Environmental Preferences and Peer Effects in the Diffusion of Solar Photovoltaic Panels *

Bryan Bollinger
Stanford University

Kenneth Gillingham
Stanford University

Revise & Resubmit at *Marketing Science*

August 2010

Abstract

As solar photovoltaic (PV) technology diffuses in California, a clear geographic clustering pattern has emerged. This paper documents the pattern of clustering and explores several reasons why it may come about. Specifically, we examine the influence of environmental preferences and peer effects due to social learning or image motivation. We present quasi-experimental evidence suggesting that peer effects play a critical role in contributing to the clustering. We then develop a hazard model of technology adoption, and find that a 1% increase in the number of installations in a zip code decreases the time until the next adoption by roughly 1%, even after controlling for possibly changing unobserved heterogeneity across zip codes and time. These results provide evidence supporting the importance of peer effects, perhaps along with localized marketing to take advantage of these peer effects, in the diffusion of solar PV technology. Greater environmental preferences also correspond to increased adoption, but appear to reduce the importance of peer effects. The positive influence of previous adoptions on the decision to adopt holds even at the street-level. These results provide insight into the influences of peer effects and environmental preferences in the adoption of a key fledgling energy technology.

*The authors would like to thank Hunt Allcott, Tim Bresnahan, Jesse Cunha, Misha Dworsky, Wesley Hartmann, Sridhar Narayanan, Peter Reiss, Arthur van Benthem, Ali Yurukoglu, the participants in the Stanford Economics Department Environmental Reading Group, Stanford Economics IO Workshop, the Stanford MS&E Policy and Economics Research Roundtable, the UC-Berkeley ARE seminar series, the Energy Institute at Haas seminar series, and faculty and students from the Stanford Graduate School of Business for their valuable comments and insights. We would like to especially thank Adam Leising for providing the street-level data. Any errors are solely the responsibility of the authors.

1 Introduction

The diffusion of solar photovoltaic (PV) technology is of critical interest to policy-makers invested in promoting residential adoption of solar PV panels. Germany, Japan and Spain are just a few of the countries that have taken major policy actions to encourage the diffusion of solar PV technology for environmental and national energy security reasons. California followed a similar path in 2006 with the $3.3 billion California Solar Initiative (CSI), which provides 10 years of subsidies for solar PV panels. Along with substantial Federal subsidies for solar PV panels, the CSI rebates dramatically lower the cost to a household of a solar PV installation. The resulting increase in solar adoption in California has been dramatic, with nearly a doubling of new installations per year between 2006 and 2009.

This paper documents an intriguing pattern of geographic clustering in the diffusion of solar PV panels in California, and explores the factors that may lead to such a pattern. Environmental preferences are one obvious reason why we might see clustering in the adoption of any fledgling green technology, such as solar PV panels, hybrid vehicles, or LEED buildings (Kahn and Vaughn 2009). However, our empirical results suggest that there is more to this pattern of clustering than just environmental preferences. In particular, we find evidence suggestive of peer effects, whereby one household's choice to adopt may be influenced by the previous adoption decisions of other nearby households. Peer effects are a class of social interactions that have been widely recognized in the economic, marketing, and sociological literature. These effects may occur due to information sharing in the process of social learning, or image motivation in which households receive utility from the conspicuous consumption of an environmentally-friendly good (Griskevicius et al. 2010). In either case, peer effects could be driven by actual communication between agents or simply through the visibility of the installations.

The possibility that peer effects may be important in the adoption of solar PV panels is anecdotally well-documented in the marketing reports of both the CSI administrators as well as many of the solar installers in California. The primary marketing strategy of several key players in the solar PV panel installation business explicitly aim to use peer effects to leverage their marketing in order to increase sales. The idea that social interactions influence technology diffusion and growth also has a long history in the academic literature. For example, knowledge spillovers have played a key role in endogenous growth theory (Romer 1986; Lucas 1988; Aghion and Howitt 1998) and much of the theory of technology diffusion (Griliches 1957; Frank et al. 1964; Arndt 1967; Bass 1969; Rogers 1995). Several studies empirically examine the quantitative importance of learning from others in the adoption of new agricultural technology or practices (Foster and Rosenzweig 1995; Munshi 2004; Bandiera and Rasul 2006; Conley and Udry 2010). Social interactions have also been studied in the diffusion of hybrid vehicles (Axsena et al. 2009) and the adoption of technologies to phase out lead (Newell and Kerr 2003). In a study even more closely related to this one, Lessem and Vaughn (2009) find that the political ideology of a neighborhood affects the adoption of solar PV installations in Sacramento, providing the first suggestive evidence of possible peer effects due to image motivation in solar PV installations in California.

Despite the considerable interest in identifying peer effects in the diffusion of new technologies, such identification is notoriously fraught with difficulty. There are three well-known issues that often confound identification of peer effects: endogenous group formation leading to self-selection of peers, correlated unobservables, and a particular type of simultaneity called "reflection" (Manski 1993; Brock and Durlaf 2001; Moffitt 2001; Soetevent 2006; Hartmann et al. 2008). Endogenous group formation occurs when agents with similar tastes form social groups, so that any correlation in their decisions may be due to common preferences, rather than peer effects. Correlated unobservables may be an issue when there are unobserved common group effects that influence all agents in the group simultaneously. Common environmental preferences would fall in this category. "Reflection," first described by Manski (1993), occurs when decisions are made by all individuals in a group simultaneously, so that correlation in subsequent actions could simply reflect the fact that one agent's decision affects the decisions of others in the group, just as the decisions of others may affect that agent's decision.

Our empirical strategy is designed to mitigate these issues by exploiting both quasi-experimental variation from differing incentives in zip codes split by the CSI administration zone boundary and zip code-time fixed effects at both the zip code and street level. Our quasi-experimental variation allows us to very cleanly control for changing heterogeneity in environmental preferences or any other correlated unobservable. However, it is limited to a particular subset of zip codes. Our zip code-time fixed effects approach allows us to examine evidence for peer effects at the broader scale in California, while still controlling for many possibly correlated unobservables, such as differing or changing environmental preferences. Each of our empirical specifications are designed to examine the effect of *past* installations on the next decision to adopt, so reflection is not a concern. When we examine the effect of demographics on adoption, we find that adoption of another "green" good, hybrid vehicles, can help explain solar PV panel adoption, suggesting that environmental preferences play an important role. Interestingly, greater environmental preferences appear to associated with a smaller effect of previous nearby adoptions on the decision to adopt, perhaps suggesting that environmentalists are early adopters who do not need to be influenced as much by their neighbors.

The rest of the paper is organized as follows. Section 2 documents the pattern of geographic clustering of solar PV panels in California using our rich installation-level dataset. Section 3 provides quasi-experimental evidence suggestive of the importance of the presence of nearby adoptions on the rate of adoption – a result we interpret as suggestive of peer effects, perhaps along with localized marketing to take advantage of such peer effects. Section 4 contains our zip code-level analysis, in which we develop and estimate a hazard model of technology adoption in order to quantify the effect of previous adoptions in a zip code on the decision to adopt. We go on in that section to further decompose the determinants of these effects to see how environmental preferences may interact with peer effects to influence adoption. Section 5 presents an alternative street-level model to estimate the effect of previous adoptions on the probability to adopt at a finer geographic level, using a subset of our dataset that includes detailed address-level data. Section 6 concludes.

2 Clustering of Solar PV Panels in California

While solar PV technology has had a long history in California, it was not until the late 1990s and early 2000s that the California solar PV panel market really gained a foothold. In 1997 the California Energy Commission (CEC) Emerging Renewables Program subsidized solar PV installations with a $3 per Watt (W) rebate, to be renewed year-by-year. In 1998, California added "net metering," allowing owners of solar PV systems to receive credit for electricity sold back to the grid. In 2001, an up to 15% state tax credit was added.[1] In January 2006, the California Public Utilities Commission (CPUC) established the California Solar Initiative (CSI), the $3.3 billion, 10-year rebate program aiming to "install 3,000 MW of new solar over the next decade and to transform the market for solar energy by reducing the cost of solar" (CPUC 2009).[2] On top of this, the Federal Energy Policy Act of 2005 created a 30% tax credit for residential and commercial solar PV installations, but with a $2,000 limit, which was subsequently lifted in 2008. These substantial subsidies have contributed to the dramatic growth in annual solar PV adoptions over the past decade, from less than 1,000 residential installation per year in 2001 to almost 12,000 per year in 2009. To explore the pattern and determinants of this growth, we assemble an installation-level dataset of residential solar PV installations in the three large investor-owned utility (IOU) regions from 2001 to 2009.[3]

Prior to 2007, our data are from the CEC Emerging Renewables Program, and after January 2007 our data are from the CSI database. The data include the zip code of the

[1] The state tax credit remained in place through the end of 2005.

[2] For an overview of the history of solar PV policy in California, see Taylor (2008) and for further information regarding the CSI incentives, see Bollinger and Gillingham (2010).

[3] The three investor owned utilities Pacific Gas and Electric Company (PG&E), Southern California Edison (SCE) and San Diego Gas and Electric (SDG&E), cover nearly the entire state and over 90% of the solar PV market (CPUC 2009). Each of the municipal utilities are required to have a rebate program similar in generosity to the CSI, but we do not observe installations from these programs.

customer, utility, size of the installation and incentive, PV installer and manufacturer, the date when the customer reserved solar incentives for an installation, the date payment was submitted for the installation, and the date of completion. Our Emerging Renewables Program data also include the address of the installation, an essential component for examining patterns of clustering and for our street-level analysis. We further augment the installation data with zip code-level demographic data from Sourcebook America and American FactFinder, as well as data on hybrid vehicle registrations from R.L. Polk and Company. The cleaned dataset includes 33,685 completed residential installations between January 2001 and December 2009, 26,111 of which are from after the inception of the CSI. Further details on the cleaning of the dataset are included in an online appendix.

Table 1 contains zip code summary statistics for the total residential installations and demographics. The average number of residential installations in a zip code is 24.1.[4] The average sizes and prices of the residential installations are shown in Table 2. The installation price is adjusted by the CPI to real 2009 dollars per W and all Watts in this paper are direct current Watts. The average size of an installation is 4.9 kW, with an average pre-incentive price of $8.40 per W. This corresponds to an average system price in the range of $40,000 before incentives. Figures 1(a) and 1(b) show histograms for the time between completion dates for residential installations within a zip code and the logarithm of this time which is used in estimation.

With our data, we can explore patterns of adoption over time, both at the regional and neighborhood level. Figure 2(a) shows the initial pattern of clustering of solar PV panel installations in the San Francisco Bay Area from 2001 to 2003. More densely populated zip codes tend to have more installations, but by no means does the pattern simply follow the density of the zip codes. There are densely populated zip codes that have few installations, and less densely populated areas with many installations. Moreover, this pattern of clustering appears to build upon itself. Figure 2(b) shows the same map of the San Francisco Bay Area with installations from 2003 to 2006 also included. While there are more installations everywhere by 2006, the pattern of clustering very clearly continues.

[4]Conditional on having at least one installation.

Even more striking is that this pattern of clustering is clearly exhibited even within zip codes, at the neighborhood level. For example, Figure 3(a) shows a breakdown of neighborhoods in Berkeley with a color-coding for how environmental each neighborhood is.[5] Figure 3(b) shows the location of solar installations in these neighborhoods by the end of 2006. These regional and neighborhood patterns of clustering are representative of the pattern of diffusion of solar PV technology throughout California. The remainder of the paper explores why we might see such a pattern of clustering. Possible reasons include differences in environmental preferences or wealth across zip codes, localized marketing by installers, and peer effects.

The marketing by some successful solar installers suggests that there is more than environmental preferences and wealth contributing to this clustering behavior. For example, SolarCity is the largest installer in California and bases much of its marketing on finding one or two vocal members of the neighborhood and giving the entire neighborhood a slightly lower price if more adoptions are made within that neighborhood. SolarCity has used this strategy to grow rapidly, despite having some of the highest installation prices in the industry. Other companies have not been as successful with this particular marketing strategy, but even the PG&E CSI administrators note the value of peer effects by establishing "Solar Champion" training sessions for "SF citizens interested in helping spread the word about solar in their neighborhoods." These types of localized marking are primarily leveraging social learning by using neighbors to provide information, although they also appear to be taking advantage of peer pressure to be environmental. This is underscored by a popular type of localized advertising intended to leverage image motivation: putting up a sign indicating that a solar PV panel has been installed at that location. In these cases, the distinction between peer effects and localized marketing is blurred, since it is not clear whether additional installations induced by viewing this sign are induced due to peer effects or due to localized advertising. In the following analysis, we disentangle the effect on diffusion of possibly changing environmental preferences, other demographics, and peer effects.

[5] Courtesy of Bernt Wahl and the Neighborhood Team at the University of California-Berkeley

3 Quasi-Experimental Evidence

We begin our exploration by looking for the cleanest evidence possible that there is something more than environmental preferences or demographics driving the clustering behavior: a quasi-experiment. Ideally to test for peer effects we would want to have two geographic areas with identical environmental preferences, demographics, and macroeconomic shocks, so that we could randomly place a few solar PV installations in one of the areas and see whether clustering would occur. Our quasi-experimental design is a second-best option that alleviates any concerns of unobserved changing heterogeneity in preferences or demographics. On the other hand, it focuses on a relatively limited number of zip codes.

We focus on zip codes that are split by CSI administrative zones. There are eight zip codes split by the border between PG&E and SCE and 13 split by the border between SCE and CCSE. After examining maps of these zip codes carefully, we find that the utility regions seem to cut through zip codes in a quite random manner – even seeming to cut through neighborhoods. Thus we make the identifying assumption that within each split zip code, changes in demographics and environmental preferences are the same on each side of the zip code. The only time-varying difference between the two parts of each zip code results from the difference in the utility region. There are very few substantial differences between each of the utilities in terms of prices or quality of service, yet there is an important difference relating to the structure of the CSI incentives. The CSI incentives are on a "step" schedule, whereby the incentives drop to a lower level once a certain number of cumulative megawatts (MW) of solar PV technology have been installed in that administrative region. This implies that there are times when one CSI administrative zone meets the target MW and for a limited period has lower incentives than the other CSI administrative zone. Figure 4 shows this in a schematic. We can use this discrepancy in the incentive step level as a treatment effect, for it effectively acts as a "shock" that places more installations on one side of the zip code than the other.

We are most interested in what happens when the incentive steps are back in synch.

Specifically, we would like to know whether previous adoptions in the surrounding area influence the rate of adoption after this "shock" of additional installations. Thus, we use the log of the time between adoptions as our dependent variable. If there is a positive, causal relationship between the number of previous installations and the rate of future adoption, then we would expect the rate of adoption to remain higher in the side of the zip code that temporarily had higher incentives. This result could be viewed as evidence of peer effects or perhaps inertia in localized marketing that moved entirely to the side of the zip code with higher incentives and did not realign. Of course, the inertia in this localized marketing may occur exactly because of peer effects, if the marketing was aiming to leverage such effects. If there is no causal effect, then we would expect the adoption rates to realign.

The transition from incentive step four to five happened first for PG&E, then for SCE and finally by SDG&E.[6] To examine the treatment effect we perform an ordinary least squares estimation with the following specification:

$$\log(\Delta t) = \beta_0 + \beta_1 Util + \beta_1 S + \beta_2 A + \beta_3 Util \cdot S + \beta_4 Util \cdot A + \eta_z + \xi_t + \varepsilon,$$

where Δt is the time between the completion dates of the solar PV installations, $Util$ is an indicator variable for whether the utility of the observation is the one that received the "shock," B is an indicator for the period when the second utility had a higher incentive value (was still on step four), A is an indicator for the period after the "shock" when the incentives were re-equalized (both on step five). η_z are zip code indicators and ξ_t are month indicators, which are included to control for unobservable trends and differences across zip codes that might be confound our results. ε is assumed to be a mean-zero stochastic error term. An observation in this estimation is an installation. The results are given in Table 3. Along both borders, the adoption rate increases (time between adoptions decreases) for the administrative district which has not changed step (i.e., received the

[6]All three of these utility districts moved from incentive step two to three and three to four at approximately the same time.

shock), as shown by the negative coefficient on the interaction β_3. This result is exactly as expected. What is more interesting is that in both cases the adoption rate remains higher after the incentives are realigned, as shown by the coefficient on the the second interaction β_4. Thus, there were more installations in the side of the zip code that received the shock *and this effect continued even after the incentives were realigned*.

4 Zip Code-Level Analysis

Our primary empirical approach analyzes how the rate of diffusion of solar PV panels in each zip code is influenced by the cumulative number of previous adoptions in that zip code, while controlling for possibly correlated time and area-specific unobservables. This approach also allows for an exploration of the effect of environmental preferences and demographics on the rate of adoption. We use a hazard function approach, perhaps the most common empirical method used to examine the determinants of technology adoption (e.g., see Hannan and McDowell (1984), Mulligan (2003), Baker (2001), and Engers et al. (2009)), and particularly well-suited to shed light on our questions of interest.[7] Many diffusion models, such as the Bass model (Bass 1969), can be derived from hazard models and are based on the idea that the probability of that any one consumer adopts depends on the number of previous adoptions. Importantly, a hazard function approach can capture the dynamics of diffusion in the early stage of the "S-shaped" adoption curve, since it quantifies the accelerating rate of adoption at this early stage. Adoptions of solar PV panels in California appear to be following the first stage of the "S-shaped" adoption curve, as they are increasing rapidly and appear to be accelerating. Furthermore, the market is far from saturated, as the average number of installations in a zip code with at least one

[7]Other common approaches include: using a normal, censored regression for the date of adoption (Lee and Waldman 1985; Genesove 1999), using a linear model if the outcome variable is continuous (e.g., Moretti (2008) examines the effect of movie reviews on box office revenues), or by assuming a linear probability model (Goolsbee and Klenow 2002; Duflo and Saez 2003; Gowrisankaran and Stavins 2004; Oster and Thorton 2009). Recent structural work incorporating demand-side spillovers (usually network effects) include Hamilton and McManus (2005), Lenzo (2006), Schmidt-Dengler (2006), Tucker (2004), Gowrisankaran and Stavins (2004), Tucker (2008) and Shriver (2010).

installation is around 24, out of an average of about 5,000 owner-occupied homes. This provides further evidence suggesting a hazard model is a useful approach to modeling this early stage of the diffusion of solar PV panels in California.

The most common hazard function approach assumes a homogenous Poisson process. Estimation then proceeds using maximum likelihood, based on an assumed distribution of the hazard function (e.g., exponential, weibull, lognormal). There are two concerns with using this approach in our context. First, assuming a homogenous Poisson process imposes the restriction that the mean of the distribution must equal the variance. We see no reason why this should be true in our context. Indeed, there large variation in the data of the time between adoptions, which implies a standard deviation of 329 days, which is much larger than the mean of 250 days. Second, Heckman and Singer (1984) demonstrate that if there is unobserved heterogeneity in the hazard rate then failing to account for this unobserved heterogeneity can lead to severely biased estimates of the coefficients of interest. Furthermore, Heckman and Singer also show that if there is unobserved heterogeneity in the hazard function, the hazard function parameter estimates are sensitive to the distributional assumptions on that heterogeneity. One way to model heterogeneity parametrically is to condition the survival function on an unobserved heterogeneity component. Heckman and Singer argue that this approach tends to over-parameterize the survival distribution and may lead to biased estimates.

We believe that modeling unobserved heterogeneity in the hazard rate may be important in our context. Thus, rather than a homogenous Poisson process, we assume a non-homogenous Poisson process (i.e., the hazard function is a function time). Specifically, we assume a doubly-stochastic or Cox Poisson process. This choice allows the mean and variance of the distribution of unobserved heterogeneity to differ. In addition, using a doubly-stochastic Poisson process allows us to directly include fixed effects and indicator variables to control for the possibility of the unobserved heterogeneity in the hazard rate. We feel that this approach is more benign than assuming away unobserved heterogeneity or simply modeling heterogeneity parametrically by conditioning the survival function on the unobserved heterogeneity.[8]

4.1 Hazard Model of Technology Diffusion

We assume the below exponential form for the hazard rate of solar adoption within a zip code z, because it both allows us to include a rich set of controls and ensures a positive hazard rate:

$$\lambda_z(t) = \lambda_{z0} \exp(X_{zt}\beta + f(z, t, \epsilon_{zt})). \tag{1}$$

Here X_{zn} are zip code and time-specific explanatory variables that do not vary between installations. $f(z, t, \epsilon_{zt})$ includes zip code and time control variables and a stochastic term capturing the unobserved heterogeneity in the hazard rate which may change on a daily basis. The heterogeneity captured by ϵ_{zt} includes unobserved factors influencing both the utility received from an installation and the frequency at which consumers make the decision whether to adopt or not. These factors may be due to zip code specific marketing, changing preferences of consumers in a zip code, or anything else that is not picked up by the explanatory variables or the zip code and time control variables.

This is a non-homogenous Poisson process since the hazard rate depends on time, due to its dependence on the time control variables as well as the stochastic shocks. A non-homogenous Poisson process can be treated as a homogenous process with hazard rate equal to $\lambda_{zn} = \int_{t_{n-1}}^{t_n} \lambda_z(t)dt$ for the period between installation $n-1$ and n. By assuming a Poisson arrival process, we have exponentially distributed waiting times between adoptions:

$$f(\Delta t_{zn}) = \frac{1}{\lambda_{zn}} \exp\left(-\Delta t_{zt}/\lambda_{zn}\right), \tag{2}$$

where we define $\Delta t_{zn} = t_{zn} - t_{zn-1}$ as the time between adoptions in the zip code. The

[8] Unfortunately, since the distributions of the hazard function and the unobserved heterogeneity are not separately identified, there is no way to test whether this is the appropriate assumption, but it appears very plausible to us that there are zip code-specific factors determining the hazard rate that we cannot control for. The use of a doubly-stochastic Poisson process to model the adoption decisions within a geographic region can be justified from a continuous-time model of individual-level discrete choices, as shown in our online Appendix.

conditional log likelihood function (conditional on λ_{zn}) takes the following form:

$$L = -\sum_{z,n} (\Delta t_{zn}/\lambda_{zn} + \log(\lambda_{zn})). \tag{3}$$

Maximizing this likelihood function with respect to λ_{zn} yields the following first-order condition:

$$\frac{d}{d\lambda_{zn}} - \sum_{z,n} (\Delta t_{zn}/\lambda_{zn} + \log(\lambda_{zn})) = 0. \tag{4}$$

Evaluating the derivative yields the following expression:

$$\log(\Delta t_{zn}) = \log(\lambda_{zn}) = \log(\lambda_{z0}) + X_{zn}\beta + \log\left(\int_{t_{n-1}}^{t_n} \exp f(z,t,\epsilon_{zt})\right). \tag{5}$$

This final equality follows since λ_{z0} and X_{zn} do not change over time. Note that for the time control variables that change in between installations, we need to integrate over the time since the last adoption. However, since the coefficients on these variables are not of interest to us, we replace the integrand expression with a linear function which includes zip code indicator variables and time variables which are equal to the fraction of Δt_{zn} spent in that month, as well as a linear stochastic shock, ϵ_{zn}, which includes the aggregate effect of the daily stochastic shocks in between adoptions.[9] Mathematically, we assume that:

$$\log\left(\int_{t_{n-1}}^{t_n} \exp f(z,t,\epsilon_{zt})\right) = \eta_{zn} + \epsilon_{zn}, \tag{6}$$

where η_{zn} either equals both zip and time control variables or zip-time control variables.[10] In the specification in which we include zip-specific quarter indicator variables, we drop observations when the previous adoption in the zip code occurred in the previous quar-

[9]If, for example, the last installation occurred a week before the end of last month and the current installation occurred a week into the current month, both month variables would have values of 0.5. If the entire time between installations is in a single month, the time control variables result in standard indicator variables. The time variables may or may not be zip-code specific e.g., zip-quarter indicator variables.

[10]As a robustness check to make sure that our results are not driven by these modeling assumptions, we run the regression using only the observations in which the time between adoptions occurs within a single month, using standard month dummies. The results are almost exactly the same.

ter, so that the zip-quarter indicator variable remains the same over the period of time between adoptions. These are technical details which are only needed to correctly account for the changing hazard rate, due to the fact that the Poisson process is non-homogenous.

Since $\log(\lambda_{z0})$ is subsumed in the constant, we can rewrite the expression for log time between adoptions as:

$$\log(\Delta t_{zn}) = X_{zn}\beta + \eta_{zn} + \epsilon_{zn}. \tag{7}$$

This expression is our primary specification for estimation, which models the natural logarithm of the time between successive adoptions as a linear function of explanatory variables that remain constant between successive installations, zip code and time control variables, and the stochastic term capturing unobserved heterogeneity. This stochastic term is by definition heteroskedastic, and may exhibit autocorrelation. Our primary explanatory variable of interest is the number of previous adoptions in a zip code, which is either log-transformed, or included with a quadratic term to allow for the possibility of a diminishing effect with more adoptions. Other relevant explanatory variables include indicator variables for the electric utility and indicator variables for the current incentive step in that IOU. To control for (pre-incentive) price levels, we also examine a specification that includes the price of the previous installation in the zip code.[11] Finally, we examine two closely related specifications in order to explore the demographic factors that influence adoption.

4.2 Identification

The hazard model developed above is ideally suited to examine whether more than just environmental preferences and demographics influence the rate of adoption of solar PV panels. In particular, it allows us to examine the effect of the stock of previously installed solar PV panels in a zip code on the rate of adoption of solar PV panels in that zip code. A key question is how to interpret this effect. Can we truly identify peer effects? Peer effects

[11] As a robustness check, we also estimate the model using a zip code price index of the average price of the last ten installations and find similar results.

are notoriously difficult to identify, and there is an extensive literature on the difficulties of such identification, starting with Manski (1993). There are three major issues in general with identifying peer effects, and while our particular specification does not suffer from all of these, there is a fourth issue that may be a concern. The three general issues with identifying peer effects are the "reflection" simultaneity problem, endogenous group formation, and correlated unobservables. The fourth potential issue is that our specification is closely related to a specification with a lagged dependent variable.

The first two potential issues with identifying peer effects do not appear to be a concern at all in our study. Our specification is based on the effect of *past* installations in the relevant area of study. A reflection issue could occur when one's peers simultaneously influence the person at the same time as the person influences his or her peers. In our study the covariate of interest is determined *before* the adoption decision – entirely mitigating the possibility of a reflection issue. Angrist and Lang (2004) and Ammermueller and Pischke (2009) use a similar argument for why reflection does not confound identification in their studies. Endogenous group formation is also not likely to be a problem in our context because we are examining where people live, and we find it highly unlikely that that the choice of where to live is related at all to the decision of whether to adopt a solar PV system. This is an important point, for endogenous group formation can be extremely important in impinging upon the identification of peer effects (e.g., see Christakis and Fowler (2007) and the follow-up critiques Cohen-Cole and Fletcher (2008a) and Cohen-Cole and Fletcher (2008b)).

Of course, correlated unobservables may well be an issue. Environmental preferences and demographics are the most obvious possibilities. For instance, zip codes that are more environmental or wealthier can certainly be expected to have a larger installed base and a greater rate of adoption. Similarly, if there are localized marketing campaigns focusing on a single zip code or neighborhood, that zip code may have both a larger installed base and greater rate of adoption. First, as described above, we include month indicator variables and zip code fixed effects in our primary specifications. These flexibly control for different preferences for solar across zip codes and any changes that occur across all zip codes over time. However, they do not control for changing heterogeneity of preferences for solar *within* a zip code over time that may be correlated with the installed base and rate of adoption. For example, if there was a changing distribution of environmental preferences within a zip code over time that happened to be correlated with the zip code installed base of solar PV panels, then the coefficient on the installed base may be picking up some of this effect. We recognize that these potential correlated unobservables may be an issue and thus provide serval robustness checks which will help control for these factors.

The fourth potential identification issue is that our dependent variable is the log of the time between adoptions in a zip code, while our key explanatory variable is the sum of the previous adoptions in that zip code. Thus, our key explanatory variable can be thought of as containing much of the information of a lagged dependent variable, even if it is not exactly a lagged dependent variable. If autocorrelation exists in our error terms, then there may be an endogeneity issue, for the error term would be correlated with the lagged error term. In this case, the lagged dependent variable would not be orthogonal to the error term, which would lead to biased and inconsistent estimates of our coefficients. Moreover, there is a well-known issue, sometimes called the Nickell-Hurwicz bias, in dynamic models with fixed effects (Hurwicz 1950; Nickell 1981). The intuition behind this issue is that in the fixed effects (demeaning) transformation, the lagged dependent variable will by construction not be orthogonal to the lagged term contained in the error mean. The Nickell/Hurwicz bias also leads to a biased and inconsistent estimate of the coefficients. Our cumulative adoptions variable is not exactly a lagged dependent variable, so both of these two issues may not be as important in our case.

To address the fourth potential identification concern, we perform an estimation where we instrument for our variable of interest. We use the cumulative number of new vehicles registered in the zip code after 2001 as the instrument. The cumulative number of vehicles registered in the zip code will capture previous shocks that influence durable goods purchase behavior that may affect both solar PV installations and new vehicle purchases. At the same time, we find it very reasonable that the cumulative new vehicle registrations satisfies the exclusion restriction.

4.3 Primary Results

In our primary set of estimation results, we estimate (7) using a log specification for the effect of the installed base of solar PV panels on time between adoptions. The results can be found in Table 4. For each of the regressions, we test for autocorrelation by regressing the residuals on the lags of the residuals. We find no evidence of statistically significant autocorrelation,[12] which implies that endogeneity from our key variable of interest being similar to a lagged dependent variable would only be an issue due to a Nickell-Hurwicz-type bias. For additional confidence, we report asymptotic standard errors which are robust to both heteroskedasticity and autocorrelation, and we more closely examine the possibility of endogeneity in our robustness checks.

In column one of Table 4, we control for zip code heterogeneity and find the first evidence of a large and highly statistically significant decrease in the time between adoptions when the cumulative adoptions of solar PV panels is larger. In column two, we include time control variables for each month which equal the fraction of time between adoptions spent in that month, i.e. the average value of month indicator variables, averaged over the time since the last adoption. We do not present results using standard month indicator variables, but we find extremely similar results.[13] The result in column 2 indicates that a 1% increase in the installed base would be associated with a roughly 1% decrease in the time between adoptions, even after controlling for changes in the solar market over time (e.g., changes in the cost of a solar installation, macroeconomic shocks) and differences across zip codes (e.g., differing wealth and environmental preferences). This result is highly statistically significant, and shows a considerable effect on the rate of adoption.

[12]The coefficient on the first lag is tiny and is not statistically significant from zero (p-value of 0.44).

[13]Including standard month indicator variables would assume that hazard rate remains constant over each month, which we view as an unlikely circumstance.

In column three, we allow the effect of installed base to differ across CSI administrative zones. We find a similar effect of the installed base on the rate of adoption in PG&E and SCE, but the effect appears to be slightly smaller in SDG&E. Column four includes the most recent installation price within the same zip code as the installation in order to control for the possibility of localized price shocks that might be correlated with the installed base in the zip code. The tiny and highly insignificant coefficient, along with nearly no change in our coefficient of interest, suggests that localized price shocks are not likely to be an issue confounding our identification. An identical result holds if we include a zip code price index made up of the average price of the past ten installations rather than the most recent price.[14] This does not mean that price does not affect the rate of adoption: the coefficients on the incentive step dummy variables become more positive as the incentives decrease indicating that the time between adoptions becomes longer.

In column five we instrument for our variable of interest to address the possibility of a bias similar to the Nickell-Hurwicz bias. As mentioned before, the instrument used here is the cumulative number of new vehicles registered in the zip code after 2001. The results roughly correspond to the primary OLS results, with a positive and highly statistically significant coefficient on the log installed base variable. The magnitude of the coefficient is smaller than before, at -0.7, perhaps indicating a slight bias in our OLS results. But the general magnitude, sign, and significance are not appreciably different, which in general corroborates our OLS specification results.

4.4 Robustness Checks

Now we present several pieces of evidence that shed light on the degree to which correlated unobservables may be influencing identification in our primary specification. The first piece of evidence is our quasi-experimental result suggesting that we are finding more than simply changing heterogeneity of environmental preferences. The second piece of evidence is our estimation results with zip code-quarter fixed effects (Table 5).

[14] We do not report these results, but they are available upon request.

The estimates using zip code-quarter fixed effects control for any number of potentially correlated unobservables that vary within a zip code over time. For instance, there could be a short-term localized marketing campaign using radio advertisements or billboards. Alternatively, there could be different demand shocks over time in each local market. Or each zip code may experience its own "S-shaped" diffusion curve, so that it is not only the adoption rate, but also the rate of change in the adoption rate, that is zip code dependent – and also somehow correlated with the cumulative number of adoptions in a way that varies over time. The results in Table 5 show a negative and statistically significant coefficient on the installed base of solar PV panels, which is just a bit larger than in the primary results. We interpret this evidence as consistent with correlated unobservables not dramatically confounding our results.[15]

Next we examine evidence for a *contractor-specific* effect. We run a similar specification as those in Table 4, only we use the log of the time between adoptions by the same contractor as our dependent variable and include variables for the the log of the cumulative installation performed by the same contractor as well as the log of the cumulative installations performed by all other contractors in the zip code. We include contractor-zip fixed effects to control for localized, contractor-specific demand and supply effects, as well as month indicator variables. We find a highly statistically significant coefficient of -1.23 for the log contractor cumulative installed base[16] and an equally statistically significant coefficient of 0.13 for the log competitors' installed base.[17] What is most notable is that the magnitude of the contractor-specific effect is about the same size as the entire effect in the previous specifications. We view this as evidence that the effects we are finding are highly

[15]One possible criticism of using zip code-quarter fixed effects is that they operate in a discrete fashion. Since identification in this case relies upon within-zip code-quarter variation in the time between adoptions, then if the hazard rate is varying locally over time in a smooth fashion due to correlated unobservables, we may still be attributing the effect of these correlated unobservables to peer effects. At the same time, quarter-zip code fixed effects are highly demanding of the data, so that we may be running the risk of overfitting. Thus, we follow Cohen-Cole and Fletcher (2008b) and perform a robustness check by including zip code indicator variables interacted with a time trend and quadratic time trend. These zip code-specific time polynomials flexibly control for unobservables which vary smoothly over time, including changing heterogeneity and localized marketing. We find nearly identical results.

[16]Standard error of 0.06

[17]Standard error of 0.04

contractor-specific, implying that changing environmental preferences are very unlikely to be an important confounding unobservable in our previous regression results. This result could imply an informational component to the effects we are finding.

We also perform a robustness check where we use the same specification as our primary specification in column two of Table 4, only with the log of the time between hybrid vehicle adoptions as the dependent variable. If we find a statistically significant negative coefficient on the cumulative adoptions of solar PV panels in the zip code, then we might be concerned that there is changing heterogeneity of environmental preferences driving our results. We find an extremely small and insignificant coefficient on the installed base of solar PV panels.Finally, we perform several additional robustness checks that include specifications using different transformations of the installed base variables of interest, such as a quadratic specification rather than a log specification. We again find results very similar to our primary specification results, suggesting that the model is reasonably specified. Taken together, these robustness checks indicate that there is an economically important and statistically significant effect of nearby previous adoptions on the rate of adoption of solar PV panels – consistent with peer effects.

4.5 Decomposing the Determinants of Adoption

To shed light on how these peer effects are changed by different demographics, we regress the log of the time between adoptions on the log zip code installed base and a variety of demographic variables interacted with the log of the zip code installed base. One of the more striking findings is that the fraction of people who drive hybrids in the zip code (averaged over the full time period) has a large and positive coefficient, which implies that it reduces the effect we are finding. This result suggests that more environmentally conscious consumers are less affected by their neighbors in the decision of whether to install a solar PV panel. Perhaps they are already better informed about the benefits of solar PV installations. Alternatively, they may simply be less susceptible to peer effects. The fraction of the population that is white also seems to reduce the effects, while the median home value appears to enhance the effects. All of these interactions are statistically significant to at least the 10% confidence level.

A second estimation looks at the direct effect of zip code demographics on the adoption of solar PV panels. We regress the zip code control variables in the previous regression (i.e., a variable made up of the coefficients on the zip code control variables from the regression in column two of Table 4) on the demographic variables. Here there is one observation per zip code. We use bootstrapped standard errors since the dependent variable is a fitted value from a previous estimation. We find that household size and the fraction of the population who drive hybrids both lead to a faster adoption rate, while the fraction of population who are male, fraction of the population who work at home or walk to work, and the number of owner occupied homes lead to a smaller adoption rate. These effects are meaningful in size and statistically significant. For example, the coefficient on the fraction of the population who own a hybrid is -3.24 which means that for a 0.01 increase in the fraction of people who drive a hybrid (the average fraction in a zip is 0.021), the time between adoption of solar in the zip decreases by 3.24%. This provides clear evidence that stronger environmental preferences are indeed associated with faster diffusion of solar PV panels.

5 Street-Level Analysis

The above analysis provides evidence for an effect of previous nearby adoptions on the rate of adoption of solar PV panels. However, we would also expect this effect to be at an even more localized level than the zip code, for a zip code contains approximately 5,000 households on average. With our address-level data from the CEC Emerging Renewables Programs, we can go even further, by examining how solar PV system adoption decisions are affected by the previous decisions of others *on the same street*. We define a street here as a street within a zip code, so that a long street that is in several zip codes is considered several separate streets.

For this analysis, we create a panel dataset from our 2001-2006 data, where each observation is a street-month. Our dependent variable of interest is an indicator variable for an installation occurring in that street-month. The key explanatory variables are an indicator variable for whether an installation has already taken place and the log of the installed base of solar PV systems in the zip code the street lies in.[18] Many streets are in the relatively early stages of adoption, so we have sufficient variation for our empirical analysis. Table 6 provides summary statistics for the constructed street-level dataset, showing the number of new installations and previous installations in a street-month, as well as the installed base and number of completed contracts within the zip code the street is located in.[19] Table 7 shows a cross tabulation of whether a new installation occurs on the street and whether there was a previous installation on the street. The first row indicates that in 3.9% of the street-months where there were no new installations, there had been a previous installation. The second row shows that in 15% of the street-months when there was a new installation, there had already been an installation.

As in the zip code analysis, we are not concerned about the reflection problem of simultaneity, since the decision to adopt is made *after* the decisions of the previous potential adopters. Similarly, endogenous group formation is not likely to be an issue, for it is difficult to imagine that decisions about whether to live on a certain street are related at all to solar PV installations. We address correlated unobservables in a flexible manner with a variety of fixed effects in different specifications to show the effect of the additional controls.

5.1 Results

In column one of Table 8, we examine only the effect of a previous installation on the likelihood of another installation on the same street. We include zip code fixed effects and time indicator variables in this specification. We find a statistically significant positive effect of a previous installation on the likelihood of another installation, providing

[18]As in the previous analysis, we also examine a quadratic specification in zip code installed base and are able to confirm the concave functional form.

[19]We drop streets with one or less adoptions, leaving 1,233,111 street-month observations.

suggestive evidence of highly localized peer effects at the street level. The result indicates that if there was a previous installation on the same street, the likelihood of an adoption increases by four percentage points.

In the second specification, we include the log of the zip code installed base to see how the decision to adopt is influenced by neighbors in a slightly larger community. In this specification, we also include the number of contracts completed in the zip code that month along with zip code fixed effects and time indicator variables. The effect of a previous installation is nearly the same as in the first specification. The effect of the zip code installed base is positive and significant but very small. The third specification controls for time-varying unobservables by including zip code-month fixed effects. Here identification is based on within zip code-year-month and across-street variation.[20] With these additional controls, the effect of previous installations dummy does not change, but installations elsewhere in the zip code now have a meaningful effect.

Finally, we are concerned that since each street has its own characteristics, it is possible that some of these unobserved characteristics are correlated with the installed base. For example, streets that are more residential are more likely to have more residential solar installations than more commercial streets. Our final, preferred specification controls for unobserved heterogeneity across streets by including an indicator variable for each street. We also include zip code-month fixed effects, just as before. Due to extremely large number of variables, we run the analysis on streets with at least three installations. Streets with fewer than three installations have little identifying power when we include street indicator variables, so we feel that this is a reasonable simplification.[21] When controlling for the heterogeneity in streets, the effect of a previous installation on the street and installations elsewhere have larger effects. If there has been a previous installation on the street, the probability of installations in future months increases by 16.5 percentage points. The coefficient on the log zip code installed base variable suggests that a 1% in-

[20] We are in effect comparing the likelihood of adoption in streets within a zip code in a particular year-month, conditional on a previous adoption and after controlling for the time-invariant, street-specific probability of adoption.

[21] We find that the results are fairly robust to keeping streets with at least three, four, or five installations.

crease in adoptions in a zip code increases the likelihood of adoption by 0.78 percentage points – consistent with our zip code-level results. For a zip code with the average number of installations of 24, one additional installation elsewhere in the zip code changes the probability of future installations by 3.25 percentage points, one fifth of the effect of an installation on the same street.

These results provide evidence suggestive of peer effects – or localized marketing to take advantage of peer effects – that decrease with distance and operate at both the street-level and zip code-level, a useful result for understanding the geographic nature of how decisions about adoption solar PV panels are made. These results also indicate that the effects that are leading to the clustering pattern of solar adoption appear to work at several geographic levels. This finding gives us additional confidence that in addition to environmental preferences, the previous decision of neighbors to adopt solar PV plays a role in decision-making about installing solar today.

6 Conclusions

In this paper we document a distinctive pattern of geographic clustering in the diffusion of solar technology in California. The geographic clustering appears to occur at both a zip code and neighborhood level, and does not simply match the density or the "greenness" of the zip code. Furthermore, industry reports and contractor marketing strategies all indicate that there may be more than simply environmental preferences or demographics underlying this pattern of clustering. Specifically, there may be peer effects, whereby previous choices of one's neighbors influence the decision to adopt a solar PV panel. We use a rich installation-level dataset to explore whether there is quantitative evidence to support this contention using three complementary methodologies.

The first uses quasi-experimental variation from the unique nature of the CSI subsidy schedule. The evidence from this analysis is very small scale, given the limited number of zip codes that are suited to the analysis, but it benefits from clean identification. The results provide initial support for a causal effect of nearby previous adoptions on the rate of adoption. Our second methodology uses a hazard model to examine evidence for the influence of previous adoptions in a zip code on the rate of adoption in that zip code. The analysis is broader based, but requires careful controls for a variety of plausible correlated unobservables. Yet the results are highly statistically significant and correspond closely with those from the quasi-experiment. This effect is reduced by the fraction of people in a zip code who drive hybrids, suggesting that stronger environmental preferences may lessen the influence of neighbors in the decision to adopt. On the other hand, the fraction of hybrids in a zip code is strongly associated with faster adoption of solar PV panels, underscoring the important role that environmental preferences play in the diffusion of solar PV technology. Our final methodology takes advantage of our detailed address-level installation data to explore the degree to which the effects found in the previous sections are highly localized. We find evidence both that previous adoptions on the same street and previous adoptions in the same zip code increase the likelihood of an adoption on a street. Again, the results are highly statistically and economically significant. These results are all consistent with consumers making adoption decisions based in part by what they see and hear from their neighbors.

Our results strongly indicate that previous adoptions in the same localized area play a role in the decision of a household to adopt solar. There are a variety of ways by which this could occur. One possible way is through image motivation, whereby households see their neighbors installing solar and choose to make an installation to be able to also project a green image. The work by Lessem and Vaughn (2009) in Sacramento suggests that this may indeed be an important factor. Another important way is through information-transfer in the process of social learning. Our results indicating that the effects are contractor-specific seem to support this mechanism. One could imagine neighbors discussing their experience with a particular contractor contributing to another person's decision to install. In addition, localized marketing could be designed to take advantage of peer effects. Anecdotal evidence, such as SolarCity's marketing strategy, suggests that for at least part of the market, this is an important factor. Our dataset does not allow us to empirically distinguish between such localized marketing leveraging on peer effects and peer effects alone. However, we might expect even localized marketing of this sort to be more short-term, and thus would be at least partly controlled for in our zip code-time fixed effects specifications. Our belief is that we are identifying a combination of all of the effects discussed here: social learning, image motivation, and localized marketing designed to take advantage of peer effects.

Our results provide insight into the nature of diffusion of emerging green products. They also point to leveraging such effects as a useful approach to marketers working to speed the adoption of a new technology. We view the results as an important first step in convincingly demonstrating peer effects in solar PV technology. The results are only suggestive of the mechanisms by which peer effects work, and thus understanding how these mechanisms work is a prime topic for future research – one likely to require a carefully developed experimental research design at the neighborhood scale.

References

Aghion, P. and Howitt, P. (1998). *Endogenous Growth Theory*, MIT Press.

Ammermueller, A. and Pischke, J.-S. (2009). Peer effects in european primary schools: Evidence from pirls, *Journal of Labor Economics* **27**: 315–348.

Angrist, J. and Lang, K. (2004). Does school integration generate peer effects? evidence from boston's metco program, *American Economic Review* **94(5)**: 1613–1634.

Arndt, J. (1967). Role of product-related conversations in the diffusion of a new product, *Journal of Marketing Research* **4**(3): 291–295.

Axsena, J., Mountain, D. C. and Jaccard, M. (2009). Combining stated and revealed choice research to simulate the neighbor effect: The case of hybrid-electric vehicles, *Resource and Energy Economics* **31**(3): 221–238.

Baker, L. C. (2001). Managed care and technology adoption in health care: evidence from magnetic resonance imaging, *Journal of Health Economics* **20**(3): 395–421.

Bandiera, O. and Rasul, I. (2006). Social networks and technology adoption in northern mozambique, *Economic Journal* **116(514)**: 869–902.

Bass, F. M. (1969). A new product growth model for consumer durables, *Management Science* **15**: 215–227.

Bollinger, B. and Gillingham, K. (2010). Learning-by-doing spillovers in the solar photovoltaic industry, *Stanford University Working Paper* .

Brock, W. and Durlaf, S. (2001). *Handbook of Econometrics, Vol 5*, Elsevier, chapter Interaction-based Models, pp. 3297–3380.

Christakis, N. and Fowler, J. (2007). The spread of obesity in a large social network over 32 years, *The New England Journal of Medicine* **357(4)**: 370–379.

Cohen-Cole, E. and Fletcher, J. (2008a). Detecting implausible social network effects in acne, height, and headaches: Longitudian analysis, *British Medical Journal* **337**: a2533.

Cohen-Cole, E. and Fletcher, J. (2008b). Is obesity contagious? social netwoks versus environmental factors in the obseity epidemic, *Journal of Health Economics* **27(5)**: 1382–1387.

Conley, T. and Udry, C. (2010). Learning about a new technology: Pineapple in ghana, *American Economic Review* **100(1)**: 35–69.

CPUC (2009). California solar initiative annual program assessment, *California Public Utilities Commission Go Solar California Report* .

Doraszelski, U. and Judd, K. L. (2008). Avoiding the curse of dimensionality in dynamic stochastic games, *working paper* .

Duflo, E. and Saez, E. (2003). The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment, *Quarterly Journal of Economics* **118**(3): 815–842.

Engers, M., Hartmann, M. and Stern, S. (2009). Are lemons really hot potatoes?, *International Journal of Industrial Organization* **27**(2): 250–263.

Foster, A. and Rosenzweig, M. (1995). Learning by doing and learning from others: Human capital and technical change in agriculture, *Journal of Political Economy* **103**(6): 1176–1209.

Frank, R., Massy, W. F. and Morrison, D. G. (1964). The determinants of innovative behavior with respect to a branded, frequently purchased food product, *in* L. Smith (ed.), *Proceedings of the American Marketing Association*, American Marketing Association, Chicago.

Genesove, D. (1999). The adoption of offset presses in the daily newspaper industry in the united states, *NBER Working Paper No. 7076* .

Goolsbee, A. and Klenow, P. (2002). Evidence on learning and network externalities in the diffusion of home computers, *Journal of Law and Economics* **45**: 317–344.

Gowrisankaran, G. and Stavins, J. (2004). Network externalities and technology adoption: Lessons from electronic payments, *The RAND Journal of Economics* **35**(2): 260–276.

Griliches, Z. (1957). Hybrid corn: An exploration in the economics of technological change, *Econometrica* **25**(4): 501–522.

Griskevicius, V., Tybur, J. M. and den Bergh, B. V. (2010). Going green to be seen: Status, reputation, and conspicuous conservation, *Journal of Personality and Social Psychology* **98(3)**: 392–404.

Hamilton, B. H. and McManus, B. (2005). Technology diffusion and market structure: Evidence from infertility treatment markets, *working paper* .

Hannan, T. H. and McDowell, J. M. (1984). Market concentration and the diffusion of new technology in the banking industry, *The Review of Economics and Statistics* **66**: 686–691.

Hartmann, W., Manchanda, P., Nair, H., Bothner, M., Dodds, P., Godes, D., Hosanagar, K. and Tucker, C. (2008). Modeling social interactions: Identification, empirical methods and policy implications, *Marketing Letters* **19**(3): 287–304.

Heckman, J. and Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data, *Econometrica* **52**(2): 271–320.

Hurwicz, L. (1950). *Statistical Inference in Dynamic Economic Models*, Wiley: New York, chapter Least Squares Bias in Time Series, pp. 365–383.

Kahn, M. and Vaughn, R. (2009). Green market geography: The spatial clustering of hybrid vehicles and leed registered buildings, *B.E. Journal of Economic Analysis and Policy* **9(2)**: 1–22.

Lee, R. H. and Waldman, D. M. (1985). The diffusion of innovations in hospitals: Some econometric considerations, *Journal of Health Economics* **4**(4): 373–380.

Lenzo, J. (2006). Market structure and profit complementarity:the case of spect and pet, *working paper* .

Lessem, N. and Vaughn, R. (2009). Image motivation in green consumption, *Manuscript, UCLA Economics Department* .

Lucas, R. (1988). On the mechanics of economic development, *Journal of Monetary Economics* **22(1)**: 3–42.

Manski, C. (1993). Identification of endogenous social effects: The reflection problem, *Review of Economic Studies* **60**: 531–542.

Moffitt, R. (2001). *Social Dynamics*, MIT Press, chapter Policy Interventions, Low-level Equilibria, and Social Interactions, pp. 45–82.

Moretti, E. (2008). Social learning and peer effects in consumption: Evidence from movie sales, *working paper* .

Mulligan, J. G. (2003). Market segmentation and the diffusion of quality-enhancing innovations: The case of downhill skiing, *Review of Economics and Statistics* **85**(3): 493–501.

Munshi, K. (2004). Social learning in a heterogeneous population: Technology diffusion in the indian green revolution, *Journal of Development Economics* **73(1)**: 185–213.

Newell, R. and Kerr, S. (2003). Policy-induced technology adoption: Evidence from the u.s. lead phase-down, *Journal of Industrial Economics* **51**(3): 317–343.

Nickell, S. (1981). Biases in dynamic models with fixed effects, *Econometrica* **49**: 1417–1426.

Oster, E. and Thorton, R. (2009). Determinants of technology adoption: Private values and peer effects in menstrual cup take-up, *NBER Working Paper 14828* .

Rogers, E. (1995). *Diffusion of Innovations*, The Free Press.

Romer, P. (1986). Increasing returns and long-run growth, *Journal of Political Economy* **94(5)**: 1002–1037.

Schmidt-Dengler, P. (2006). The timing of new technology adoption: The case of mri, *Working Paper* .

Shriver, S. (2010). Network effects in alternative fuel adoption: Empirical analysis of the market for ethanol, *Stanford University GSB Working Paper* .

Soetevent, A. (2006). Empirics of the identification of social interactions: An evaluation of the approaches and their results, *Journal of Economic Surveys* **20(2)**: 193–228.

Taylor, M. (2008). Beyond technology-push and demand-pull: Lessons from california's solar policy, *Energy Economics* **30**(6): 2829–2854.

Tucker, C. (2004). Two-sided network effects in the adoption of electronic payments, *Stanford Institute for Economic Policy Research Discussion Paper 03-11, Stanford, CA* .

Tucker, C. (2008). Identifying formal and informal influence in technology adoption with network externalities, *Management Science* **55**(12): 2024–2039.

Wiser, R., Barbose, G., Peterman, C. and Darghouth, N. (2009). Tracking the sun ii: The installed cost of photovoltaics in the us from 1998-2008, *LBNL Working Paper 2674E. Lawrence Berkeley National Laboratory* .

A Solar PV Installation Dataset

The dataset used in this analysis is installation-level data from both the CPUC CSI database (2007-2009) and the CEC Emerging Renewables Program (2001-2007). This appendix describes pertinent details of cleaning and preparing the data. First, the CEC data include the same variables as the CSI data with the exception of the installation type. To assign an installation type, we classify installations that are smaller than 10 kW as residential and larger than 10 kW as commercial.[22] For this study, we restrict our sample to only residential installations. Second, we note that 99.5% of the installations in the CEC data are in the three IOU regions, while all of the installations in the CSI data are in the three IOU regions. Thus for consistency with the CSI data, we only include installations in the IOU regions in our analysis. Third, we recognized several instances of community-based purchases of installations, where a small group would choose to purchase solar PV systems at the same time (e.g., SolarCity is known for this type of marketing). Often community-based groups negotiate a better price with the installer in exchange for the volume. There were only 249 instances of multiple installations in the same zip code at the same date in our dataset, so rather than separately model the bargaining process, we randomly selected one of the installations in each of the instances to include in our dataset. In this process, we dropped 833 installations (less than 2% of installations).

Anecdotal evidence from conversations with members of the solar industry suggests that social learning and peer effects are critical to making community-based purchases work. To the extent that this is true, our empirical estimates will underestimate the importance of peer effects. In our regressions, we also drop observations with miscoded prices or system sizes (roughly 2,500 dropped).[23] It should be noted that we do include these installations in the construction of all variables that depend only on the existence of

[22]The level of 10 kW has been used reports such as Wiser et al. (2009) as a suitable cutoff level and it is the 95th percentile of the residential installations in the CSI data. Since 10 kW is just below the 50th percentile for our CSI commercial installations, some commercial installations will inadvertently be coded as residential in the pre-CSI period.

[23]Specifically, we drop prices and sizes with missing values, prices less than $2 per W and greater than $15 per W, and system sizes greater than 50 kW.

the installation, such as the installed base and time between installations variables. It is the price and/or size variables which are unreliable, not whether the installation actually occurred, and so it would be incorrect to not include these installations in the construction of these variables. The cleaned installation dataset includes 43,570 residential installations between January 2001 and December 2009, 33,685 of which are completed installations.

B Single Agent Model

This appendix shows how the use of a hazard model of technology adoption can be justified by a discrete choice model of consumer choice with continuously arriving consumers. Consider a (residential) consumer choosing whether to hire a contractor to install a solar PV panel at time t. The consumer bases his or her decision on the installation price, the incentives in place, electricity prices, and the expectations of the future values of these variables. If there are localized demand effects, such as peer effects, the decision of whether to install will also depend on the number of solar PV panel adoptions in the relevant area around the consumer. The most relevant area for consumers is most likely to be quite localized, such as at the zip code or street level. The geographic unit of analysis in our primary analysis is the zip code level which is a small enough geographic unit so that we would expect to find evidence of localized peer effects through the effect of the zip code installed base variable. Since we primarily focus our analysis at the zip code level, we denote the geographic area of interest with subscript z. Allow the utility of an installation for consumer i in period t to be given as $u_{it} = u(X_{it}, \zeta_{it}; \theta_i)$. The consumer thus adopts a solar PV installation when $u_{it} > 0$. In most discrete choice models, time is discretized and the decision of a consumer to adopt in each period is often assumed to be independent of the decisions in adjacent periods, by assumimg iid errors. This is done for convenience in estimation as well as the fact that the data are often aggregated over some period of time, often at the monthly or quarterly level. Recently, there has been increased interest in modeling discrete choice with continuous time in dynamic discrete choice models, primarily due to the computational benefits associated with this formulationDoraszelski and Judd (2008). In these models, conditional choice probabilities are modeled using the familiar discrete choice random utility framework, but the choice is conditional on the agent making a choice at time t. The probability of the agent making a choice at time t is modeled through the use of a hazard model. This approach is especially suited to our application since we have continuous time data. Let $h(X_{it}, \xi_{it}; \gamma_i)$ be the hazard function that determines consumer i's decision times. As in Doraszelski and Judd (2008), we assume a non-homogenous Poisson process for consumer decision times, which is desirable for its memoryless property. From Bayes' rule, the probability of consumer i making an installation at any time is now also governed by a non-homogenous Poisson process with the following hazard rate:

$$\lambda_i(t) = 1\{u(X_{it}, \zeta_{it}; \theta_i) > 0\} h(X_{it}, \xi_{it}; \gamma_i) \tag{8}$$

The probability of an adoption in a geographic region z is determined by the hazard rate computed by integrating over the distribution of consumer heterogeneity within the area of interest. The rate at which new adoptions occur at time t in region z is given as:

$$\lambda_z(t) = \int 1\{u(X_{it}, \zeta_{it}; \theta_i) > 0\} h(X_{it}, \xi_{it}; \gamma_i) dF(\theta_i, \gamma_i). \tag{9}$$

For simplicity, assume consumers are homogenous in their utility except for an intercept term, and the ζ_{it} enter linearly and are distributed type 1 extreme value with scale factor normalized to one. Let the hazard rate determining consumers' decision time be an exponential function $h(X_{it}, \xi_{it}; \gamma_i) = \lambda(0) \exp(X_{zt}\gamma + \gamma_i^0 + \eta_{it} + \zeta_{it})$. The zip code hazard function of adoption becomes:

$$\lambda_z(t) = \exp\left(X_{zt}\gamma + \int \eta_{it} dF_z(\eta_{it})\right) \frac{\exp(X_{zt}\theta)}{1 + \exp(X_{zt}\theta)} N_z \int \lambda_i(0) \exp \gamma_i^0 dF_z(\lambda_i(0), \gamma_i^0), \tag{10}$$

where N_z is the number of consumers in the zip code. Since the probability of purchasing a solar installation is very low, it is reasonable to assume that the utility of installing is much lower than the outside alternative, i.e., $1 \gg \exp(X_{zt}\theta)$. We can then make the approximation that the denominator in the logit expression is equal to one. This gives us the following exponential expression for the hazard rate:

$$\lambda_z(t) = \lambda_z(0)\exp(X_{zt}\beta + \eta_{zt} + \epsilon_{zt}), \tag{11}$$

where X_{zt} are zip code and time-specific explanatory variables, $\lambda_z(0) = I_z \int \lambda_i(0)\exp\left(\gamma_i^0\right) dF_z(\lambda_i(0), \gamma$ $\beta = \gamma + \theta$, η_{zt} are zip code and time control variables, and $\epsilon_{zt} = \int \eta_{it} dF_z(\eta_{it})$ is a stochastic term capturing the unobserved heterogeneity in the hazard rate aggregated over individuals in the zip code, which may change on a daily basis. This may be due to zip code specific marketing, continually changing preferences of consumers in a zip code, or any other factor that is not picked up by the explanatory variables or the zip code and time control variables. The rate of adoption varies over time, since it is dependent on time-varying regressors as well as the stochastic shocks. The daily stochastic shocks are likely to be correlated over time. Although some fairly strong assumptions were made in the simplification of the expression, this section is merely intended to show how a hazard model of adoption can be justified from the primitives of a latent utility model of individual choice.

Table 1: Zip code-level summary statistics over 2001-2009

Variable	Mean	Std. Dev.	Min.	Max.	N
Zip code number of residential installations	24.13	37.14	1	404	1,496
Zip code MW of residential installations	0.11	0.186	0.001	1.816	1,496
population (millions)	0.02	0.02	0	0.11	1,219
household size	2.81	0.594	0	5.2	1,219
median income ($100,000)	0.64	0.29	0	3.75	1,219
pop male	0.502	0.031	0.34	0.97	1219
pop white	0.655	0.199	0.089	0.952	1,219
pop with college degrees	0.38	0.17	0.052	0.957	831
pop between 20 and 45	0.33	0.07	0.039	0.796	1,219
pop over 65	0.124	0.06	0	0.80	1,219
pop who drive to work	0.86	0.1	0.203	1	1,221
pop who carpool	0.145	0.06	0.005	0.559	1,202
pop using public transit	0.03	0.05	0.001	0.426	980
pop who work at home or walk to work	0.08	0.067	0.016	0.615	1,139
pop with over a 30 min commute	0.38	0.12	0.054	0.809	1,070
pop who drive a hybrid	0.04	0.06	0	1	1,301
number of owner occupied homes (1000s)	5.11	4.20	0	18.965	1,219
median value owner occupied home (millions $)	0.54	0.26	0	1	1,219
home loan (SPI)	123.16	68.54	0	576	1,219
home repair (SPI)	124.79	71.14	0	585	1,219
fraction of homes worth 0-50K	2.58	3.71	0	53.2	1,219
fraction of homes worth 50-90K	1.99	2.56	0	37.3	1,219
fraction of homes worth 90-175K	5.65	7.33	0	50	1,219
fraction of homes worth 175-400K	29.95	23.18	0	89.7	1,219
fraction of homes worth 400K+	59.56	29.61	0	100	1,219

Table 2: Residential installation size and price

Variable	Mean	Std. Dev.	Min.	Max.
size (kW)	4.873	2.752	0.119	48.3
price ($/W)	8.395	1.451	3.018	14.997
N		36,112		

Table 3: Preliminary regressions for transitions from incentive step four to five, where transition order is PG&E, CSE, SDG&E.

	PG&E and SCE border	**SCE and SDG&E border**
second region	-0.925	2.044
	(0.707)	(0.616)
after first region changes step	-0.009	-0.027
	(0.592)	(0.645)
after both regions change step	0.000	1.126
	(0.000)	(0.641)
second region x after first region changes step	-2.475	-1.061
	(0.605)	(0.705)
second region x after both regions change step	-0.968	-1.817
	(0.541)	(0.589)
Zip Code Controls (η_z)	Y	Y
Month Controls	N	Y
R-squared	0.603	0.376
N	72	213

Table 4: Log time regressions with zip code FE

	(OLS 1)	**(OLS 2)**	**(OLS 3)**	**(OLS 4)**	**(IV 1)**
log zip installed	-0.509	-0.955	-0.953	-0.948	-0.656
	(0.019)	(0.036)	(0.037)	(0.037)	(0.031)
SCE x log zip code installed base			-0.013	-0.012	
			(0.031)	(0.031)	
SDG&E x log zip code installed base			-0.015	-0.020	
			(0.038)	(0.038)	
previous installation price				-0.001	
				(0.004)	
Zip FE (η_z)	Y	Y	Y	Y	Y
Month Controls	N	Y	Y	Y	Y
R-squared	0.157	0.159	0.159	0.160	0.146
N	31,613	29,172	29,172	28,404	29,083

Table 5: Log time regressions with zip code-quarter FE

	(1)	(2)	(3)
log zip installed base	-1.409	-1.366	-1.360
	(0.174)	(0.180)	(0.182)
SCE x log zip code installed base		-0.171	-0.162
		(0.176)	(0.189)
SDG&E x log zip code installed base		-0.109	-0.104
		(0.210)	(0.218)
previous installation price			-0.000
			(0.007)
Zip-quarter FE	Y	Y	Y
R-squared	0.072	0.075	0.076
N	18,936	18,936	18,602

Table 6: Summary statistics at the street-month level

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**
new installation	0.015	0.121	0	1
previous installation	0.04	0.196	0	1
zip installed base (100s)	0.189	0.321	0	3.24
zip contracts (100s)	0.388	0.345	0.01	1.83
N		1,400,117		

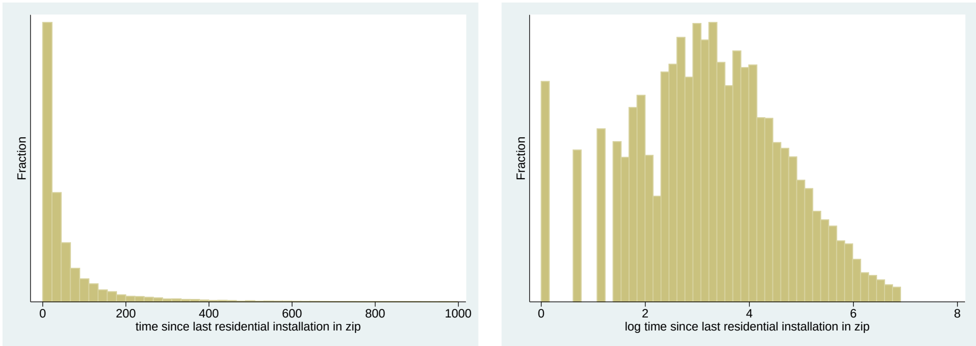
Table 7: Tabulations of new installations on the same street

	previous installations		
	no	yes	Total
no new installation	1,326,465	52,990	1,379,455
new installation	17,644	3,018	20,662
Total	1,344,109	56,008	1,400,117

Table 8: Street-Level Results

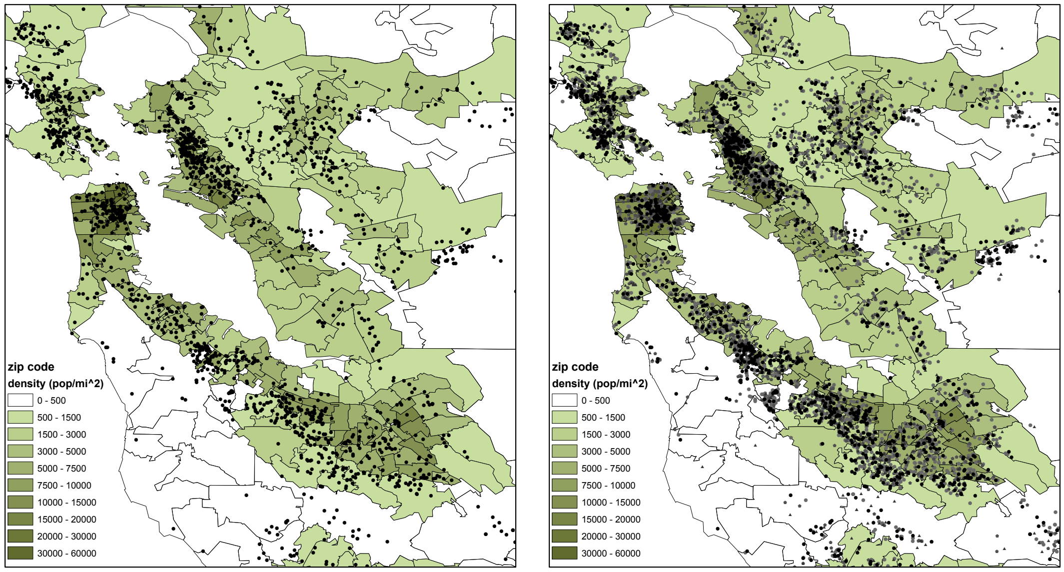
	(1)	**(2)**	**(3)**	**(4)**
previous installation	0.037	0.036	0.037	0.165
	(0.001)	(0.001)	(0.001)	(0.036)
log zip code installed base		0.022	0.329	0.775
		(0.001)	(0.009)	(0.143)
constant	0.000	0.416	0.752	1.156
	(0.000)	(0.004)	(0.021)	(0.219)
Zip FE (η_z)	Y	Y	N	N
Month Controls	Y	Y	N	N
Zip-Month FE	N	N	Y	Y
Street Controls	N	N	N	Y
R-squared	0.011	0.012	0.026	0.740
N	1,400,117	1,121,324	1,121,324	3,688

Dependent variable: Adoption occurs on that street
Heteroskedasticity and cluster-robust standard errors in parentheses



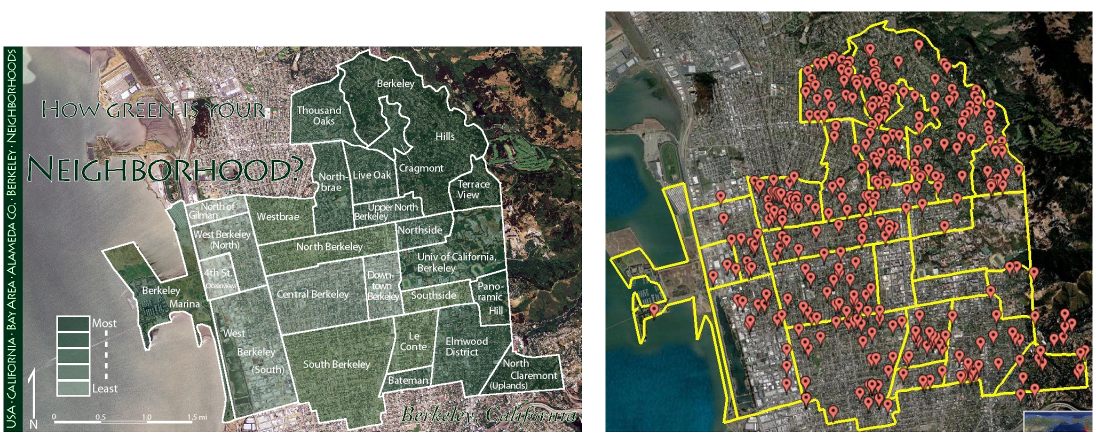
(a) Time (b) Log time

Figure 1: Time and Log Time Between Zip Code Installations



(a) 2001-2003 (b) 2004-2006

Figure 2: Clustering in Solar PV installations in the San Francisco Bay Area



(a) Berkeley neighborhoods (b) Berkeley installations

Figure 3: Google Earth view of Berkeley Installations by Neighborhood 2001-2006

Figure 4: The quasi-experiment exploits changing incentives within the same zip code