospatial data with relatively coarse spatial resolution (e.g., ZIP code) is not particularly useful for representing variations in higher resolution PV adoption trends. This suggests that the types of data that are useful for informing and constraining PV diffusion dynamics for high spatial resolution models could be fundamentally limited compared to the data that could be used to constrain lower resolution models.

Third, we find that PV diffusion characteristics are regional, and the predictive performance of regression models can be improved by regionally constraining fit parameters in a model. However, we do find that within California, the association between historical PV diffusion trends and population statistics are similar enough that the models trained in one region are reasonably representative of different regions. One exception to this was the SDG&E region, where regionally-trained models performed much better, possibly because of the relative low number of ZIP codes and high homogeneity between ZIP codes within that region.

While some of these best performing variables are consistent with the existing literature on the demographic characteristics of green technology adopters, and, more specifically, PV adopters (namely education, race and home value) most variables have not been previously been explored in the context of PV adoption. Number of rooms, heating source and house age were key variables that had not been previously explored in the literature, but are intuitively consistent with the expected profile of a PV adopter. The strong relationship provided by foreclosure indicators and mortgage status have less of a clear relationship to PV adoption, but may be highly correlated with characteristics inherent in PV adopters.

This analysis excluded several key datasets that likely drive adoption. These include data characterizing the range in value for PV-generated electricity both within and between regions in California based on the incentives available when the PV systems were installed, the cost of PV systems, and household electricity costs. Future research aims to further refine parameters that may feed into diffusion models by evaluating diffusion dynamics outside of California as well as if, and how, diffusion dynamics have evolved over time. Improving upon current models would serve to better inform multiple solar stakeholders including utility generation planners, regulators, policy-makers, and solar companies.

### Acknowledgments

This work was supported by the US Department of Energy under contract number DE-AC36-08GO28308. The authors would like to thank the following individuals and organizations for their contributions to and review of this work: Michael Gleason, Dylan Hettinger and David Keyser.

### Appendix. Description of variables and summary statistics

| Variable name | Variable description | Unit | Native resolution |
|---------------|----------------------|------|-------------------|
| percent TPO   | Percentage of TPO systems | Percent | Address |
|CSI data      |                       |      |                   |
| total         | Total population     | Total number | Tract |
| under20       | Percent of total population under 20 | Percent | Tract |
| age20to30     | Percent of total population age 20–29 | Percent | Tract |
| age30to40     | Percent of total population age 30–39 | Percent | Tract |
| Variable name | Variable description | Unit | Native resolution |
|---------------|----------------------|------|------------------|
| age40to50     | Percent of total population age 40–49 | Percent | Tract |
| age50to60     | Percent of total population age 50–59 | Percent | Tract |
| age60to70     | Percent of total population age 60–69 | Percent | Tract |
| age70to80     | Percent of total population age 70–79 | Percent | Tract |
| over80        | Percent of total population age 80+ | Percent | Tract |
| white         | Percent of total population, white | Percent | Tract |
| black         | Percent of total population, black | Percent | Tract |
| child         | Percent households with one or more children | Percent | Tract |
| husbandwife  | Percent of husband–wife households | Percent | Tract |
| famsize       | Family size | Average | Tract |
| occupied      | Total housing units—occupied | Percent | Tract |
| ownerocc      | Total housing units that are occupied by owner | Percent | Tract |

ACS data

| Variable name | Variable description | Unit | Native resolution |
|---------------|----------------------|------|------------------|
| toownerocc    | Owner-occupied housing units | Total number | Tract |
| samehouse     | Same house 1 year ago | Percent of total population | Tract |
| car           | Car transportation to work | Percent of total working population | Tract |
| wfh           | Work from home | Percent of total working population | Tract |
| traveltime    | Travel time to work | Average time, total working population | Tract |
| oneworkers    | Households with one worker in house | Percent of households | Tract |
| twoworkers    | Households with two workers in house | Percent of households | Tract |
| threeworkers  | Households with three workers in house | Percent of households | Tract |
| aggearn       | Aggregate earnings in past 12 months | Average $ total population | Block group |
| rooms         | Median number of rooms | Median, owner-occupied houses | Block group |
| mortgage      | Have mortgage | Average, owner-occupied houses | Block group |
| built2000s    | Owner occupied houses built after 2000 | Percent of owner-occupied houses | Block group |
| built90s      | Owner occupied houses built 1990-99 | Percent of owner-occupied houses | Block group |
| built80s      | Owner occupied houses built 1980-89 | Percent of owner-occupied houses | Block group |
| Variable name | Variable description                                                                 | Unit | Native resolution |
|---------------|---------------------------------------------------------------------------------------|------|-------------------|
| built70s      | Owner occupied houses built 1970–79                                                    | Percent of owner-occupied houses | Block group      |
| built60s      | Owner occupied houses built 1960–69                                                    | Percent of owner-occupied houses | Block group      |
| built50s      | Owner occupied houses built 1950–59                                                    | Percent of owner-occupied houses | Block group      |
| built40s      | Owner occupied houses built 1940–49                                                    | Percent of owner-occupied houses | Block group      |
| utilgas       | House heating fuel supplied by utility                                                 | Percent of total owner-occupied houses | Tract          |
| wood          | Wood heating                                                                           | Percent of total owner-occupied houses | Tract          |
| avgvehicles   | Vehicles available                                                                     | Average, owner-occupied houses | Block group      |
| upperquint    | Upper quintile house value                                                             | Upper quintile house value | Block group      |
| mortgage2orHE | Mortgage status: houses with either a second mortgage or home equity loan, but not both | Percent of total houses with mortgages | Block group      |
| mortgage2andHE | Mortgage status: houses with both a second mortgage and a home equity loan            | Percent of total houses with mortgages | Block group      |
| neithermortgage2norHE | Mortgage status: houses with neither a second mortgage or home equity loan       | Percent of total houses with mortgages | Block group      |
| mortgage.under20pctincome | Mortgage under 20% of household income                                               | Percent of total houses with mortgages | Block group      |
| mortgage.20to40pctincome  | Mortgage between 20% and 40% of household income                                      | Percent of total houses with mortgages | Block group      |
| mortgage.over40pctincome   | Mortgage over 40% of household income                                                 | Percent of total houses with mortgages | Block group      |
| value.under.100 k | Houses valued under $100 000                                                           | Percent of total owner-occupied houses | Block group      |
| Variable name       | Variable description                  | Unit                                   | Native resolution |
|---------------------|---------------------------------------|----------------------------------------|-------------------|
| value.100to200 k    | Houses valued $100 K – $199 K         | Percent of total owner-occupied houses | Block group       |
| value.200to300k     | Houses valued $200 K – $299 K          | Percent of total owner-occupied houses | Block group       |
| value.300to500 k    | Houses valued $300 K – $499 K          | Percent of total owner-occupied houses | Block group       |
| value.500 kto1 mil  | Houses valued $500 000–$999 999        | Percent of total owner-occupied houses | Block group       |
| value.over1 mil     | Houses valued over $100 0000           | Percent of total owner-occupied houses | Block group       |
| hrsworked           | Mean hours worked                      | Mean hours for population 16 to 64     | Tract             |

Canopy cover

| canopy              | Percent canopy cover                  | Percent canopy cover                  | 30 m               |

Foreclosure data

| foreclosure         | Foreclosure risk score                | Normalized score by state ranging from 1–100 | ZIP code           |

Insolation

| insolation                  | Global horizontal insolation          | Global horizontal insolation           | 10 km              |

Polk

| hev                        | Number of registered hybrid electric cars | Number of vehicles | ZIP code |
|----------------------------|-------------------------------------------|--------------------|----------|
| elec                       | Number of registered electric cars        | Number of vehicles | ZIP code |
| diesel                     | Number of registered diesel cars          | Number of vehicles | ZIP code |

Summary Statistics

| Variable | Mean   | Median | SD     | Minimum | Maximum |
|----------|--------|--------|--------|---------|---------|
| numinstalls | 50.18  | 22.50  | 69.75  | 1.00    | 358.00  |
## Demographic data

| Variable       | Mean | Median | SD  | Minimum | Maximum |
|----------------|------|--------|-----|---------|---------|
| numinstalls-lg | 3.10 | 3.11   | 1.35| 0.00    | 5.88    |
| percentTPO     | 53%  | 53%    | 21% | 6%      | 100%    |

| Total          | 4942.53 | 4817.86 | 1105.25 | 2208.00 | 8714.00 |
| under20        | 32%     | 33%     | 5%      | 18%     | 43%     |
| age20to30      | 15%     | 15%     | 3%      | 8%      | 28%     |
| age30to40      | 13%     | 13%     | 2%      | 8%      | 18%     |
| age40to50      | 13%     | 13%     | 1%      | 9%      | 16%     |
| age50to60      | 12%     | 12%     | 2%      | 7%      | 20%     |
| age60to70      | 8%      | 7%      | 3%      | 4%      | 19%     |
| age70to80      | 5%      | 4%      | 2%      | 2%      | 11%     |
| over80         | 3%      | 2%      | 1%      | 1%      | 7%      |
| white          | 55%     | 53%     | 17%     | 7%      | 91%     |
| black          | 9%      | 6%      | 12%     | 0%      | 78%     |
| child          | 34%     | 35%     | 5%      | 21%     | 44%     |
| husbandwife    | 53%     | 54%     | 8%      | 27%     | 65%     |
| famsize        | 3.68    | 3.75    | 0.43   | 2.70    | 4.59    |
| occupied       | 88%     | 91%     | 9%      | 59%     | 97%     |
| ownerocc       | 59%     | 61%     | 12%     | 28%     | 85%     |

## ACS data

| Variable          | Mean     | Median    | SD       | Minimum | Maximum   |
|-------------------|----------|-----------|----------|---------|-----------|
| totownerocc       | 880.47   | 851.07    | 308.10   | 267.13  | 1791.67   |
| samehouse         | 85%      | 86%       | 5%       | 64%     | 96%       |
| car               | 88%      | 89%       | 5%       | 65%     | 95%       |
| wfh               | 5%       | 4%        | 2%       | 1%      | 15%       |
| traveltime        | 26.93    | 26.70     | 5.83     | 14.43   | 45.53     |
| oneworkers        | 27%      | 27%       | 7%       | 13%     | 50%       |
| twoworkers        | 39%      | 39%       | 4%       | 26%     | 47%       |
| threeworkers      | 26%      | 26%       | 4%       | 16%     | 38%       |
| ageearn           | $16 772  | $16 020   | $6450    | $4826   | $37 128   |
| upperquint        | $86 581  | $86 629   | $14 799  | $48 381 | $130 116  |
| rooms             | 5.66     | 5.69      | 0.36     | 4.56    | 6.85      |
| mortgage          | 71%      | 72%       | 9%       | 42%     | 91%       |
| built2000s        | 13%      | 11%       | 11%      | 0%      | 69%       |
| built80s          | 17%      | 15%       | 11%      | 1%      | 60%       |
| built90s          | 13%      | 12%       | 8%       | 0%      | 34%       |
| built70s          | 14%      | 14%       | 7%       | 2%      | 54%       |
| built60s          | 11%      | 11%       | 5%       | 1%      | 43%       |
| built50s          | 15%      | 11%       | 11%      | 0%      | 62%       |
| built40s          | 9%       | 6%        | 8%       | 0%      | 34%       |
| yrmoved           | 1996     | 1996      | 3        | 1988    | 2002      |
| utilgas           | 62%      | 70%       | 23%      | 3%      | 91%       |
| wood              | 6%       | 2%        | 8%       | 0%      | 48%       |
| avgvehicles       | 1.41     | 1.39      | 0.38     | 0.62    | 2.98      |
| upperquint        | $404 464 | $400 166 | $107 700| $215 086| $918 300  |
| mortgage2orHE     | 15%      | 15%       | 5%       | 2%      | 30%       |
| mortgage2andHE    | 1%       | 1%        | 1%       | 0%      | 5%        |
| neithermortgage2norHE | 55% | 55% | 8% | 35% | 74% |
| mortgaugeunder20ptinctime | 23% | 22% | 6% | 8% | 44% |
| mortgage.20to40ptinctime | 29% | 28% | 5% | 9% | 47% |
| mortgage.over40ptinctime | 36% | 36% | 8% | 2% | 66% |
| value.under.100 k | 30% | 29% | 7% | 10% | 53% |
| value.100to200 k  | 21% | 22% | 11% | 1% | 48% |
| value.200to300 k  | 21% | 21% | 6% | 5% | 38% |
| value.300to500 k  | 29% | 28% | 13% | 4% | 56% |
| value.500 kto1 mil | 13% | 10% | 10% | 1% | 53% |
| value.over1 mil   | 2%  | 1%  | 2%  | 0%  | 18%      |
| hrsworked         | 38.74    | 38.54    | 1.44     | 34.95   | 43.85     |
| Variable    | Mean | Median | SD  | Minimum | Maximum |
|-------------|------|--------|-----|---------|---------|
| Canopy cover|      |        |     |         |         |
| canopy      | 6.27 | 1.51   | 11.09 | 0.00    | 62.17   |
| Foreclosure data | |        |     |         |         |
| foreclosure | 22.31| 16.45  | 21.30 | 0.10    | 100.00  |
| Insolation  |      |        |     |         |         |
| insolation  | 4.99 | 4.98   | 0.27 | 3.93    | 5.76    |
| Polk        |      |        |     |         |         |
| diesel      | 5.70 | 1.00   | 10.94 | 0.00    | 80.00   |
| elec        | 367.76| 250.00 | 363.81 | 16.00 | 2028.00 |
| hev         | 20 358.17| 16 215.00 | 19 979.09 | 186.00 | 103 241.00 |

References

Beck R W 2009 Distributed renewable energy operating impacts and valuation study (Prepared for Arizona Public Service by R W Beck)

Bollinger B and Gillingham K 2012 Peer effects in the diffusion of solar photovoltaic panels Mark. Sci. 31 900–12

Cai D W H et al 2013 Impact of residential PV adoption on retail electricity rates Energy Policy doi:10.1016/j.enpol.2013.07.009

Darghouth Naim R, Galen Barbosa and Ryan H Wiser 2014 Customer-economics of residential photovoltaic systems (part 1): the impact of high renewable energy penetrations on electricity bill savings with net metering Energy Policy 67 290–300

Denholm P, Drury E and Margolis R 2009 The Solar Deployment System (SolarDS) Model: Documentation and Sample Results (Golden, CO: National Renewable Energy Laboratory) NREL/TP 6A2-45832

Diamantopoulos A, Schlegelmilch B B, Sinkovics R R and Bohlen G M 2003 Can socio-demographics still play a role in profiling green consumers? a review of the evidence and an empirical investigation J. Bus. Res. 56 465–80

Drury E et al 2012 The transformation of Southern California’s residential photovoltaics market through third-party ownership Energy Policy 42 681–90

EIA 2012 NEMS Residential Demand Module Documentation Report (Washington, DC: US Department of Energy)

Faiers A, Neame C and Cook M 2007 The adoption of domestic solar-power systems: do consumers assess product attributes in a stepwise process? Energy Policy 35 3418–23

Greene W 2011 Econometric Analysis 7th edn (Englewood Cliffs, NJ: Prentice-Hall)

Guidolin M and Mortarino C 2010 Cross-country diffusion of photovoltaic systems: modelling choices and forecasts for national adoption patterns Technol. Forecast. Soc. Change 77 279–96

Gujarati D 2011 Essentials of Econometrics (New York: McGraw-Hill)

Hines J, Hungerford H and Tomera A 1986 Analysis & synthesis of research on responsible environmental behavior: a meta-analysis J. Environ. Educ. 18 1–8

Kutner M, Nachtstein C and Neter J 2004 Applied Linear Regression Analysis 4th edn M-Hill/Irwin (New York: McGraw-Hill/Irwin)

Kwan C L 2012 Influence of local environmental, social, economic and political variables on the spatial distribution of residential solar PV arrays across the US Energy Policy 47 332–44

Paidipati J, Frantzis L, Sawyer H and Kur rash A 2008 Rooftop Photovoltaics Market Penetration Scenarios NREL/Sr-581-42306 (Burlington, MA: Navigant Consulting)

Rai V and Robinson S 2013 Effective information channels for reducing costs of environmentally-friendly technologies: evidence from residential PV markets Environ. Res. Lett. 8 014044

Rai V and Sigrin B 2013 Diffusion of environmentally-friendly energy technologies: buy versus lease differences in residential PV markets Environ. Res. Lett. 8 014022

Zhang Y, Song J and Hamori S 2011 Impact of subsidy policies on diffusion of photovoltaic power generation Energy Policy 39 1958–64