# Peer Effects in the Diffusion of Solar Photovoltaic Panels*

Bryan Bollinger

NYU Stern School of Business†

Kenneth Gillingham

Yale School of Forestry & Environmental Studies‡

November 16, 2011

Abstract

Social interaction (peer) effects are recognized as a potentially important factor in the diffusion of new products. In the case of environmentally friendly goods or technologies, both marketers and policy makers are interested in the presence of causal peer effects due to the social spillovers that can expedite adoption. We study the diffusion of solar photovoltaic (PV) panels in California, and find evidence of a causal peer effect in which a one percent increase in the zip code installed base leads to just over a one percent increase in the adoption rate. Our approach addresses the key difficulties in identifying peer effects. We provide a methodology for the simple, straightforward identification of peer effects with sufficiently rich data, avoiding the biases that occur with traditional fixed-effects estimation when using the past installed base of consumers in the reference group.

*The authors would like to thank Hunt Allcott, Tim Bresnahan, Jesse Cunha, Misha Dworsky, Wesley Hartmann, Sridhar Narayanan, Harikesh Nair, Peter Reiss, Arthur van Benthem, Ali Yurukoglu, and 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, and the Energy Institute at Haas seminar series 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.

†40 West 4th St, New York, NY 10012, bbolling@stern.nyu.edu.

‡195 Prospect Street, New Haven, CT 06511, kenneth.gillingham@yale.edu.

# 1 Introduction

Factors affecting the adoption of new products have long been of critical interest to marketers. The idea that social interactions influence technology diffusion and growth has a long history in the academic literature, and social interactions have been studied in the fields of marketing, industrial organization, development economics, and sociology. Knowledge spillovers between agents 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). Classic aggregate diffusion marketing models such as the canonical Bass model (1969) and its variants (Norton and Bass 1987; Danaher et al. 2001; Van den Bulte and Joshi 2007; Mahajan et al. 1990) allow for social contagion and indeed rely upon it to explain the S-shaped diffusion curve commonly observed in the diffusion of many types of products. Recently, social interactions have received increased attention at the micro level, in part due to the increased availability of disaggregate data allowing for better identification of the relevant peer group either through self-elicitation (Conley and Udry 2010; Kratzer and Lettl 2009; Iyengar et al. 2011; Nair et al. 2010), social, demographic or cultural proximity (Bertrand et al. 2000; Sacerdote 2001; Duflo and Saez 2003; Sorensen 2006; Munshi and Myaux 2006) or geographic proximity (Topa 2001; Arzaghi and Henderson 2007; Bell and Song 2007; Manchanda et al. 2008; Choi and Bell 2010; McShane et al. 2010; Nam et al. 2010; Narayanan and Nair 2011).

A key implication of social interactions is the potential for social spillovers, in which a marketing action that affects one agent can also indirectly influence other agents through the social spillover. Marketers attempt to account for these spillovers when determining the intensity of marketing activity. Asymmetric interactions or other sources of heterogeneity in the level of the social interaction can directly affect the optimal targeting of marketing efforts. However, this presence of a spillover effect depends on a *causal* social interaction effect. In the classic aggregate diffusion models, the estimated coefficient of contagion may capture a variety of other factors along with causal social interaction effects.

This paper makes several contributions to the literature on social interaction effects. Substantively, we document and measure the size of peer effects in the diffusion of an environmentally beneficial technology, solar photovoltaic (PV) panels. The diffusion of solar PV technology is of great interest to policy-makers since increased adoption leads to reduced greenhouse gas emissions. Methodologically, we provide an empirical framework for the quick and straight-forward estimation of peer effects using daily adoption data which leverages the difference in the dates an installation is requested and completed with a first-differences estimation approach in order to i) maximize the efficiency of our estimates while avoiding the restrictive functional and distributional assumptions on the data generating process and unobservables often seen in the literature, ii) control for endogenous group formation and correlated unobservables with a rich set of fixed effects, iii) avoid aggregation bias, and iv) avoid the biases present in traditional fixed effects estimation with endogenous or predetermined regressors.

The three well-known issues that often confound identification of peer effects are endogenous group formation leading to self-selection of peers (homophily), correlated unobservables, and a particular type of simultaneity called "reflection" in which agents' behaviors affect each other (Manski 1993; Brock and Durlaf 2001; Moffitt 2001; Soetevent 2006). Hartmann et al. (2008) discusses some of the different modeling and estimation strategies to avoid these issues. The third issue, reflection, is eliminated by focusing on the effect of the installed base on future adoption where the installed base is defined as the set of agents in the relevant peer group who have *previously* adopted the product, just as in the classical aggregate diffusion models. The other main issues, homophily and correlated unobservables, can be controlled for with sufficient data using a rich set of random or fixed effects. In the case of random effects, misspecification of the distribution can lead to severely biased estimates, and the use of these estimators rely on the questionable assumption that the random effects are independent of the included regressors. Often fixed effects are preferred to random effects since they do not rely on these distributional or independence assumptions. However, as discussed in Narayanan and Nair (2011), the traditional fixed effects estimator leads to inconsistent estimates when the regressors are either correlated with or direct functions of the lagged dependent variable, as is the case for the installed base (Nickell 1981).

An undesirable tradeoff results. With the inclusion of random effects only, controls for homophily and correlated unobservables are limited and the inference of causality for the effect of the installed base is questionable. With a rich set of fixed effects, which are especially necessary in the presence of peer group-specific, time-varying unobservables, such as localized marketing campaigns, the estimated effect of the installed base is biased. Most of the aforementioned papers in marketing ignore this issue; Narayanan and Nair (2011) address it in two ways, first using an instrumental variables approach and second by estimating the size of the bias. They find that without the bias correction, even the sign of the estimated coefficient of interest can be reversed. Unfortunately, the second method fails to account for endogeneity that would arise if there is autocorrelation in the unobservables. In addition, the instrumental variables approach is often not possible to implement due to the difficulty of finding appropriate instruments that satisfy the exclusion restriction (instruments are especially difficult to find for installed base since it is a stock variable rather than a flow variable). Our approach leads to consistent estimates of the peer effect even in the presence of correlated unobservables. We leverage the fact that in our application, the decision to install solar panels does not lead to an instantaneous installation, due to the time needed to then perform the installation. This means that a demand shock yesterday that leads more people to request a solar PV installation would not be immediately reflected in the installed base. So by using a first-differences estimator with observations at the daily level, we can avoid a correlation between the first-differenced installed base and the first-differenced unobservables.

The rest of the paper is organized as follows. In section 2 we provide background on the industry, describe the data, and document the pattern of geographic clustering of solar PV panels in California. In Section 3, we develop our zip-code level adoption model and describe the estimation procedure as it compares to other methods. We also present our main results regarding the effect of the zip code installed base on the probability of adoption, as well as on installation size, and we further decompose the determinants of the peer effects to see how demographic variables may interact with peer effects to influence adoption rates. In Section 4, we present an alternative street-level model to estimate the effect of previous adoptions on the probability of adoption at a finer geographic level, using a subset of our dataset that includes detailed address-level data. We discuss the feasibility of identifying peer effects and provide further, quasi-experimental evidence for peer effects in our application in Section 5. We also compare our results to those obtained using alternative methods, and discuss the implications of peer effects. Section 6 concludes.

## 2 Background and Data

While solar PV technology has 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 terms of government support and consumer adoption. 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] Finally, in January 2006, the

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

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] These substantial subsidies have contributed to the dramatic growth in annual solar PV adoptions in California over the past decade, from less than 1,000 residential installations in 2001 to over 17,000 in 2010.

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 January 2001 to August 2011.[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 customer, IOU, size of the installation and incentive, PV installer and manufacturer, the date when the customer requested (and 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 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

[2]In addition, 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 removed in 2008. Germany, Japan and Spain are a few other countries that have taken major policy actions to encourage the diffusion of solar PV technology for environmental and national energy security reasons. For an overview of the history of solar PV policy in California, see Taylor (2008) and for further information regarding the CSI incentives, see **?**

[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.

from R.L. Polk and Company. The cleaned dataset includes 79,101 requested residential installations between January 2001 and August 2011. Table 1 contains zip code level summary statistics for residential installations and key demographics.

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. In a phone survey of 639 participants and 601 non-participants in the CSI program, 16% of participants and 13% of nonparticipants had heard of the program through word of mouth.[4] Peer effects that operate through information exchange have been shown to exist in the domain of agricultural technologies and practices (Foster and Rosenzweig 1995; Munshi 2004; Bandiera and Rasul 2006; Conley and Udry 2010). Image motivation is another possible mechanism for peer effects, whereby households receive utility from the conspicuous consumption of an environmentally-friendly good (Griskevicius et al. 2010). Lessem and Vaughn (2009) find that the political ideology of a neighborhood affects the adoption of solar PV installations in Sacramento, providing suggestive evidence of possible peer effects due to image motivation in solar PV adoption.[5] If image motivation is important, peer effects might be larger for environmentally-friendly goods than other goods, with ramifications for marketers and policymakers who may be able to exploit them to expedite and increase diffusion.

Two patterns should emerge in the data if there are indeed peer effects in the diffusion

[4]Over one third of the nonparticipants surveyed had not heard of the CSI initiative; 37% of participants heard about the program through their contractor compared to 4% for non-participants.

[5]Social interactions have also been studied in the diffusion of hybrid vehicles (Axsena et al. 2009; Narayanan and Nair 2011) and the adoption of technologies to phase out lead (Newell and Kerr 2003).

of residential PV panels. First, we should see a clustering of installations because peer effects increase the probability of nearby adoptions. We explore patterns of adoption over time, both at the regional and neighborhood level. Figure 1(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, yet there are densely populated zip codes with few installations, and less densely populated ones with many installations. This could be indicative of peer effects – or spatially correlated preferences. For example, greater concern for the environment in some areas than others may lead to clustering in the adoption of any fledgling green technology, such as solar PV panels, hybrid vehicles, or LEED buildings (Kahn and Vaughn 2009). In the aforementioned survey, 52% of participating consumers reported financial reasons and the primary reason for their installation, 26% cited a concern for the environment, and 11% to save energy.

In addition to the clustering of installations, the presence of peer effects would imply *accelerating* adoption in regions with more installations. In the maps, the pattern of clustering appears to build upon itself. Figure 1(b) shows the same map of the San Francisco Bay Area with installations from 2004 to 2006 also included. While there are more installations everywhere by 2006, the level of clustering very clearly increases. The acceleration of installations can be seen in the empirical hazard rates. Figure 2 shows the median empirical hazard rates for different zip codes quantiles, which are grouped based on the total number of installations at the end of our data panel. This is a useful way of classifying the quantiles because it allows us to see whether zip codes with higher initial adoption rates continue to have an accelerating rate of adoption through the end of our data. The graphs clearly show that this is indeed what happens, providing suggestive evidence consistent with peer effects.

Some marketing efforts by firms aim to leverage social interaction effects. For example, one of the strategies employed by SolarCity (the largest installer in California) involves finding one or two vocal solar advocates in a neighborhood and giving the entire neighborhood a slightly lower price if enough adoptions are made within that neighborhood. Some firms try to increase the visibility of the installations by putting up a sign indicating that a solar PV panel has been installed at that location. The PG&E CSI administrators also note the value of peer effects by establishing "Solar Champion" training sessions for "citizens interested in helping spread the word about solar in their neighborhoods."

Previous studies attempting to identify social interactions have relied heavily on detailed marketing data or data on the peer network itself. Manchanda et al. (2008) and Nair et al. (2010) make use of pharmaceutical detailing data, and Nair et al. (2010) and Iyengar et al. (2011) use self reported measures of social interaction. One drawback of relying exclusively on specific marketing data without allowing for other localized, time-varying factors is that other correlated unobserved variables may also be important in the diffusion process. Time-varying correlated unobservables (which often exhibit auto-correlation) are be particulary important when maximum likelihood estimation with the assumption of i.i.d. errors is used, as in Manchanda et al. (2008) and Iyengar et al. (2011).

In a different approach, Bell and Song (2007) have zip code-level data on the adoption of online grocery retailing, and cleverly use order statistics to derive an expression for the probability that at least one individual within a region makes use of the new technology. This method also relies heavily on the assumption of i.i.d. errors across individuals within the same market as well as the number of potential adopters within the market. These assumptions are necessary in Bell and Song's setting because by looking only at the first adoption within each market, the authors are limited in the number of control variables that can be used to capture unobserved effects.

Our dataset does not include detailed localized marketing data for the large number of firms in the market. Yet in our empirical setting, we recognize that preferences for environmentally-friendly products could be changing over time and at different rates in different areas. Moreover, we feel it is important in our empirical setting to impose as few assumptions as possible, while still controlling for time-varying and area-specific unobserved factors. These considerations motivate our empirical approach.

# 3 Zip Code Analysis

## 3.1 Model and Estimation

In this section, we model 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. Specifically, we assume that the probability that a household in zip code $z$ adopts solar in any given month can be modeled as:

$$Pr_{zt} = \alpha_z \log(b_{zt}) + X_{zt}\beta + \eta_t + \xi_{zt} + \epsilon_{zt}, \tag{1}$$

where $b_{zt}$ is the zip code installed base of solar panels, $X_{zt}$ includes potentially time-varying explanatory variables such indicator variables for the different level of incentives, $\eta_t$ include time indicator variables for each month, the day of the month, and the day of the week, and $\xi_{zt}$ are zip code-quarter fixed effects to control for changing environmental preferences and localized marketing. We define the probability of adopting, $Y_{zt}$, as the fraction of households in zip code $z$ who adopt solar at time $t$.

There are two advantages to modeling the effect of the installed base using its logarithm. First, it assumes a declining effect of additional installations on the probability that households in the zip code adopt solar, which is what we would expect. Second, it allows the model to be agnostic regarding the scaling of the installed base, i.e., whether the regressor is the number of installations per zip code, per acre, per person, or per household, since any zip code-specific scaling variable within the logarithm is absorbed in the zip code fixed effect.

The installed base variable includes all installations that have been completed by time $t$. Clearly, there is an endogeneity issue using a traditional fixed effects estimation:

$$(Y_{zt} - \overline{Y}_z) = \alpha_z(b_{zt} - \bar{b}_z) + (X_{zt} - \overline{X}_z)\beta + (\eta_t - \overline{\eta}) + (\epsilon_{zt} - \overline{\epsilon}_z). \tag{2}$$

The mean-differenced error term is correlated with the mean-differenced installed base since large epsilon shocks in previous periods will lead to more installation requests, which will lead to a larger current installed base. However, installations typically take months to perform from the time the installation is requested and the incentives are locked in to the time it is completed. The median time to completion is 161 days, and less than half of a percent of installations are completed in a month or less. Therefore, the first-differenced regression,

$$(Y_{zt} - Y_{zt-1}) = \alpha_z(b_{zt} - b_{zt-1}) + (X_{zt} - X_{zt-1})\beta + (\eta_t - \eta_{t-1}) + (\epsilon_{zt} - \epsilon_{zt-1}), \qquad (3)$$

does not suffer from the same endogeneity issue (even when the data are aggregated at the monthly level). The term $b_{zt} - b_{zt-1}$ only includes installations that have been completed in the past day, and $Y_{zt} - Y_{zt-1}$ only includes installations that have been requested in the past day. Thus, $b_{zt} - b_{zt-1}$ is no longer a pre-determined regressor since it takes much longer than a day for a requested installation to be included in the installed base. The difference in the error term $\epsilon_{zt} - \epsilon_{zt-1}$ will not be correlated with $b_{zt} - b_{zt-1}$ so long as the duration of any autocorrelation in the errors is less than the time it takes for requested installations to enter the installed base. As is standard with first-differences, the first observation for each group is not used in estimation.

By using the daily probabilities of adoption, we avoid the time aggregation bias present when there is an underlying continuous-time data generating process as discussed in Schmittlein and Mahajan (1982), Petersen (1991), Srinivasan and Mason (1986) and Ter

Hofstede and Wedel (1998). When one of the regressors is monotonic, such as is the case with the installed base, this bias can be exacerbated. For example, if we look at the simple relationship between installed base and the probability of adoption in (1) and ignore zip code fixed effects and other covariates, the estimated effect of the installed base is $\widehat{\alpha}_z = \sum_z \sum_t b_{zt}^{-1} Y_{zt}$. If we aggregate this to the monthly level, we have $\widehat{\alpha}_z = \sum_z \sum_m \overline{b}_{zm}^{-1} \overline{Y}_{zm}$ where $\overline{Y}_{zm}$ and $\overline{b}_{zm}$ are the average values of these variables in month $m$. However, in actuality, with monthly aggregated data what is done instead is to use the installed base at the start of the month, $b_{zm0}$, since the timing of the installations within the month are not known. The bias present from this approximation is equal to $\sum_z \sum_m \left( \frac{1}{b_{zm0}} - \frac{1}{\overline{b}_{zm}} \right) y_{zm}$, which is positive since the installed base at the start of the month is always smaller than the average installed base within the month.[6] This would lead to an overestimation of the size of the peer effect, since the effect of installations occurring within the month on the adoption rate are being attributed to the smaller number of installations that had occurred by the start of the month.

Our model assumes that the peer effect resulting from an installation operates once an installation has been completed. The actual assumption needed is not this strong; since we perform the analysis at the daily level, the necessary assumption is that new, requested installations do not affect other consumers' decisions to request solar on the following day. This is a benign assumption even if the peer effect operates during the actual installation since no installations will begin to be installed until the relevant paper-

[6]The direction of the bias is unclear when using first-differences since the expression for the bias is the same but with the first-differenced variables substituted into the expressions.

work is filed, which takes on the order of months.

In summary, our approach relies three key elements: (1) data at the daily-level, (2) a lag between the time an installation is requested and completed (that is longer than the order of autocorrelation), and (3) first-differences estimation. This approach, avoids the common issue in fixed effect estimation with a pre-determined regressor, while at the same time avoids a possibly important aggregation bias. The approach does require the order of autocorrelation to be less than the lag between a request and completed installation.

## 3.2 Results

**Adoption rate** Table 3 presents the primary zip code-level estimation results. In column one, as a baseline, we estimate the model including month indicator variables without the inclusion of geographic fixed effects. We find an insignificant effect of the installed base. Including zip code fixed effects (within estimator) to control for geographically clustered preferences also show an insignificant effect. Column three includes zip-quarter fixed effects (within estimator). The result shows a significant, negative coefficient for the log installed base. However, as Narayanan and Nair (2011) point out, traditional fixed effects estimation will be biased downward. In their application, Narayanan and Nair (2011) also find a negative sign of the effect with traditional fixed effects estimation.

Column four presents our preferred specification using the first-difference estimator. The R-squared is much lower since we are only explaining differences in adoption probabilities rather than the actual probabilities, and there are slightly fewer observations because it is necessary to drop the first day in each zip-quarter. We find a positive, significant effect of the installed base on households' probability of adopting solar. To give a better sense of the magnitude of the effect, the coefficient estimate of 5.12 implies that a one percent increase in the installed base leads to a 1.15 percent increase in the daily adoption rate on average.[7]

Table 4 shows analogous results for when we aggregate the data at the monthly level (the dependent variable is still scaled in terms of a daily adoption rate for easy comparison). The differences in the results highlight how important the actual timing of the installations can be when determining whether or not a peer effect exists. Although there is no bias present from the presence of pre-determined regressors in the first-differenced estimates (column four), an aggregation bias exists because the presence of additional installations as they enter the installed base are ignored until the end of the month. The result is a coefficient with the incorrect sign, underscoring the aggregation bias.

Identification of the installed base parameter in our first-differenced estimates hinges upon the order of autocorrelation being less than the time it takes for peer effects to begin after an installation is requested. In the traditional fixed effects regression, we find statistically significant and positive (albeit small) autocorrelation for the first two lags (days), and for the first-differences regression, we find a significant, negative correlation of -0.5 for the first lag, a mechanical result of the first-differencing. We view this as comforting, since we do not expect that peer effect would begin working within days after the solar

[7] We use the elasticity of adoption $\alpha/Y_{zt}$ and take the average over all days and zip codes in the 4.43 million dataset.

PV contract was signed – long before the installation was actually put in place.

To get a better sense of which zip code characteristics lead to increases in levels of solar adoption, we run a traditional OLS regression including zip code demographic variables and their interaction with the installed base.[8] We find that zip codes with a larger population and a higher percentage of people who are male, white, college educated, have over a half hour commute, own a hybrid vehicle, and have home repairs have a higher household adoption rate of solar panels. Curiously, zip codes with a higher median income, higher proportion of people aged 20-45 and more than 65, as well as zip codes with higher valued homes exhibit lower adoption rates conditional on the other covariates.

To get a sense of how the peer effect may vary based on demographics, we run a first-differences regression including the demographic variables interacted with the installed base. We find that zip codes with higher median household sizes and fraction of people with more than a thirty minute commute have a larger peer effect, while zip codes with more people who carpool have a smaller peer effect. The significance of these three variables suggests that the mere visibility of installations may contribute to the peer effect, for larger households have more eyes per household to see other adoptions of solar, and longer commutes may imply more driving time to see other installations. Carpooling would have the opposite effect by reducing driving time. If visibility enhances the peer effect, the marketing implication is that solar installers should make an effort to increase the visibility of their installations. Indeed, this strategy can be seen with several installers putting up signs indicating that a solar PV panel has been installed.

[8]The full results are available from the authors upon request.

If the peer effect operates largely through word-of-mouth rather than through visibility, one might expect the peer effect to be contractor-specific. To test this, we run a similar set of specifications as those shown in Table 3, where the dependent variable is the probability of an installation by a specific contractor, and we include the total zip code log installed base as well as the contractor-specific log installed base as the key explanatory variables. We do this for the five contractors with the most total installations (over 1,000 residential installations each). We find no evidence that the peer effect is contractor-specific using the first-differenced regression results. This suggests that if the peer effect has an informational component, it does not in itself lead to more installations by the same contractor.[9]

**Installation size** The peer effect may work in part by reducing uncertainty in the utility a consumer will get from an installation. If this is the case and consumers are risk averse, then we would expect see the average installation size increase as more neighbors have adopted solar PV panels. Furthermore, according to the 2009 CSI Impact Report, environmental concerns were more likely to be the main factor in adoption for consumers with installations smaller than the median size, as shown in Figure 3. This suggests that the reduction of uncertainty (and thus the importance of discussions with peers) would be of even greater importance for larger installations because concern for the environment is not as important for consumers with larger installations. The average sizes and prices of

[9]When using traditional fixed-effects regression, it does appear that a contractor-specific result is present when using zip code fixed effects, further underscoring the importance of taking the bias seriously.

all residential installations are shown in Table 2.[10]

The intuition behind this prediction can be seen in the model that follows. Let $f$ be a concave function to allow for risk aversion and assume the net value of the electricity generated on day $t$ for household $h$ is equal to the price, $p_{ht}$, minus the daily cost of the installation, $c_{ht}$, with uncertainty $\sigma_{ht}\epsilon_{ht}$. Allowing for some additional potential benefit (or cost) from the installation $g_h$, possibly due to concern for the environment, the overall utility of an installation for household $h$ is given by:

$$u_{hzt} = f\left((p_{ht} - c_{ht} + \sigma_{ht}\epsilon_{ht})q_h\right) + g_h, \tag{4}$$

where $q_h$ is the size of the installation. The financial component is more important for larger installations. If the effect of the installed base decreases the uncertainty by lowering $\sigma_{ht}$, then the size of the break-even installation (across households) increases with the installed base, due to the concavity of $f$. To test this prediction, we estimate a linear model of installation size, first using OLS regression, then including zip code fixed effects, and finally with zip-quarter fixed effects. The installed base is no longer a pre-determined variable, so we can use traditional fixed effects estimation without worrying about the bias. The results are shown in Table 5. We find a positive, significant coefficient on the zip code installed base, which is what we would expect if the peer effect reduces the risk of installations through the provision of information, which then makes larger installations

[10]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 5.24 kW, with an average pre-incentive price of \$8.49 per W. This corresponds to an average system price in the range of \$40,000 before incentives.

appear more attractive.

## 4 Street-Level Analysis

While the above analysis provides evidence for an effect of previous adoptions in the zip code on the rate of adoption of solar PV panels, we might also expect the same effect at a more localized level. With address-level data from the CEC Emerging Renewables Program, we can examine 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. Due to the vast number of streets, it is not possible to perform the analysis at the street-day level, but since installations on a particular street are infrequent, there should be minimal aggregation bias. The key explanatory variables are an indicator variable for whether an installation has already taken place on the street and the log of the zip code installed base. 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 street's zip code.[11]

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

As before, we model the probability of a household adopting solar panels. However, unlike the analysis at the zip code level, we do not know the set of potential adopters on each street since we do not have information regarding the number of households on each street. However, we can overcome this issue. Let the probability of an installation on street $s$ in zip code $z$ at time $t$ be given by:

$$Pr_{szt} = \frac{1}{M_s}\gamma \log(b_{st}) + \alpha_z \log(b_{zt}) + X_{zt}\beta + \eta_t + \xi_{zt} + \epsilon_{szt}, \tag{5}$$

where $b_{st}$ is the installed base of installations on the same street and $M_s$ is the unknown number of potential adopters. In this specification, we assume in this formulation that the effect of a previous installations on the street is smaller for streets with larger number of potential adopters, due to the increased length of the street. If we multiply the equation on all sides by $M_s$, we get:

$$Y_{szt} = \gamma \log(b_{st}) + M_s * \alpha_z \log(b_{zt}) + M_s * X_{zt}\beta + M_s * \eta_t + M_s * \xi_{zt} + M_s * \epsilon_{szt}, \tag{6}$$

where $Y_{szt}$ is the number of adoptions on the street in that month. Although we do not know $M_s$, we can include interactions between street indicator variables and explanatory variables in order to control for the unobserved number of potential entrants and isolate the effect of the street-level installed base, $\gamma$. The large number of explanatory variables necessitates the use of streets with at least four installations; however, with the use of street-quarter fixed effects, these are the streets that provide the identifying variation anyway. Again we estimate the model using first-differences to control for time-varying unobservables as well as for correlated preferences. Our standard errors are again robust to heteroskedaticity, which is present by construction.

In the first column of Table 7 we show the results without the zip code installed base interacted with zip code indicator variables, and in the second column we include the interactions. The results without the interactions show that the effect of an installation on the same street has about seven times the effect an an installation elsewhere in the zip code. For the average adoption rate of 0.068, the elasticity of adoption is 9.28[12], so a one percent increase in the street-level installed base leads to a 9.28% increase in the street level adoption rate (eight times the effect of a zip-code installation in our zip code analysis at the average zip code adoption rate). These results provide evidence that the peer effect decreases with distance and operates at both the street and zip code-level, a useful result for understanding the geographic nature of how decisions about adoption solar PV panels are made.

# 5 Discussion

## 5.1 Can we truly identify peer effects?

While reflection is not a concern in our study since our specification is based on the effect of *past* installations, and endogenous group formation is completely controlled for using zip-quarter fixed effects, correlated unobservables can never be ruled out entirely

[12] Using the column two result.

in the absence of experimental variation. One could always argue that the unobservables operate at levels of finer granulation than the included controls. Ideally, to test for peer effects, we would have two geographic areas with identical environmental preferences, demographics, and macroeconomic shocks, and then randomly place a few solar PV installations in one of the areas to see whether the extra installations have a causal effect leading to more installations. Since this ideal randomized experiment is not possible, we do the next best thing: we exploit a temporary, exogenous difference in similar regions that serves to induce more installations in one region than the other. Once the difference is removed, if the region with more installations has a higher relative adoption rate, we can infer a causal effect of the extra installations on the rate of adoption.

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 cut through zip codes in a seemingly random manner – even cutting 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, and in this time period the only substantial difference between the regions is the level 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 has been installed in that administrative region. This implies that there are times when one CSI administrative zone moves to the next incentive step and for a limited period has lower incentives than the other CSI administrative zone in the zip code. Figure 4 shows this in a schematic. We 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. During the period in which the incentives are different, we obviously expect more installations to occur in the half of the zip code with higher incentives. What we are most interested in is what happens when the incentive steps are back in synch. At this point, the only difference between the two halves of a zip code are that one half received a "shock" of additional installations due to the temporarily higher incentives. If there is a positive, causal relationship between the number of previous installations and the rate of adoption, then we would expect the rate of adoption to *remain* higher in the side of the zip code that temporarily had higher incentives after the incentives realign; if there is no causal effect, then we would expect the adoption rates to also realign.

The transition from incentive step four to step five happened first for PG&E, then for SCE and finally by SDG&E.[13] 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 $\Delta t$ is the time between the completion dates of the solar PV installations, $Util$ is

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

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). $\eta_z$ are zip code indicators and $\xi_t$ are month indicators, which are included to control for unobservable trends and differences across zip codes that might confound our results. $\varepsilon$ is assumed to be a mean-zero stochastic error term. An observation in this estimation is an installation. The results are given in Table 8. Along both borders, the adoption rate increases (time between adoptions decreases) for the administrative district which has not changed step (i.e., received the shock), as shown by the negative coefficient on the interaction $\beta_3$. 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 $\beta_4$, implying a causal effect of the added installations on the adoption rate.

## 5.2 Comparison to other models

The Bass model is the classic workhorse of the marketing diffusion literature with aggregate data (Bass 1969; Mahajan et al. 1995). As a benchmark, we compare our results to those using the traditional Bass model, in which the probability of adoption at time $t$ (given that adoption has not occurred) is equal to:

$$n(t)/[m - N(t)] = p + (q/m)N(t), \tag{7}$$

where $p$ and $q$ are the coefficients of innovation and imitation, respectively, and $m$ is the market potential. Then the first order condition is given by:

$$n(t) = [m - N(t)][p + (q/m)N(t)] = pm + (q - p)N(t) - (q/m)[N(t)]^2. \tag{8}$$

We run this linear regression estimating coefficients equal to the expressions $pm$, $q-p$, and $-q/m$. The Bass model is intended to study diffusion using aggregate data, so we begin by using $n(t)$ and $N(t)$ for the number of new and cumulative installations throughout California. However, because we are interested in diffusion at a more localized level, we also run this regression using zip code new and cumulative installations for $n(t)$ and $N(t)$, respectively, with and without zip code fixed effects.[14] The regression results are in Table 9, as are the implied values for p, q and m.

We compare the effect of early adoptions since our model and the Bass model have different ways of including a declining effect of new installations: the linear probability through the additive log specification, and the Bass model through the declining number of potential adopters. Using the Bass model results in column three, the effect of an additional installation on the probability of adoption in a zip code with negligible adoption thus far, $q/m$, is $1.69$ x $10^{-5}$. In contrast, we find the effect of the first installation to be $3.55$ x $10^{-6}$ on the probability of adoption.[15] The Bass model estimate of the peer effect is five

[14] All of the three Bass model specifications yield similar results for the coefficient of imitation, q, and as we would expect, the estimated market potential is much larger for the aggregated data since it is for all of California instead of for a single zip code. The value of the coefficient of imitation falls in the range reported by Sultan et al. (1990).

[15] $(5.12$ x $10^{-6})(\log(2) - \log(1))$.

times larger than ours. This is not surprising since many unrelated factors – including increasing demand over time and clustering of installations for any reason – are attributed to peer effects in the Bass model.

We also compare our results to those using a Poisson model at the daily level, first with zip-quarter random effects and no time dummies and then with zip-quarter random effects with monthly dummies.[16] We use the number of installations in the zip code on a day as the dependent variable. The issue with this specification is that time-varying factors are assumed to affect all zip codes the same, rather than all individuals the same, as in our model. We find that the effect of the log installed base on the probability of an installation is 0.592 (including the time fixed effects). For an average zip code with 4,955 owner-occupied homes, this corresponds to a household effect of $1.195$ x $10^{-4}$, an absurdly high estimate. The issue may be either the distributional assumption on the random effects (gamma) or the unrealistic assumption that time-varying factors affect all zip codes the same.

In order to explicitly model the probability of a household's adoption, as we do in our model, and to include fixed effects instead of random effects, we can use the generalized method of moments (GMM) estimator described in Woolridge (1997) and Windmeijer (2000), which essentially uses scaled first differences of the dependent variable and lags of the regressors as instruments in order to estimate a count or probability model with multiplicative fixed effects. This method is computationally intensive even when aggregating the data at a monthly level, especially as additional explanatory variables are added. We

[16]As before, the fixed effects estimator is biased since the installed base is a predetermined regressor.

use monthly aggregated data, where the dependent variable is the probability of a household adoption that month (in millionths) and we again scale the dependent variable in terms of a daily adoption rate. The resulting estimated coefficient is 18.2 – three times the result in our first-differences specification. However, the GMM estimation strategy assumes independence of the error terms, which is a strong assumption in our setting.

### 5.3 Marketing and policy implications

The presence of a causal peer effect has important marketing implications for firms. First, prices should reflect the fact that new installations increase the likelihood of future installations. So in the presence of peer effects, prices early on should be set lower than would be the case if the objective is to maximize current period profits. However, since the the peer effect does not appear to be contractor-specific, the positive externality due to the peer effect will not likely be fully internalized in the pricing behavior of contractors performing the installations. This provides some justification for the declining CSI incentives, which are designed to expedite adoption in the early years. In addition, the significant interactions between the installed base and the demographic variables imply that the visibility of installations is important. Firms already appear to know this since efforts are made to increase the visibility of installations.

How would the diffusion process have been different in the absence of a peer effect? We simulate the diffusion of solar PV over five years using the 2003 installed base as our starting point. We do this with and without a household-level peer effect of size $5.12$

x$10^{-6}$, as estimated in our primary specification. We include a base adoption rate with a positive time trend of $1.59$ x $10^{-7}$ which we estimate from the data when including a specification with a linear time trend. We include heterogeneity in the peer effect and/or time trend across zip codes by multiplying the estimated values by a random number drawn from a uniform distribution between zero and two (so that the mean remains the same). Figure 5(a) shows the installed base over time comparing what would happen with and without peer effects, and with and without zip code heterogeneity in the base adoption rate (averaged over 100 simulations). Peer effects increase the average number of adoptions (assuming homogenous zip codes) from 904 to 999, an increase of 10.5%. In zip codes with heterogeneous base adoption rates, the increase is 14.3%, from 805 to 920. Heterogeneity in the baseline adoption rates leads to less adoption since the effect of additional installations declines with installed base size; in other words, previous installations are more valuable in increasing overall adoption when they occur in zip codes with lower baseline adoption rates.

Figure 5(b) shows simulations with the heterogeneity on the peer effect instead of the baseline adoption rate. Heterogeneity on the peer effect leads to the same level of adoption as without the heterogeneity, and the the effect of the peer effect on overall adoption is the same. Figure 5(c) includes (perfectly) positively correlated heterogeneity on both the peer effect and baseline adoption rate, and Figure 5(d) includes negatively correlated heterogeneity on both the peer effect and baseline adoption rate. The heterogeneity in the baseline adoption rate again leads to overall decreases in adoption, and this is exacerbated when the heterogeneity in adoptions is positively correlated with the heterogeneity in the peer effect. When the peer effect heterogeneity is negatively correlated with the heterogeneity in the baseline adoption rate, the negative effect of the baseline adoption heterogeneity on overall adoption is partially mitigated through the peer effect, since the peer effect is more likely to lead to more adoptions in zip codes with lower overall adoption. These results indicate that marketers should work to increase the level of adoption in zip codes with low installed bases, especially in areas with characteristics indicative of larger peer effects.

# 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 population density or the "greenness" of the zip code. Furthermore, industry reports and contractor marketing strategies all suggest that there may be more than simply geographically correlated preferences underlying this pattern of clustering. Specifically, there may be peer effects, whereby previous choices of one's neighbors influence the decision of whether to adopt a solar PV panel. We use a rich installation-level dataset to explore whether there is quantitative evidence to support this contention. Our methodology leverages the exact timing of installations with the fact that requested installations do not immediately enter the installed base, along with a first-differences estimation strategy in order to include an extremely rich set of fixed effects while avoiding both aggregation bias and the bias present in traditional fixed effects regression with predetermined or endogenous regressors.

We find strong evidence of a causal peer effect, where a one percent increase in the zip code installed base increases the adoption rate by just over five percent, with the exact value depending on the current adoption rate in the zip code. Our quasi-experimental evidence provides additional support for the causal nature of the peer effect. We further find evidence that the peer effect is even stronger at a finer geographic level, specifically at the street-level, and that key demographic variables can help explain the total level of adoption and influence the magnitude of the peer effect. There are a variety of mechanisms through which the peer effect may operate including image motivation and social learning. Since we find no evidence that the peer effects are contractor-specific, even if there is information transfer, potential adopters who are more likely to adopt due to nearby installations are not significantly influenced in their choice of contractor.

Our results provide insight into the nature of diffusion of emerging green products. In addition, we provide a straightforward methodology for the estimation of peer effects with the inclusion of a rich set of fixed effects to control for the possible confounding factors. Our belief is that the peer effects occur due to a combination of social learning and image motivation. Since our results are only suggestive of the mechanisms by which the peer effects work, disentangling the relative importance of these mechanisms is a prime topic for future research – one likely to require a carefully developed experimental research design at the neighborhood level.

# References


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

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

Arzaghi, M. and Henderson, J. V. (2007), 'Networking off madison avenue', *Review ofEconomic Studies* **75**(4), 10111038.

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.

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.

Bell, D. R. and Song, S. (2007), 'Neighborhood effects and trial on the internet: Evidence from online grocery retailing', *Quantitative Marketing and Economics* **5**, 361–400.

Bertrand, M., Luttmer, E. F. P. and Mullainathan, S. (2000), 'Network effects and welfare cultures', *Quarterly Journal of Economics* **115**(3), 1019–1055.

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

Choi, J. and Bell, S. K. H. D. R. (2010), 'Spatiotemporal analysis of imitation behavior across new buyers at an online grocery retailer', *Journal of Marketing Research* **47**(1), 75–89.

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* .

Danaher, P. J., Hardie, B. G. S. and Jr., W. P. P. (2001), 'Marketing-mix variables and the diffusion of successive generations of a technological innovation', *Journal of Marketing Research* **38**(4), 501–514.

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.

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.

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.

Hartmann, W. R., 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**, 287304.

Iyengar, R., den Bulte, C. V. and Valente, T. W. (2011), 'Opinion leadership and social contagion in new product diffusion', *Marketing Science* **30**(2).

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.

Kratzer, J. and Lettl, C. (2009), 'Distinctive roles of lead users and opinion leaders in the social networks of schoolchildren', *Journal of Consumer Research* **36**, 646659.

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.

Mahajan, V., Muller, E. and Bass, F. M. (1990), 'New product diffusion models in marketing: A review and directions for research', *The Journal of Marketing* **54**(1), 1–26.

Mahajan, V., Muller, E. and Bass, F. M. (1995), 'Diffusion of new products: Empirical generalizations and managerial uses', *Marketing Science* **14**, G79–G88. Special Issue on Empirical Generalizations in Marketing.

Manchanda, P., Xie, Y. and Youn, N. (2008), 'The role of targeted communication and contagion in product adoption', *Marketing Science* **27**(6), 961976.

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

McShane, B., Bradlow, E. and Berger, J. (2010), 'Multivariate visual diffusion for social groups: How social identity influences when and what people buy', *working paper* .

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

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

Munshi, K. and Myaux, J. (2006), 'Social norms and the fertility transition', *Journal of Developmental Economics* **80**(1), 138.

Nair, H., Manchanda, P. and Bhatia, T. (2010), 'Asymmetric social interactions in physician prescription behavior:the role of opinion leaders', *Journal of Marketing Research* **47**, Vol. XLVII (October 2010), 883895.

Nam, S., Manchanda, P. and Chintagunta, P. K. (2010), 'The effects of service quality and word-of-mouth on customer acquisition, retention and usage', *Marketing Science* **29**(4), 690700.

Narayanan, S. and Nair, H. (2011), 'Estimating causal installed-base effects: A bias-correction approach', *Stanford Graduate School of Business Research Paper Series No. 2076* .

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.

Norton, J. A. and Bass, F. M. (1987), 'A diffusion theory model of adoption and substitution for successive generations of high-technology products', *Management Science* **33**(9), 1069–1086.

Petersen, T. (1991), 'Time-aggregation bias in continuous-time hazard rate models', *Sociological Methodology* **21**, 263–290.

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.

Sacerdote, B. (2001), 'Peer effects with random assignment: Results for dartmouth roommates', *Quarterly Journal of Economics* **116**, 681–704.

Schmittlein, D. C. and Mahajan, V. (1982), 'Maximum likelihood estimation for an innovation diffusion model of new product acceptance', *Marketing Science* **1**(1), 57–78.

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.

Sorensen, A. (2006), 'Social learning and health plan choice', *RAND Journal of Economics* **37**(4), 929945.

Srinivasan, V. and Mason, C. H. (1986), 'Nonlinear least squares estimation of new product diffusion models', *Marketing Science* **5**(2), 169–178.

Sultan, F., Farley, J. U. and Lehmann, D. R. (1990), 'A meta-analysis of applications of diffusion models', *Journal of Marketing Research* **27**(1), 70–77.

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

Ter Hofstede, F. and Wedel, M. (1998), 'A monte carlo study of time aggregation in continuous-time and discrete-time parametric hazard models', *Economics Letters* **58**, 149–156.

Topa, G. (2001), 'Social interactions, local spillovers and unemployment', *Review of Economic Studies* **68**(2), 261295.

Van den Bulte, C. and Joshi, Y. V. (2007), 'New product diffusion with influentials and imitators', *Marketing Science* **26**(3), 400421.

Windmeijer, F. (2000), 'Moment conditions for fixed effects count data models with endogenous regressors', *Economics Letters* **68**(1), 21–24.

Woolridge, J. M. (1997), 'Multiplicative panel data models without the strict exogeneity assumption', *Econometric Theory* **13**(5), 667–678.

Table 1: Zip code-level summary statistics

| Variable | Mean | Std. Dev. | Min. | Max. | N |
|---|---|---|---|---|---|
| Zip code number of residential installations | 48.057 | 73.157 | 1 | 695 | 1646 |
| Zip code MW of residential installations | 0.252 | 0.411 | 0 | 3.675 | 1646 |
| population (100,000s) | 0.244 | 0.215 | 0 | 1.095 | 1287 |
| household size | 2.828 | 0.609 | 0 | 5.21 | 1287 |
| median income | 6.346 | 2.896 | 0 | 37.5 | 1287 |
| % pop male | 50.239 | 3.231 | 34.4 | 97.8 | 1287 |
| % pop who are white | 65.247 | 20.205 | 4.4 | 95.2 | 1287 |
| % pop with college degrees | 38.119 | 17.58 | 4.115 | 95.731 | 853 |
| % pop between 20 and 45 | 33.258 | 7.892 | 3.9 | 79.600 | 1287 |
| % pop over 65 | 12.38 | 6.155 | 0 | 80.900 | 1287 |
| % pop who drive to work | 86.204 | 10.367 | 4.348 | 100 | 1300 |
| % pop who carpool | 14.681 | 6.495 | 0.469 | 55.875 | 1277 |
| % pop using public transit | 3.917 | 5.712 | 0.058 | 42.593 | 1023 |
| % pop who work at home or walk to work | 8.827 | 6.888 | 1.617 | 61.496 | 1204 |
| % pop with over a 30 min commute | 38.211 | 12.732 | 5.371 | 80.881 | 1115 |
| % pop who drive a hybrid | 3.596 | 5.761 | 0 | 100 | 1354 |
| number of owner occupied homes (1000s) | 4.955 | 4.2 | 0 | 18.965 | 1287 |
| median value owner occupied home | 0.534 | 0.26 | 0 | 1 | 1287 |
| home loan | 121.155 | 67.864 | 0 | 576 | 1287 |
| home repair | 122.781 | 70.405 | 0 | 585 | 1287 |
| fraction of homes worth 0-50K | 2.595 | 3.718 | 0 | 53.2 | 1287 |
| fraction of homes worth 50-90K | 2.041 | 2.688 | 0 | 37.3 | 1287 |
| fraction of homes worth 90-175K | 5.876 | 7.678 | 0 | 61.7 | 1287 |
| fraction of homes worth 175-400K | 30.567 | 23.227 | 0 | 89.7 | 1287 |
| fraction of homes worth 400K+ | 58.609 | 29.825 | 0 | 100 | 1287 |

Note: Summary statistics for residential installations in zip codes with at least one installation.

Table 2: Residential installation size and price

| Variable | Mean | Std. Dev. | Min. | Max. |
|---|---|---|---|---|
| size (kW) | 5.24 | 3.28 | 0.11 | 48.3 |
| price ($/W) | 8.49 | 4.16 | 0.28 | 697.33 |
| N | | 79,101 | | |

Table 3: Zip Code Linear Probability Model (obs=zip-day)

| | **OLS** | **ZIP FE** | **ZIP-QUARTER FE** | **ZIP-QUARTER FD** |
|---|---|---|---|---|
| log installed base | 0.174 | -0.245 | -13.556 | 5.117 |
| | (0.405) | (0.496) | (5.076) | (2.828) |
| R-squared | 0.000 | 0.001 | 0.016 | 0.000 |
| N | 5,908,738 | 5,908,738 | 5,908,738 | 5,843,356 |

Table 4: Zip Code Linear Probability Model (obs=zip-month)

| | **OLS** | **ZIP FE** | **ZIP-QUARTER FE** | **ZIP-QUARTER FD** |
|---|---|---|---|---|
| log installed base | 0.147 | -0.216 | -5.244 | -4.003 |
| | (0.376) | (0.460) | (.) | (1.583) |
| R-squared | 0.003 | 0.027 | 0.467 | 0.002 |
| N | 194,864 | 194,864 | 194,864 | 129,482 |

Table 5: Installation Size Regressions

| | **1** | **2** | **3** |
|---|---|---|---|
| installed base | 0.0292 | 0.0481 | 0.0815 |
| | (0.0029) | (0.0045) | (0.0086) |
| price | -0.0739 | -0.0612 | -0.0529 |
| | (0.0396) | (0.0318) | (0.0246) |
| year-month dummies | Y | Y | N |
| month dummies | N | N | Y |
| zip FE | N | Y | N |
| zip-quarter FE | N | N | Y |
| R-squared | 0.057 | 0.202 | 0.382 |
| N | 47,029 | 46,752 | 38,931 |

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: First-Differenced Street-Level Linear Probability Model (obs=zip-month)

| | **FD 1** | **FD 2** |
|---|---|---|
| log number of previous installations on street | 0.660 | 0.631 |
| | (0.032) | (0.031) |
| log zip code installed base | 0.091 | street-specific |
| | (0.023) | |
| With street indicator variable interactions with zip code installed base | N | Y |
| R-squared | 0.352 | 0.395 |
| N | 6,139 | 6,139 |

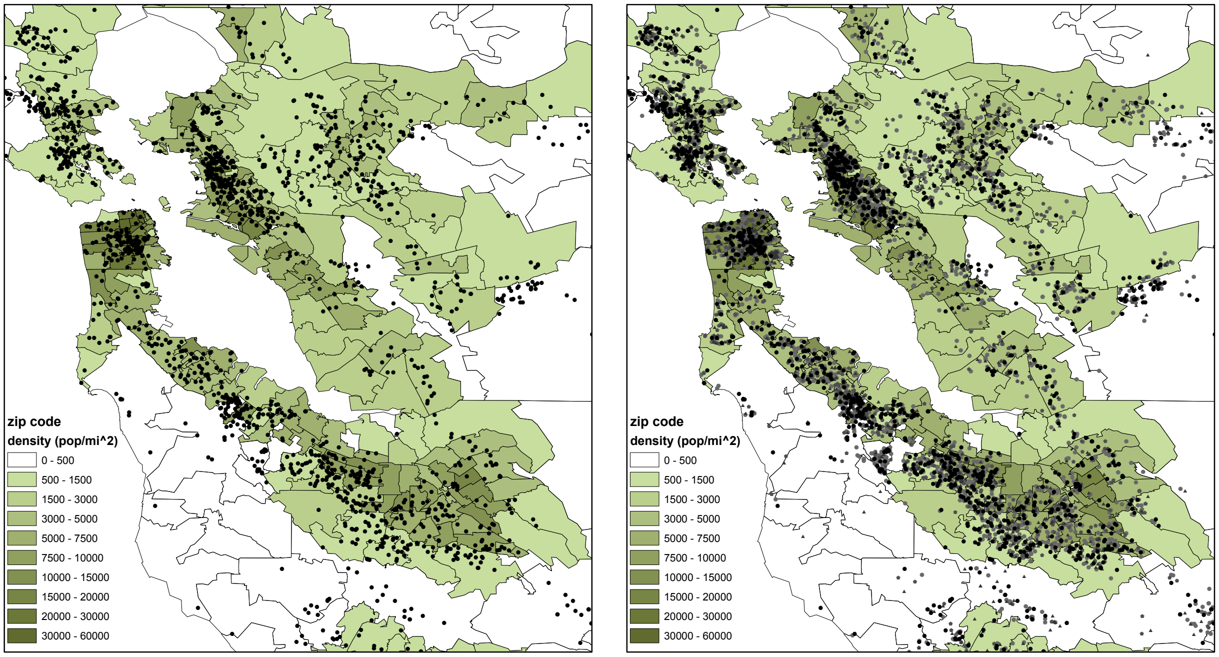
Table 8: Regressions for incentive step transitions

| | 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 ($\eta_z$) | Y | Y |
| Month Controls | N | Y |
| R-squared | 0.603 | 0.376 |
| N | 72 | 213 |

Transition from incentive step four to five, where transition order is PG&E, CSE, SDG&E.

Table 9: Bass Model

| | Aggregate | Zip no FE | Zip with FE |
|---|---|---|---|
| constant (pm) | 72.838 | 1.131 | 1.000 |
| | ( 33.796 ) | (0.014) | (0.000) |
| cumulative installations (q-p) | 0.027 | 0.023 | 0.022 |
| | (0.00418) | (0.001) | (0.001) |
| cumulative installations squared (-q/m) | $-3.27 \text{ x } 10^{-08}$ | $-1.4 \text{ x } 10^{-5}$ | $-1.7 \text{ x } 10^{-5}$ |
| | ($6.4 \text{ x } 10^{-8}$) | ($2.0 \text{ x } 10^{-6}$) | ($2.0 \text{ x } 10^{-6}$) |
| R-squared | 0.745 | 0.260 | 0.299 |
| N | 147 | 42,690 | 42,696 |
| p | $8.79 \text{ x } 10^{-5}$ | $6.61 \text{ x } 10^{-4}$ | $7.43 \text{ x } 10^{-4}$ |
| q | 0.0271 | 0.0238 | 0.0227 |
| m | 82,800 | 1,710 | 1,340 |




(a) 2001-2003

(b) 2004-2006

Figure 1: Clustering in Solar PV installations in the San Francisco Bay Area

Figure 2: Zip code empirical hazard rates

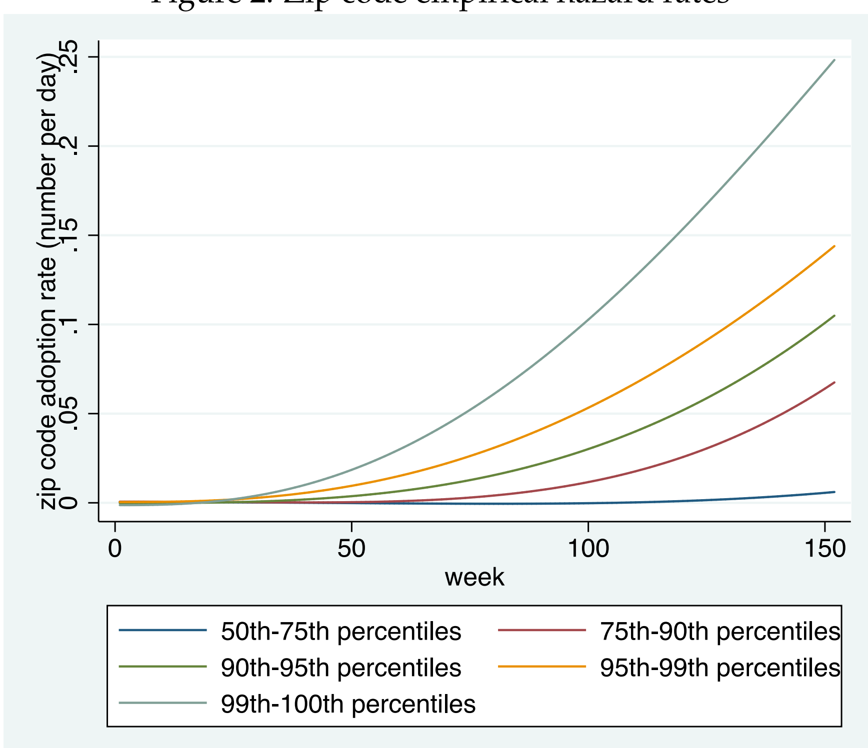


Figure 3: Consumer Reasons for Installing Solar (2009 CSI Impact Report Survey)

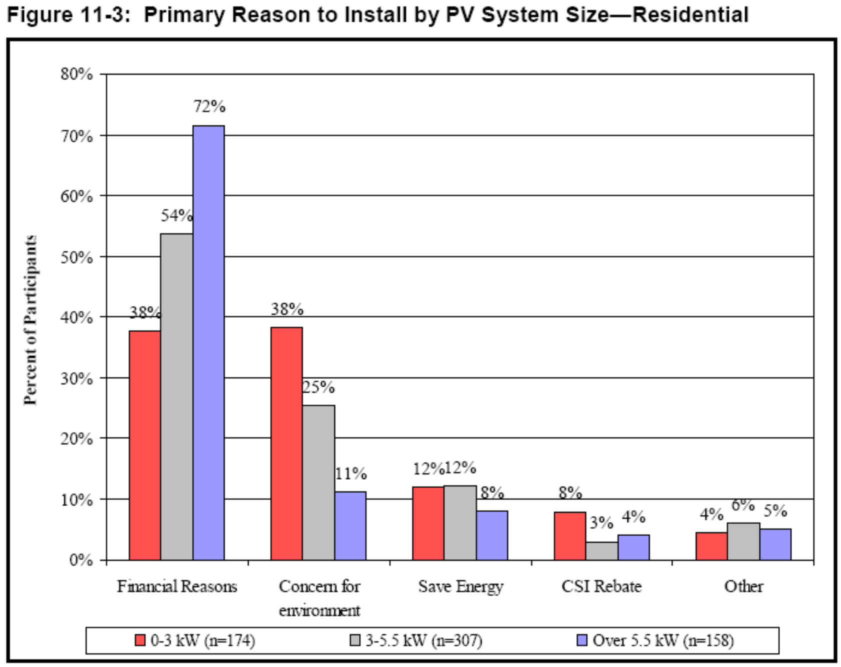
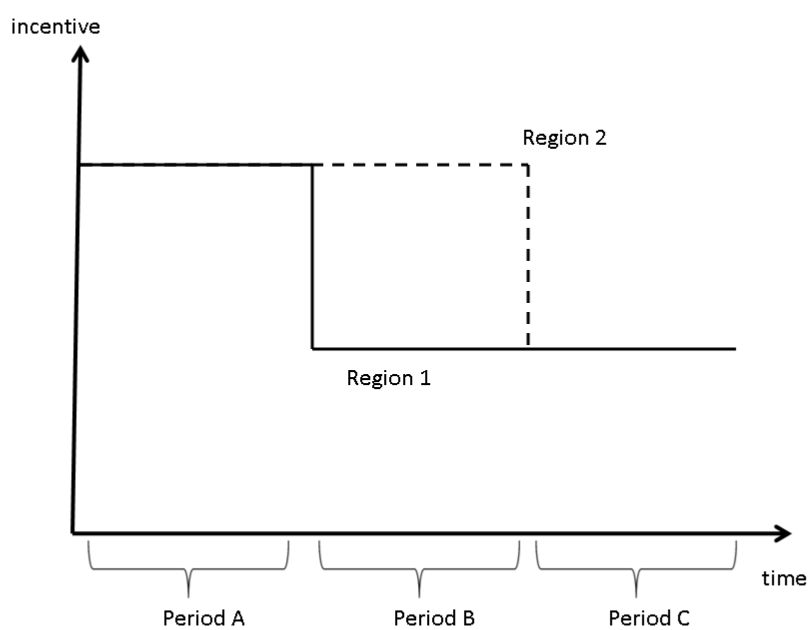
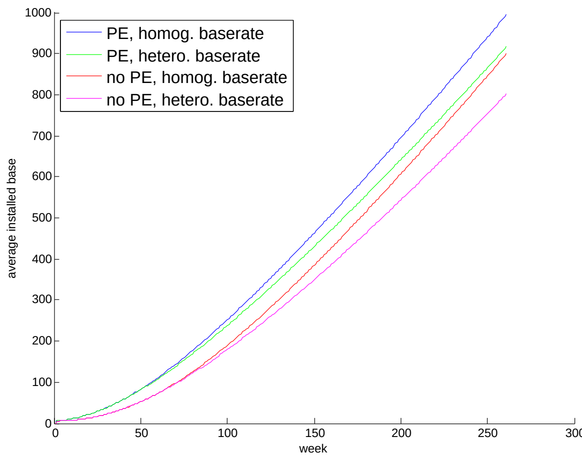
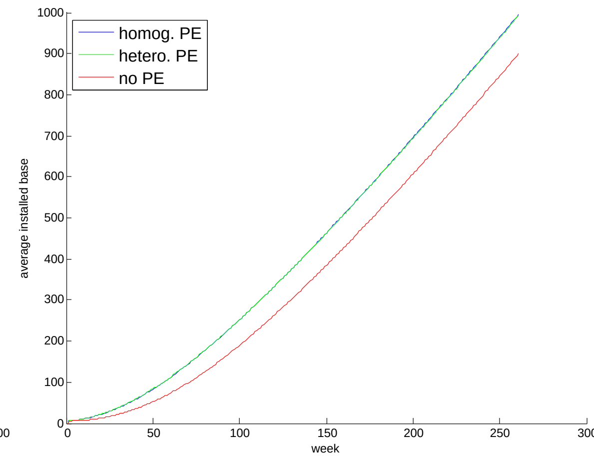



Figure 4: The quasi-experiment exploits changing incentives within the same zip code

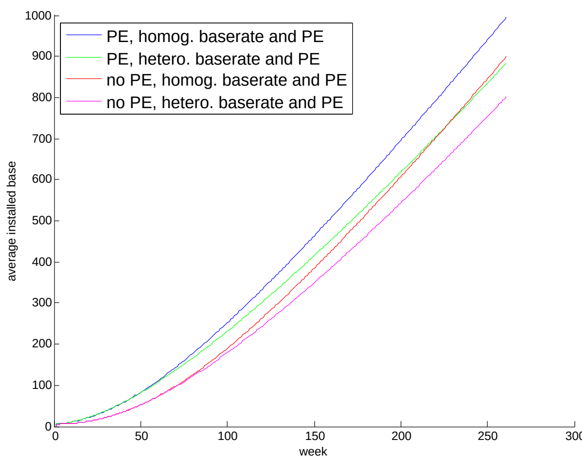




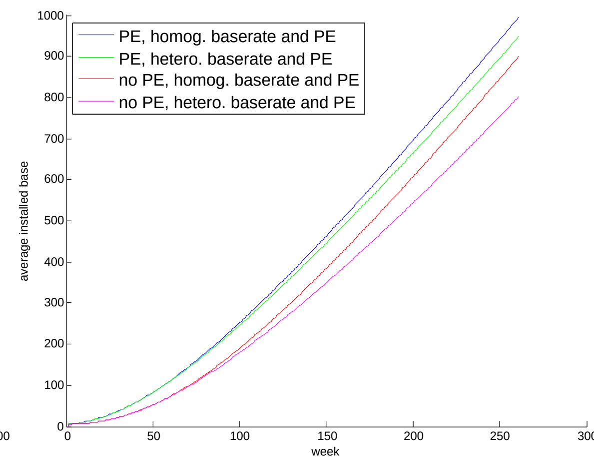

(a) Heterogeneity in baseline adoption




(b) Heterogeneity in peer effect




(c) Positively correlated heterogeneity in baseline adoption and peer effect




(d) Negatively correlated heterogeneity in baseline adoption and peer effect

Figure 5: Overall simulated adoption with and without the peer effect