Peer Effects in the Diffusion of Solar Photovoltaic Panels*

Bryan Bollinger

NYU Stern School of Business†

Kenneth Gillingham

Yale School of Forestry & Environmental Studies‡

April 25, 2012

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 since social spillovers can be used expedite adoption. 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. We study the diffusion of solar photovoltaic (PV) panels in California, and find that at the average number of owner-occupied homes in a zip code, an additional installation increases the probability of an adoption in the zip code by 0.87 percentage points. Our results provide valuable guidance to marketers designing strategies to increase referrals and reduce costs of customer acquisition. They also provide insights into the diffusion process of environmentally-friendly technologies.

*The authors would like to thank Hunt Allcott, Tim Bresnahan, Jesse Cunha, Misha Dworsky, Wesley Hartmann, Sridhar Narayana, Harikesh Nair, Peter Reiss, Arthur van Benthem, Ali Yurukoglu, the participants in the Stanford Environmental Economics seminar, Stanford Economics IO Workshop, Stanford MS&E PERR seminar, UC-Berkeley ARE seminar, Energy Institute at Haas seminar, and faculty and students from the Stanford GSB 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.

Factors affecting the adoption of new products have long been of great interest to marketers. Social interactions between consumers are of particular interest since they may result in social spillovers, in which a marketing action that affects one agent then indirectly influences other agents. Marketers often attempt to account for these social spillovers when determining the targeting and intensity of marketing activity. However, there must be a *causal* social interaction effect for the social spillovers to exist. In the classic aggregate diffusion models, such as the canonical Bass model (1969), social contagion is assumed to be a driving force behind accelerating adoption rates, but the estimated coefficient of contagion may capture a variety of factors influencing the diffusion of the technology, which may or may not include social interaction effects.

This paper makes several contributions to the literature on social interaction effects in the adoption of new technologies. Substantively, we document and estimate the magnitude of peer effects in the diffusion of an environmentally beneficial technology, solar photovoltaic (PV) panels. Policymakers are particularly interested in the diffusion of solar PV technology since increased adoption leads to reduced greenhouse gas emissions. Methodologically, we provide an empirical approach for the quick and straightforward estimation of peer effects using daily adoption data. Our methodology leverages the difference between the date an installation is requested and completed, along with a first-differences estimation method to (i) avoid 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, and (iii) avoid the biases present in traditional fixed effects estimation with endogenous or predetermined regressors. While a first-differences methodology is far from new, our estimation approach to deal with the endogeneity of predetermined regressors is novel.

The rest of the paper is organized as follows. In section 1 we discuss the difficulties of identifying causal peer effects using traditional methods in the literature and lay out our approach for addressing them. 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 present our results and evidence suggestive of different possible mechanisms behind the peer effect. We provide further evidence of causal peer effects using quasi-experimental variation in Section 4. Section 5 compares our results to those obtained using other diffusion models. Section 6 concludes with a discussion of the managerial implications of our findings.

1 Identification of Peer Effects

1.1 Perils in identifying peer effects

Causal peer effects are notoriously difficult to identify. To begin, there are three well-known issues that often confound the identification of peer effects: endogenous group formation leading to self-selection of peers (homophily), correlated unobservables, and simultaneity (Manski 1993; Brock and Durlaf 2001; Moffitt 2001; Soetevent 2006). Consider a setting where agent i in peer group z_i is making some choice (e.g., what technology to adopt) at the same time as others in the peer group.[1] Then the three issues can be easily seen in the common linear-in-means model:

$$a_{it} = \alpha + \beta \mathbb{E}[a|z_i, t] + \gamma' \mathbb{E}[\mathbf{x}|z_i, t] + u_{it}, \tag{1}$$

where a_{it} is the choice by agent i at time t, $\mathbb{E}[a|z_i, t]$ is the average number of adoptions in i's peer group z_i at time t, $\mathbb{E}[\mathbf{x}|z_i, t]$ is a vector of average characteristics in z_i at time t, and ϵ_{it} is a stochastic error term. While β is intended to capture the effect of the actions of others in z_i on the decision of agent i, it will not be consistently estimated if $\mathbb{E}[u_{it}|z_i, \mathbf{x}] \neq 0$. This is the case if there is self-selection of peers on common (unobserved) characteristics or if there are any other correlated unobservables, such as macroeconomic trends or time-varying localized marketing efforts. In addition, β may not be identified if the decision-maker is influencing others while also being influenced by them, a simultaneity issue known as "reflection." By taking the mean of (1) conditional on z_i and t and rearranging, we can see why simultaneity implies that β is not identified in this model:

$$\mathbb{E}[a|z_i, t] = \left(\frac{\alpha}{1-\beta}\right) + \left(\frac{\gamma'}{1-\beta}\right) \mathbb{E}[\mathbf{x}|z_i, t].$$

One can consistently estimate the *combined* effect $\left(\frac{\gamma'}{1-\beta}\right)$, but it is not possible to sep-

[1] The peer group can be defined using 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 et al. 2010; McShane et al. 2012; Nam et al. 2010; Narayanan and Nair 2012).

arately estimate the causal social interaction effect, β. Randomized field experiments are one way to address this issue (Bertrand et al. 2000; Sacerdote 2001; Duflo and Saez 2003). Hartmann et al. (2008) discuss many of the strategies often used with non-experimental data. One common approach used to address the self-selection of peers and correlated unobservables is to include a rich set of random or fixed effects (Manchanda et al. 2008; Nair et al. 2010). In this case, (1) can be rewritten as

$$a_{it} = \alpha + \beta \mathbb{E}[a|z_i, t] + \gamma' \mathbb{E}[\mathbf{x}|z_i, t] + \eta_{z_i,t} + u_{it}, \tag{2}$$

where $\eta_{z_i,t}$ are random or fixed effects for the peer group and time. If peer group effects flexibly control for group-specific characteristics, and if self-selection is the only issue, then $\mathbb{E}[u_{it}|z_i, \mathbf{x}] = 0$ and β can be consistently estimated. Correlated unobservables can also be addressed with a rich-enough set of random or fixed effects; if the correlated unobservables are time-varying, as in the case of targeted unobserved marketing, then peer group-time effects would be necessary.

There are several approaches to address the simultaneity issue. In some cases, detailed data on the peer group provides an exclusion restriction when there are only partly overlapping peer groups (De Giorgi et al. 2010; Nair et al. 2010). However, a more common approach is to have the agent's decision depend on *past* decisions made by agents in the peer group. A common way to implement this approach is to include the "installed base" of the product or service purchased in the relevant peer group as the explanatory variable of interest and condition x on the previous time period:

$$a_{it} = \alpha + \beta b_{z_i t} + \gamma' \mathbb{E}[\mathbf{x}|z_i, t-1] + \eta_{z_i,t} + u_{it}, \tag{3}$$

where the installed base is defined as $b_{z_i t} = \sum_{\tau=1}^{t-1} \sum_{j=1}^{m_{z_i}} a_{jt}$ and m_{z_i} is the number of potential adopters in z_i. In this context $a_{j\tau}$ is an indicator for an affirmative decision.

The installed base has been used extensively in the marketing literature in both the early aggregate diffusion models (Bass 1969; Mahajan et al. 1990, 1995) and more recently in numerous studies using disaggregate data (Sorensen 2006; Manchanda et al. 2008; Choi et al. 2010; Iyengar et al. 2011; Narayanan and Nair 2012; McShane et al. 2012). Unfortunately, there remain further challenges to using the installed base as the regressor of interest in the presence of random effects or fixed effects. The issue here relates closely to the "dynamic panel data" literature, which describes how models with fixed effects and random effects are inconsistently estimated when a lagged dependent variable is used (Nerlove 1971; Nickell 1981).[2] Since by definition b_{zt} contains lagged adoptions, inclusion of random effects or fixed effects will lead to inconsistent estimates, just as in the dynamic panel data literature. Manchanda et al. (2008) get around this by only including the past adoptions of *others* in the group, implicitly assuming that the errors are not correlated for individuals within the same peer group. Random effects have a further issue: the

[2]The issue is straightforward; consider the dynamic panel model $y_{it} = \alpha + \beta y_{it-1} + \mu_i + \epsilon_{it}$, where μ_i is a fixed effect for unit i. A model with random effects is inconsistently estimated if μ_i is correlated with y_{it-1}, yet this is the case by construction. A mean-differencing approach to fixed effects, $y_{it} - \bar{y}_{it} = \beta(y_{it-1} - \bar{y}_{it-1}) + (\epsilon_{it} - \bar{\epsilon}_{it})$, is inconsistently estimated since $(\epsilon_{it} - \bar{\epsilon}_{it})$ is correlated with $(y_{it-1} - \bar{y}_{it-1})$. A similar issue occurs using a first-differencing approach for fixed effects estimation.

installed base is the aggregation of previous adoption decisions of other consumers who share the same random effect, so the random effects and installed base are by construction correlated. Thus the random effects and installed base are not orthogonal, so random effects estimation is inconsistent.[3]

Narayanan and Nair (2012) study the diffusion of hybrid vehicles and address this issue with the installed base in two ways: an instrumental variables approach and a bias correction approach for fixed effects. Both hold promise in certain settings. When suitable instruments are available, an instrumental variables approach is quick and easy. In practice, suitable instruments tend to be extremely difficult to find, particularly because the installed base is a stock variable. The bias correction approach uses the assumption of i.i.d. errors to find a closed form solution for the bias as a function of the variance of the errors and the variance of installed base.[4] Unfortunately, the bias-adjustment is only appropriate when there is no autocorrelation in the errors.

1.2 Our methodology

Our methodology addresses the issues inherent in identifying social interactions and provides a consistent estimate of the effect of the installed base of solar PV panels on the household-level decision of whether or not to install solar, even in the presence of correlated unobservables. We leverage the fact that in the solar PV market, the decision to install solar does not lead to an instantaneous installation, due to the time needed to com-

[3]Maximum likelihood estimation with random effects can be consistent if the true distribution of the errors is known and specified correctly (Narayanan and Nair 2012).

[4]The bias is found to be $\text{plim}_{N\to\infty}(\hat{\beta}-\beta) = \frac{5}{9}\frac{\sigma_\epsilon^2}{\sigma_b^2}$, where σ_ϵ^2 is the variance of ϵ and σ_b^2 is the variance of b.

plete the necessary paperwork and perform the installation. Other situations in which there may be a time lag between the decision to adopt and the actual adoption include purchases of new vehicles, all online purchases, customized product purchases, and the adoption of regulated or incentivized products. In our situation, we would not expect social interactions to have an effect until the solar PV panels have actually been installed, after which point the adopter would have experienced the benefits of the panels and neighbors may then be able to see the installation.

Accordingly, we define the installed base in our setting as the cumulative number of *completed* installations at time t in zip code z. We drop the i subscript since we perform the analysis at the zip code level, and all households in a zip code have the same peer group: the other households in the zip code. Let $\tilde{a}_{zt}$ be an indicator variable for a completed installation (i.e., delivery of the product or service). Then $b_{zt} = \sum_{\tau=1}^{t} \sum_{j \in z} \tilde{a}_{j\tau}$. In our primary specification, we model the probability that a household in zip code z adopts solar in any given day t as:

$$Y_{zt} = \alpha + \beta b_{zt} + \gamma' X_{zt} + \eta_{zq} + \xi_t + \epsilon_{zt}, \quad (4)$$

where Y_{zt} is the fraction of owner-occupied households in z that had not previously adopted solar and decide to adopt solar on day t; η_{zq} are zip code-quarter fixed effects (q denotes a quarter); ξ_t are time indicator variables, including year-month indicators, day of the month indicators, and day of the week indicators; and ϵ_{zt} is a mean-zero stochastic error. X_{zt} contains additional explanatory variables that may vary over time, such as indicator variables for different levels of subsidies available for adopting solar.

We estimate (4) using several approaches, including traditional fixed effects estimation taking mean-differences on the zip code-quarter, zip-quarter random effects, and our preferred specification that uses first-differences to consistently estimate β, even with the predetermined b_{zt} and serial correlation.[5] The first-differences specification is:

$$(Y_{zt} - Y_{zt-1}) = \beta(b_{zt} - b_{zt-1}) + \gamma'(X_{zt} - X_{zt-1}) + (\xi_t - \xi_{t-1}) + (\epsilon_{zt} - \epsilon_{zt-1}). \tag{5}$$

The following proposition formally describes when β is consistently estimated in the first-differences model. Let l denote the time lag between the adoption decision and the installation of the solar PV panel and let ν denote the order of autocorrelation of ϵ.

Proposition 1. *If (i) $l > \nu + 1$ and (ii) the social interaction effects do not begin until the installation has been completed, then β in* (5) *is consistently estimated.*

See Appendix A for a proof. We could further relax assumption (ii), since in fact the consistency of β holds as long as the order of autocorrelation is less than the number of days after the decision to adopt before social interactions take effect.

2 Industry Background and Descriptive Evidence

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

[5]We drop the first observation in each quarter so that the zip-quarter effects cancel out in the first-differenced equation.

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).[6] 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 assembled an installation-level dataset of residential solar PV installations in the three investor-owned utility (IOU) regions from January 2001 to December 2011. 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 solar incentives for an installation, the date payment was submitted for the installation, and the date of completion. 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 85,046 requested residential installations between January 2001 and December 2011. Table 1 contains zip code level summary statistics for residential installations and key demographics.

If there are indeed peer effects in the diffusion of residential PV panels we would expect to see a spatial clustering of installations. 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 either peer effects or spatially correlated

[6] For an overview of the history of solar PV policy in California, see Taylor (2008).

preferences, perhaps due to clustering of environmental preferences (Kahn and Vaughn 2009). Moreover, we find a similar clustering pattern at the neighborhood level.

The presence of peer effects should also imply *accelerating* adoption in regions with more installations. Indeed, in our data 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 clustering pattern very clearly increases. The acceleration of installations can further be seen in the empirical hazard rate (averaged at the monthly level) shown in Figure 2.

Table 2 provides summary statistics for the probability of a household adoption in a zip code-day and the zip code installed base. To better understand the distribution of each of these, Figure 3 and Figure 4 provide histograms for the probability of adoption (conditional on there being an adoption in the zip code on that day) and the installed base. Not surprisingly, in 98% of the observations there are no adoptions in the zip code that day.

3 The Effect of the Installed Base

3.1 Effect on adoption rates

Table 3 presents our primary zip code-level results. Column one contains the OLS results from estimating (4) without zip code fixed effects, column two contains the results using traditional mean-differenced zip-quarter fixed effect, column three contains zip-quarter random effects, and column four presents the first-differenced results from estimating (5). Consistent with Narayanan and Nair (2012), the traditional mean-differenced results appear to be biased downwards. Using a Hausman test, we find that we can reject the orthogonality assumption of the random effects model at a 90% confidence level. In the traditional fixed effects regression, we find statistically significant and positive (albeit small) autocorrelation for the first two lags (days), while in the first-differences regression, we find a statistically significant, negative correlation of -0.5 for the first lag (a mechanical result of the first-differencing). This demonstrates that the order of autocorrelation is much less than the time between the adoption decision and the installation, which had a median of 163 days and is greater than 40 days 99% of the time.[7] The actual installation time (breaking ground to completion) is short, with a median of 11 days and an average of 16 days, so even if the social interaction effect began once the actual installation began, our estimation method is valid since the decision to install occurs so much earlier.

Our preferred coefficient estimate of 1.756 is statistically significant at the 1% confidence level and implies that every additional installation increases the probability of a household adoption in the same zip code by 1.756 x 10^{-6}. At the average number of owner-occupied homes in a zip code of 4,959, an extra installation in the zip code increases the probability of an adoption in the zip code by 0.87 percentage points. At the average number of installations in a zip code of 13.1, and the average household adoption rate of 4.390 x 10^{-6}, the installed base elasticity of adoption is 5.24. At the median installed base of 2 (in zip codes with at least one adoption), the elasticity is 0.80. These

[7] The mean is 195 days with a standard deviation of 121.

results provide clear evidence of a statistically and economically significant effect.[8]

Testing for a time-varying response As a robustness check, we estimate a model allowing for the response to new installations to change along with the level of the installed base. We find a small, positive, statistically significant effect of an included quadratic term, implying *increasing* effects of new installations. In another specification we interact the installed base and year. Figure 5 plots these coefficients[9] with robust standard errors, again indicating that the installed base effect increases over time. Such an increase may derive from recent marketing efforts to leverage peer effects to induce adoptions. 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 commonly increase the visibility of new installations by putting up a sign indicating that a solar PV panel has been installed at that location. The PG&E CSI administrators have even established "Solar Champion" training sessions for "citizens interested in helping spread the word about solar in their neighborhoods."

Demographic interactions A critical question for practitioners is where to target marketing efforts to best leverage social spillovers. As a starting point, we include zip code demographics and demographics interacted with the installed base in an OLS estimation

[8] We also perform the first-differences regression including zip code fixed effects which essentially includes a linear, deterministic time trend for each zip code (Amiti and Wei 2006), and our results do not change significantly.

[9] The estimate for 1999 is not shown for visibility issues, as it is large and insignificant.

of (4) to examine the determinants of adoption (Table 4).[10] Our proxy for environmental preferences, hybrid vehicle adoption, has a large and highly statistically significant effect on the rate of adoption in a zip code. Higher adoption rates are also associated with the percentage of the population who are male, white, have over a thirty minute commute, and have home repairs. Lower adoption rates are associated with variables such as the zip code population, percentage of people aged 20-45 and over 65, and zip codes with higher value homes.

To assess the conditions under which social spillovers are likely to be most effective, we estimate (5) including interactions between the installed base and demographic variables.[11] We find that zip codes with larger household sizes and fraction of people with more than a thirty minute commute have a larger peer effect, while zip codes with higher median household income and more people who carpool have a smaller peer effect. The significance of the household size and commuting variables may suggest that the mere visibility of installations may contribute to the peer effect; 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 may have the opposite effect by reducing driving time.

The interaction with median income is economically significant: running the regression on only the lower half of the zip code income distribution leads to a coefficient and standard error of 2.478 (0.680), while for the top half it is 1.243 (0.454). So while house-

[10]Due to the inclusion of demographics, we do not include zip-quarter fixed effects in this estimation, and thus we view these results as suggestive.

[11]The "main effects" from the demographics are differenced out.

holds with lower income may be less likely to adopt solar, they are more influenced by the peer effect. This is consistent with the high percentage of consumers naming financial reasons as a primary reason for adopting solar, especially if the financial benefits of adoption are learned through information exchange.[12]

The effect of installation size If the visibility of installations matters, are peer effects stronger when the existing installations are larger? Table 5 provides evidence suggesting that this is likely to be the case. We estimate (5) using the installed base measured in megawatts as the explanatory variable of interest, to allow larger installations to have a greater impact, and we again find a positive, significant effect of the installed base (column two).[13] To assess the relative explanatory power of the installed base measured in terms of number of installations versus megawatts of installations, we estimate a model including both. We find that both become statistically insignificant (although jointly significant), so we cannot determine which is more important.[14] We hypothesize that the size of the installation may matter since larger installations are slightly more visible, but that spreading out the same number of megawatts over more homes increases the visibility even more.

[12]In a phone survey of 639 participants and 601 non-participants in the CSI program, 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 (CPUC 2009).

[13]When using the number of installations per household (column three), or the number of megawatts per household (column four), we find positive but insignificant effects.

[14]Details can be found in an online appendix.

Is the peer effect contractor-specific? If the peer effect operates in part through word-of-mouth, one might expect the peer effect to be contractor-specific. To test this, we estimate (5) using the probability of an adoption in a zip code by a *specific contractor* as the dependent variable. We also include both the contractor-specific installed base and the total zip code installed base (both first-differenced) as explanatory variables. We estimate the model for the top five contractors by total installations (over 1,000 residential installations each). The coefficient on the contractor-specific installed base in all five of the regressions turns out to be statistically insignificant, suggesting that peer effects in solar PV have a limited contractor-specific effect.[15]

3.2 Effect on installation size

Peer effects that work through information transfer may serve to reduce the uncertainty in the expected value of installations. For example, consumers may discuss tax incentives, rebates or the actual production of a solar PV system. If consumers are risk averse, we would expect to see installation sizes increase as more neighbors adopt solar PV systems, since larger installations are riskier (assuming the risk is per-MW).

The summary statistics for the size and price of all residential installations are shown in Table 6.[16] To test our prediction, we estimate a linear model of installation size (conditional on an installation occurring), first using OLS regression, then including zip code

[15]Details can be found in an online appendix.

[16]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.

fixed effects, and finally with zip-quarter fixed effects. Since the installed base is no longer a predetermined variable, traditional fixed effects estimation provides a consistent estimate. We find a positive, significant coefficient on the zip code installed base when using zip code fixed effects, as shown in Table (7).[17] This is consistent with peer effects reducing the uncertainty of the solar PV investment, leading to larger installations on average.

3.3 Effect on adoption at the street-level

If peer effects are major factor underlying the clustering pattern in the data, we might expect to see evidence of peer effects at a more localized level than the zip code. With 2001-2006 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. We create a panel dataset where an observation is a street-month.[18] Table 8 provides summary statistics, 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 each street's zip code.

We regress an indicator variable for an adoption on an indicator for an installation already having occurred on the street and the zip code installed base. Since many streets

[17]The significance disappears with zip-quarter fixed effects, but this is not surprising since we have only a single observation per installation.

[18]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.

are in relatively early stages of adoption, we have sufficient variation for our analysis. Appendix B contains further details of our specification, and the results are shown in Table 9. The second column adds interactions between the installed base and zip code indicator variables to nonparametrically control for the unknown number of potential adopters on each street. The results indicate that each installation increases the monthly probability of an additional installation by approximately 15 percentage points. The effect of an installation elsewhere in the zip code is not statistically significant. These results provide evidence that the peer effect decreases with distance, with stronger peer effects at the street-level.

4 Further Evidence of Causal Peer Effects

Ideally, to test for causal 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 exogenously placed installations lead to additional installations. Since this ideal randomized experiment is not possible, our above approach flexibly controls potential correlated unobservables and homophily. Our empirical setting also allows for an alternative approach to estimate causal peer effects motivated by the ideal experiment: two similar regions and a temporary, exogenous difference in incentives that serves to induce more installations in one region than the other. If the region with more installations continues to have a higher adoption rate once the difference is removed, then we can infer a causal effect of the extra installations on the rate of adoption.

We use a difference-in-differences approach, focusing on zip codes located around the border between IOUs. We use 33 zip codes along the border between Pacific Gas & Electric (PG&E) and Southern California Edison (SCE), and 174 along the (more populous) border between SCE and San Diego Gas & Electric (SDG&E).[19] 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, as illustrated in Figure 6. To examine the treatment effect of the extra installations, we estimate:

$$Y_{zt} = \beta_0 + \beta_1 S_i + \beta_2 B_i + \beta_3 C_i + \beta_4 (S_i \cdot B_i) + \beta_5 (S_i \cdot C_i) + \zeta_z + \zeta_z \cdot t + \varepsilon_i, \tag{6}$$

where Y_{zt} is defined as before, S_i is an indicator for whether the electric utility received a "shock" of relatively higher incentives (e.g., SDG&E in Figure 6), B_i is an indicator for the period of the higher incentive, C_i is an indicator for the period afterwards when the incentives were realigned. ζ_z are zip code indicators, t is a linear time trend, and ε_i is a mean-zero stochastic error.

We use the transition from step five to six for the PG&E/SCE border and the transition from step two to three for the SCE/SDG&E border for the analysis. These are the first

[19] Details can be found in an online appendix.

steps with a large difference in the step transition times (62 and 181 days, respectively).[20] The results are given in Table 10. As expected, the adoption rate is larger on the side of the border that did not change step (i.e., had a relatively larger incentive). More importantly, the adoption rate *remained* higher after the incentives are realigned, providing further evidence for a causal effect of additional installations.

5 Comparison to Other Diffusion Models

As a benchmark and robustness check we compare our results to those from other common models of technology diffusion. We begin with the canonical Bass model commonly used to study product diffusion with aggregate data. Appendix C provides details on the Bass model estimation. Table 11 presents the estimation results, which are performed at the daily level. Column one shows the results using data aggregated over all zip codes. Columns two and three report the results at the zip code-level, with and without zip code fixed effects. All three specifications yield similar results for the coefficient of imitation, q. The estimated value of q falls in the range reported by Sultan et al. (1990). The estimated market potential is much larger for the aggregated data since it corresponds to all of California instead of a single zip code.

In the aggregate Bass model we find that one additional installation increases the probability a household adopts (q/m) by 4.51 x 10^{-9}. This cannot be directly compared to our first-differences estimate (1.76 x 10^{-6}) since the aggregate model assumes that an instal-

[20] Previous steps are small – only fifteen days between the transitions from step four to five for the PG&E/SCE border and one day for the SCE/SDG&E border – or the incentive transitions switch.

lation affects the probability of adoption everywhere in California, rather than simply in the zip codes where adoption is occurring. This difference underscores why aggregate diffusion models are unsuitable for the estimation of *localized* peer effects. Interestingly, our results from the zip code-level Bass model turn out to be quite close to our preferred estimate of (5), and in fact, are not statistically significantly different. Yet the average market size estimates from the disaggregate Bass model results are over twice the average number of owner-occupied homes in a zip code, suggesting that we may have reason to be concerned about the zip code-level Bass model results.[21]

While the Bass model is a useful benchmark, it may be more suitable to compare our approach to recent work in the marketing literature using disaggregated data. Bell and Song (2007) use a clever approach to identify social interactions in the adoption of online grocery retailing. They assume i.i.d. errors and use order statistics to derive an expression for the probability that at least one individual within a region makes use of the new technology. The authors estimate the model using maximum likelihood estimation. One concern about their approach is the reliance on i.i.d. errors across individuals, an assumption that would not hold if there is are shocks that affect many individuals in a particular area. An advantage of this method is that only information on the first adoption in each market is necessary. However, this also limits the number of control variables; only two-digit zip code fixed effects can be used. This limitation is recognized by the authors, who also estimate a mixed proportional hazards model, with a zip code random effect.

[21]The likely reason for this is that the acceleration of adoption is attributed entirely to the coefficient of contagion, when in fact other macro shocks not accounted for in the Bass model help explain some of it.

In this spirit, we estimate a hazard model with zip-quarter random effects and monthly indicators. This follows Iyengar et al. (2011), who use city and monthly indicators. The use of time controls is more flexible than a mixed proportional hazards model, since it allows the hazard rate to vary as a flexible function of time rather than a predetermined parametric function of time. Using a Poisson model with the number of daily installations as the dependent variable, we find that adding an installation to the installed base increases the probability of a household installation by 0.0109 (standard error of 0.0001). For an average zip code with 4,959 owner-occupied homes, this corresponds to a household effect of 2.198 x 10^{-6}. This is quite close to our first-differenced results. One concern however is the assumption of orthogonality between the random effects and covariates.

We can include fixed effects rather than random effects in a model using the generalized method of moments (GMM) estimator described in Woolridge (1997) and Windmeijer (2000). This method is computationally intensive and again assumes i.i.d. errors. We estimate the model using monthly aggregated data, where the dependent variable is the probability of a household adoption (in millionths).[22] The resulting estimated coefficient is insignificant, possibly due to the computational necessity of monthly aggregation.

6 Conclusions and Implications for Practitioners

This paper provides a straightforward methodology for consistently estimating peer effects in the adoption of new technologies that display a lag between the time of adoption

[22]Details can be found in an online appendix.

and delivery, and then quantifies the magnitude of the peer effects. We find strong evidence for causal peer effects, indicating that an extra installation in a zip code increases the probability of an adoption in the zip code by 0.87 percentage points when evaluated at the average number of owner-occupied homes in a zip code. These peer effects appear to be increasing in magnitude over time and are greater for larger installations and at the more localized street level. Our findings also indicate that peer effects appear to lead to larger installations. The results are suggestive of the pathways by which peer effects work: both visibility of the panels and word-of-mouth appear to contribute to social interactions that lead to further adoption. Accordingly, both image motivation and information transfer are likely to underlie the peer effects. These findings have clear implications for marketers who are striving to reduce the high cost of consumer acquisition in the solar PV market.

A first implication is that since visibility appears to enhance the peer effect, increasing the visibility of adoptions would be expected to increase the rate of adoption. Indeed, this strategy can be seen with several installers putting up signs indicating that a solar PV panel has been installed. Our finding that a larger installed base in a zip code leads to larger installations suggests that other methods of information provision may also lead to increased adoption levels. The use of demonstration sites has been shown to have positive effects on the adoption of green technologies (Bollinger 2012), although Kalish and Lilien (1986) caution us that such demonstrations for solar PV should only be used when the information to be learned is positive. Programs such as PG&E's "Neighborhood Solar

Champions" training program aim to leverage peer effects to provide such positive information to neighbors. Our finding of an increasing effect of new installations in a zip code suggests that targeting marketing efforts in areas that already have some installations is a promising strategy. Our demographic interaction results further suggest that efforts to leverage peer effects can be targeted to particular areas based on demographics, such as those with larger household sizes and many commuters.

The results in this paper are consistent with the latest developments in the solar PV market. Companies such as SolarCity and Sungevity have recognized that reducing the consumer uncertainty about installing solar is critical to expanding the market. These companies lease solar panels to consumers; they perform the installation for free and take care of all the maintenance, and they then guarantee the system will perform as promised or they will pay the consumer the difference. By transferring the risk of installation to the installer, this may reduce moral hazard issues and lead to the installation of larger (and riskier) installations.

Many of the results in this paper are also likely to apply to the diffusion of other visible green technologies, such as hybrid vehicles, electric vehicles, geothermal heating, and outdoor high efficiency lighting. A decision support system for using this "analogue approach" to forecast diffusion is described in Choffray and Lilien (1986). We see many promising opportunities to further explore the influence of social interactions in other solar PV markets and other markets for green products.

References

Amiti, M. and Wei, S.-J. (2006), 'Service offshoring and productivity: Evidence from the united states', *NBER Working Paper No. 11926* .

Arzaghi, M. and Henderson, J. V. (2007), 'Networking off Madison Avenue', *Review of Economic Studies* **75**(4), 1011–1038.

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.

Bollinger, B. (2012), 'Technology adoption in regulated markets: An empirical study of the southern california dry cleaning industry', *working paper* .

Brock, W. and Durlaf, S. (2001), *Handbook of Econometrics*, Vol. 5, Elsevier, chapter 4, Interaction-based Models, pp. 3297–3380.

Choffray, J. M. and Lilien, G. L. (1986), 'A decision-support system for evaluating sales prospects and launch strategies for new products', *Industrial Marketing Management* **15**(1), 75–85.

Choi, J., Huiand, S. K. and Bell, 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* .

De Giorgi, G., Pellizzari, M. and Redaelli, S. (2010), 'Identification of social interactions through partially overlapping peer groups', *American Economic Journal: Applied Economics* **2(2)**.

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.

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**, 287–304.

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.

Kalish, S. and Lilien, G. L. (1986), 'A market entry timing model for new technologies', *Management Science* **32**(2), 194–205.

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**, 646–659.

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), 961–976.

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

McShane, B., Bradlow, E. T. and Berger, J. (2012), 'Visual influence and social groups', *forthcoming at Journal of Marketing Research* .

Moffitt, R. (2001), *Social Dynamics*, MIT Press, chapter 3, Policy Interventions, Low-level Equilibria, and Social Interactions, pp. 45–82.

Munshi, K. and Myaux, J. (2006), 'Social norms and the fertility transition', *Journal of Developmental Economics* **80**(1), 1–38.

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**, 883–895.

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), 690–700.

Narayanan, S. and Nair, H. (2012), 'Estimating causal installed-base effects: A bias-correction approach', *forthcoming at Journal of Marketing Research* .

Nerlove, M. (1971), 'Further evidence on the estimation of dynamic economic relations from a time series of cross sections', *Econometrica* **39**(2), 359–387.

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

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

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), 929–945.

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.

Topa, G. (2001), 'Social interactions, local spillovers and unemployment', *Review of Economic Studies* **68**(2), 261–295.

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.

A Proof of Proposition 1

This appendix contains a proof of Proposition 1. To begin, recall our definition of the installed base: $b_{zt} = \sum_{\tau=1}^{t} \sum_{j=1}^{m_{zt}} \tilde{a}_{j\tau}$, where $\tilde{a}_{jt}$ is an indicator for a *completed* installation by household j at time t. In contrast, Y_{zt} is based on the fraction of households who *decide to adopt* a solar PV system in zip code z at time t, with the time lag between an adoption and installation of l.

A necessary and sufficient condition for consistency and unbiasedness is that the vector of covariates is orthogonal to the error term. In particular, we are most concerned that the installed base in the first-differenced model, $b_{zt} - b_{zt-1}$, is orthogonal to the first-differenced error term. We then have that

$$\begin{aligned}\mathbb{E}[(b_{zt} - b_{zt-1})(\epsilon_{zt} - \epsilon_{zt-1})] &= \mathbb{E}[(\sum_{\tau=1}^{t}\sum_{j=1}^{m_z} \tilde{a}_{j\tau} - \sum_{\tau=1}^{t-1}\sum_{j=1}^{m_z} \tilde{a}_{j\tau})(\epsilon_{zt} - \epsilon_{zt-1})] \\ &= \mathbb{E}[\epsilon_{zt}\sum_{\tau=1}^{t}\sum_{j=1}^{m_z} \tilde{a}_{j\tau}] - \mathbb{E}[\epsilon_{zt-1}\sum_{\tau=1}^{t}\sum_{j=1}^{m_z} \tilde{a}_{j\tau}] \\ &\quad - \mathbb{E}[\epsilon_{zt}\sum_{\tau=1}^{t-1}\sum_{j=1}^{m_z} \tilde{a}_{j\tau}] + \mathbb{E}[\epsilon_{zt-1}\sum_{\tau=1}^{t-1}\sum_{j=1}^{m_z} \tilde{a}_{j\tau}] \\ &= 0.\end{aligned}$$

To see why the final equality holds, it suffices to focus on $\mathbb{E}[\epsilon_{zt-1} \sum_{\tau=1}^{t} \sum_{j=1}^{m_z} \tilde{a}_{j\tau}]$, for if this term equals zero, then the others also equal zero. When $l \leq \nu + 1$, then $\mathbb{E}[\epsilon_{zt-1}\epsilon_{zt-\nu}] = \mathbb{E}[\epsilon_{zt}\epsilon_{zt-\nu+1}] \neq 0$ so by definition of Y_{zt}, $\mathbb{E}[\epsilon_{zt-1} \sum_{j=1}^{m_{zt}} \tilde{a}_{jt}] \neq 0$. However, with $l > \nu + 1$, $\mathbb{E}[\epsilon_{zt-1} \sum_{j=1}^{m_z} \tilde{a}_{jt}] = 0$.

B Details of Street-level Analysis

As with the zip code model, 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 a household adopting solar on street s in zip code z in month μ be given by:

$$Y_{sz\mu} = \frac{1}{m_s}\beta_s b_{s\mu} + \beta_z b_{z\mu} + \gamma' X_{z\mu} + \eta_{z\mu} + \xi_\mu + \epsilon_{sz\mu},$$

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

$$m_s \cdot Y_{sz\mu} = \beta_s b_{s\mu} + m_s \cdot \beta_z b_{z\mu} + \gamma' m_s \cdot X_{z\mu} + m_s \cdot \eta_{z\mu} + m_s \cdot \xi_\mu + m_s \cdot \epsilon_{sz\mu}.$$

Now $\mathbb{E}(N_{sz\mu}) = m_s \cdot Y_{sz\mu}$, where $N_{sz\mu}$ is the number of installations on street s and zip z in month μ. Therefore, we can write:

$$N_{sz\mu} = \beta_s b_{s\mu} + m_s \cdot \beta_z b_{z\mu} + \gamma' m_s \cdot X_{z\mu} + m_s \cdot \eta_{z\mu} + m_s \cdot \xi_\mu + \varepsilon_{sz\mu}.$$

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, β_s. 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.

C Details of Bass Model Estimation

In the Bass model, the probability of adoption at time t (given that adoption has not occurred) is modeled as:

$$Y_{zt} = (\sum_{\tau=1}^{t} \sum_{j=1}^{m_z} \tilde{a}_{j\tau})/(m_z - b_{zt}) = p + (q/m_z) b_{zt},$$

where p and q are the coefficients of innovation and imitation, respectively. Recall that m_z is the total market potential and z denotes the market (e.g., all of California or a zip code). The first order condition shows that the number of new adoptions is given by a quadratic function of b_{zt}:

$$\sum_{j=1}^{m_{zt}} \tilde{a}_{jt} = [m_{zt} - b_{zt}][p + (q/m_{zt}) b_{zt}] = p m_{zt} + (q - p) b_{zt} - (q/m_{zt}) b_{zt}^2. \quad (7)$$

The common approach, which we take, is to run a linear regression estimating coefficients equal to the expressions pm_{zt}, $q - p$, and $-q/m_{zt}$. We define t in our estimation as a month (i.e., μ as in the street-level estimation), but it is sometimes defined as a year in the previous literature.

Tables

Table 1: Zip code-level summary statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Zip code number of residential installations	51.481	78.262	1	713	1652
Zip code MW of residential installations	0.27	0.44	0	3.783	1652
population (100,000s)	0.244	0.216	0	1.095	1290
household size	2.829	0.611	0	5.21	1290
median income	6.34	2.896	0	37.5	1290
% pop male	50.237	3.231	34.4	97.8	1290
% pop who are white	65.205	20.244	4.4	95.2	1290
% pop with college degrees	38.124	17.57	4.115	95.731	854
% pop between 20 and 45	33.261	7.885	3.9	79.600	1290
% pop over 65	12.375	6.156	0	80.900	1290
% pop who drive to work	86.218	10.36	4.348	100	1304
% pop who carpool	14.691	6.496	0.469	55.875	1280
% pop using public transit	3.92	5.711	0.058	42.593	1026
% pop who work at home or walk to work	8.819	6.881	1.617	61.496	1207
% pop with over a 30 min commute	38.216	12.73	5.371	80.881	1117
% pop who drive a hybrid	3.596	5.761	0	100	1354
number of owner occupied homes (1000s)	4.959	4.204	0	18.965	1290
median value owner occupied home	0.534	0.26	0	1	1290
home loan	121.047	67.833	0	576	1290
home repair	122.663	70.379	0	585	1290
fraction of homes worth 0-50K	2.59	3.715	0	53.2	1290
fraction of homes worth 50-90K	2.04	2.686	0	37.3	1290
fraction of homes worth 90-175K	5.893	7.715	0	61.7	1290
fraction of homes worth 175-400K	30.623	23.231	0	89.7	1290
fraction of homes worth 400K+	58.543	29.836	0	100	1290

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

Table 2: Zip code-day summary statistics

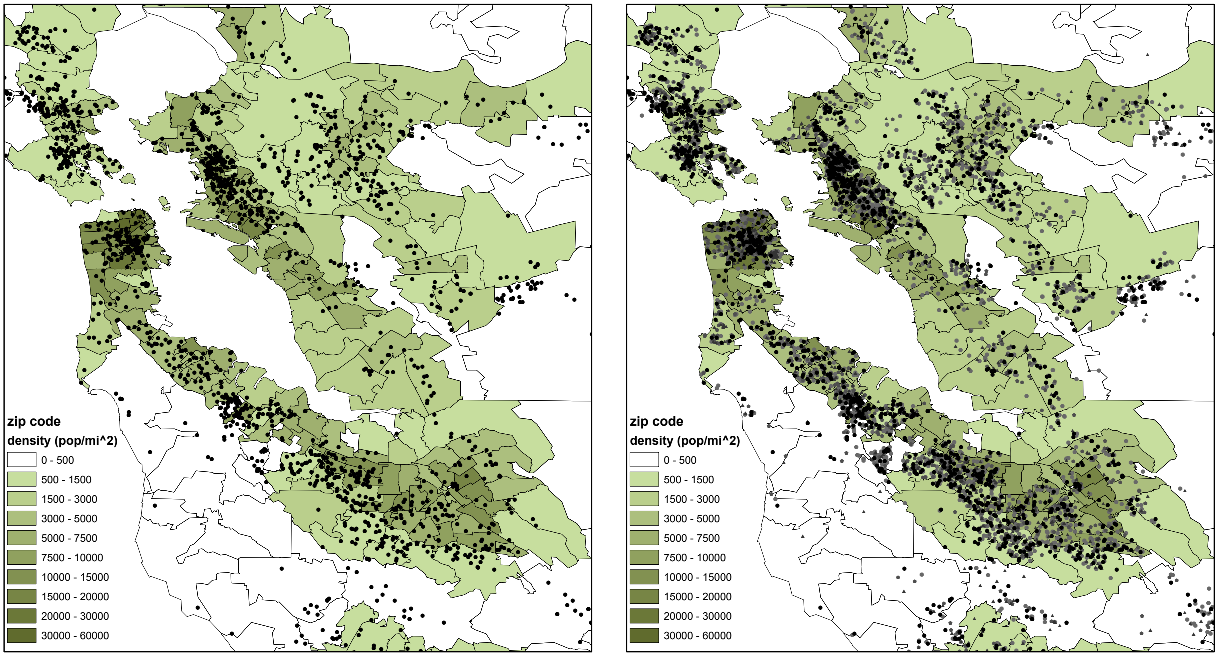
Variable	Mean	Std. Dev.	Min.	Max.	N
Household prob. of adoption (in 10^{-6})	4.39	276.49	0	333,333.34	6,102,465
Zip code installed base (count)	13.10	31.07	0	657	6,102,465
Zip code installed base (MW)	61.83	158.26	0	3,332.06	6,102,465

Table 3: Zip code model (obs=zip-day)

	OLS	**FE**	**RE**	**FD**
installed base	0.049	0.126	0.049	1.567
	(0.003)	(0.099)	(0.004)	(0.376)
quarter-zip effects	N	Y	Y	Y
year-month indicators	Y	Y	Y	Y
day of week indicators	Y	Y	Y	Y
day of month indicators	Y	Y	Y	Y
incentive step indicators	Y	Y	Y	Y
R-squared	0.000	0.013	0.000	0.000
N	6,102,465	6,102,465	6,102,465	6,034,360

Robust standard errors in parentheses.

Figures



(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

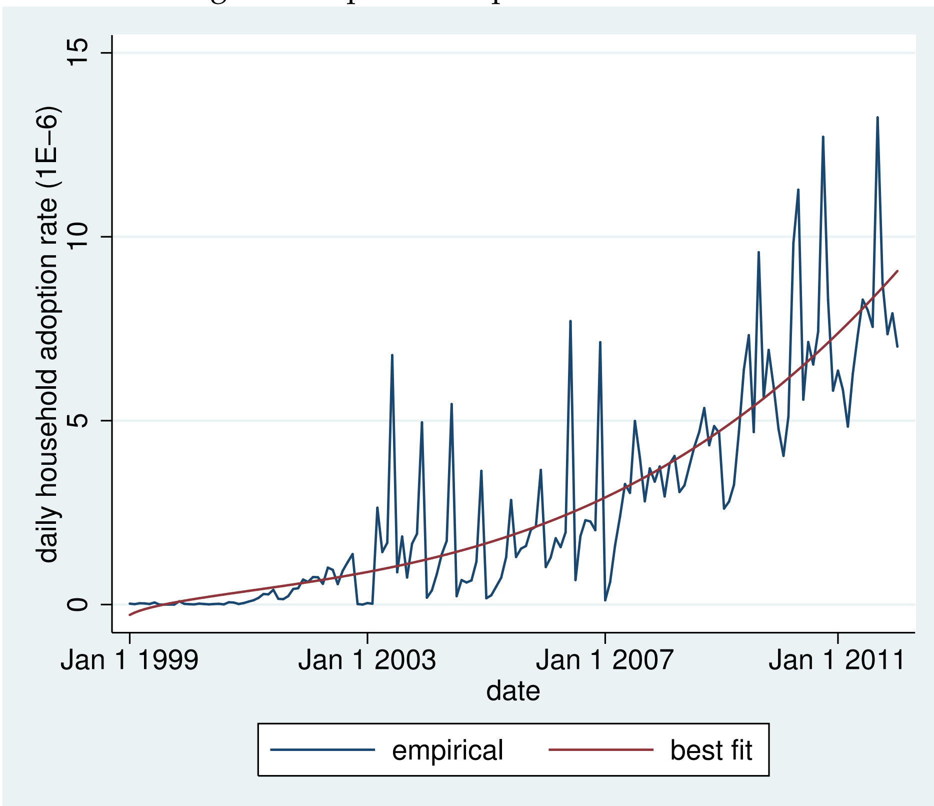


Figure 3: Non-zero household probability of adoption

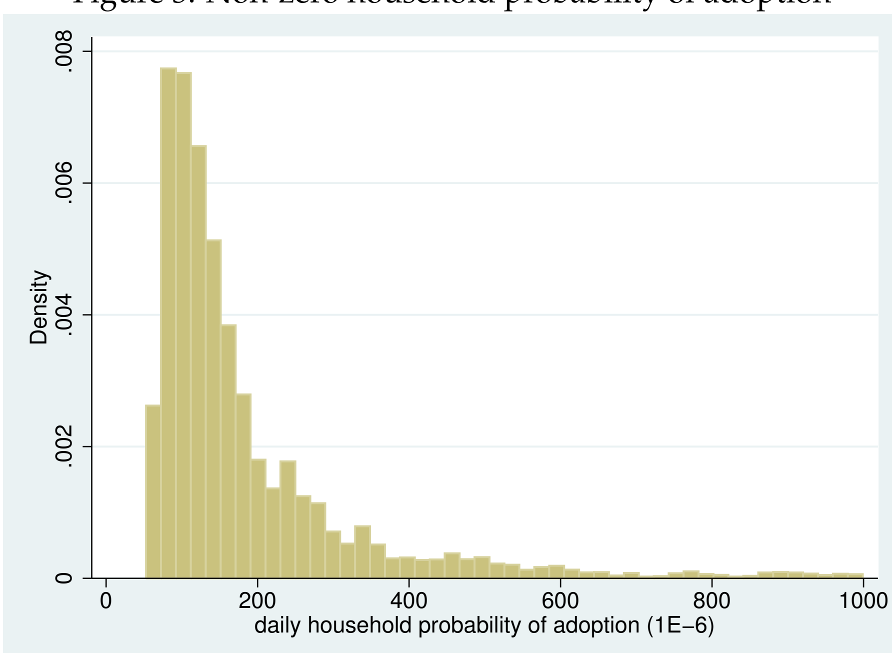


Table 4: Adoption model with demographics (obs=zip-day)

	OLS	**FD**
installed base	0.002	-5.526
population (100,000s)	-1.675***	
household size	-0.072	
median income	-0.658***	
% pop male	0.226***	
% pop who are white	0.010***	
% pop with college degrees	0.025*	
% pop between 20 and 45	-0.135***	
% pop over 65	-0.168***	
% pop who drive to work	-0.061	
% pop who carpool	-0.017	
% pop using public transit	-0.083	
% pop who work at home or walk to work	0.046	
% pop with over a 30 min commute	0.021***	
% pop who drive a hybrid	0.484***	
home loan	-0.021	
home repair	0.039**	
fraction of homes worth 90-175K	-0.069***	
fraction of homes worth 175-400K	-0.038***	
fraction of homes worth 400K+	-0.031**	
population x zip code installed base	-0.202***	-1.810
household size x zip code installed base	0.061***	3.448*
med income x zip code installed base	0.006*	-2.521*
% pop male x zip code installed base	-0.000	0.232
% white x zip code installed base	-0.000	-0.004
% college x zip code installed base	-0.001***	0.062
% pop between 20 and 45 x zip code installed base	0.003***	-0.002
% over 65 x zip code installed base	0.004***	-0.101
% drive x zip code installed base	-0.001*	0.042
% carpooling x zip code installed base	-0.002***	-0.602**
% public transit x zip code installed base	-0.001	0.070
% work at home or walk x zip code installed base	0.002*	0.233
% pop with over a 30 min commute x zip code installed base	0.000**	0.087*
% driving hybrids x zip code installed base	-0.006***	0.430
median home value x zip code installed base	0.009	3.109
home loan x zip code installed base	0.001***	0.147
home repair x zip code installed base	-0.001***	-0.082
fraction of homes worth 90-175K x zip code installed base	-0.008	-0.168
fraction of homes worth 175-400K x zip code installed base	0.022	-0.058
fraction of homes worth 400K+ x zip code installed base	0.039***	-0.110
R-squared	0.008	0.001
N	3,817,392	3,775,584

***=significant at the 1% confidence level, **=5%, *=10%.

Table 5: First-differenced model with different measures of installed base (obs=zip-day)

	counts	**MW**	**counts per household**	**MW per household**
installed base (count)	1.567	0.272	0.125	0.018
	(0.376)	(0.062)	(0.089)	(0.013)
R-squared	0.000	0.000	0.000	0.000
N	6,034,360	6,034,360	6,034,360	6,034,360

Robust standard errors in parentheses.

Table 6: Residential installation size and price summary statistics

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
size (kW)	5.241	3.272	0.114	49.92	85,046
price ($/W)	8.494	4.16	0.279	697.333	47,029

Table 7: Installation size regressions

	OLS	**FE**	**FE**
installed base	1.6722	1.7406	2.5641
	(0.3480)	(0.5184)	(9.6849)
price	-0.0738	-0.0606	-0.0517
	(0.0397)	(0.0318)	(0.0246)
zip FE	N	Y	N
quarter-zip FE	N	N	Y
year-month indicators	Y	Y	N
day of week indicators	Y	Y	Y
day of month indicators	Y	Y	Y
incentive step indicators	Y	Y	Y
R-squared	0.055	0.199	0.377
N	47,028	46,754	38,930

Robust standard errors in parentheses.

Table 8: 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 9: First-Differenced Street-Level Model (obs=zip-month)

	FD 1	FD 2
number of previous installations on street	0.127	0.151
	(0.011)	(0.012)
zip code installed base (100s)	0.074	street-specific
	(0.046)	
street-quarter effects	Y	Y
month of year indicators	Y	Y
street indicators interacted with month of year indicators	Y	Y
street indicators interacted with zip installed base	N	Y
R-squared	0.275	0.327
N	7,585	7,585

Robust standard errors in parentheses.

Table 10: Regressions for incentive step transitions

	PG&E/SCE step 5 to 6	**SCE/SDG&E step 2 to 3**
S (higher incentive utility)	2976.480	449.341
	(1910.597)	(253.365)
Period B (period of shock)	-7.916	0.904
	(5.470)	(0.292)
Period C (realigned incentives)	-15.964	1.254
	(6.333)	(0.752)
S x Period B	5.790	6.604
	(5.882)	(4.130)
S x Period C	18.725	9.581
	(7.638)	(5.685)
zip code FE	Y	Y
zip indicators interacted with time trend	Y	Y
R-squared	0.011	0.008
N	14,100	55,577

Robust standard errors in parentheses.

Figure 4: Zip code installed base

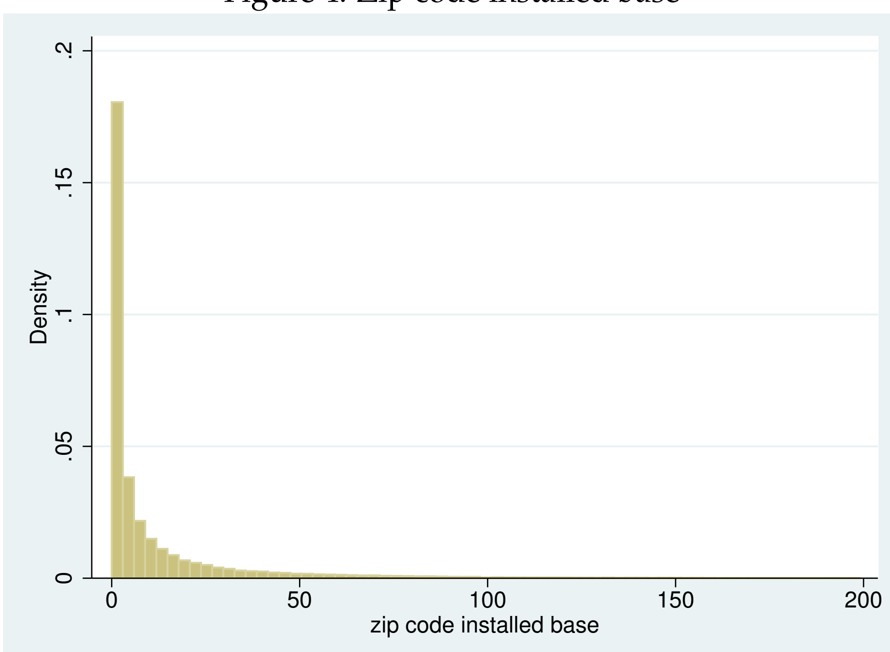


Table 11: Bass model estimation results

	Aggregate	**Zip no FE**	**Zip with FE**
constant (pm_{zt})	2.275	1.158 x 10^{-3}	1.771 x $10^{-3\dagger}$
	(0.433)	(0.046 x 10^{-3})	
cumulative installations ($q-p$)	1.089 x 10^{-3}	1.055 x 10^{-3}	1.002 x 10^{-3}
	(0.063 x 10^{-3})	(0.008 x 10^{-3})	(0.010 x 10^{-3})
cumulative installations squared ($-q/m_{zt}$)	-4.51 x 10^{-9}	-1.41 x 10^{-6}	-1.48 x 10^{-6}
	(1.21 x 10^{-9})	(0.051 x 10^{-6})	(0.0562 x 10^{-6})
zip code FE	N	N	Y
R-squared	0.745	0.260	0.299
N	147	42,690	42,696
p	6.02 x 10^{-7}	9.74 x 10^{-8}	1.60 x 10^{-7}
q	0.0170	0.0167	0.0163
m_{zt}	3.78 million	11,890	11,030

Robust standard errors in parentheses.
† Denotes the average value of the zip code fixed effect.

Figure 5: Estimated installed base coefficients by year, along with the 95% confidence intervals

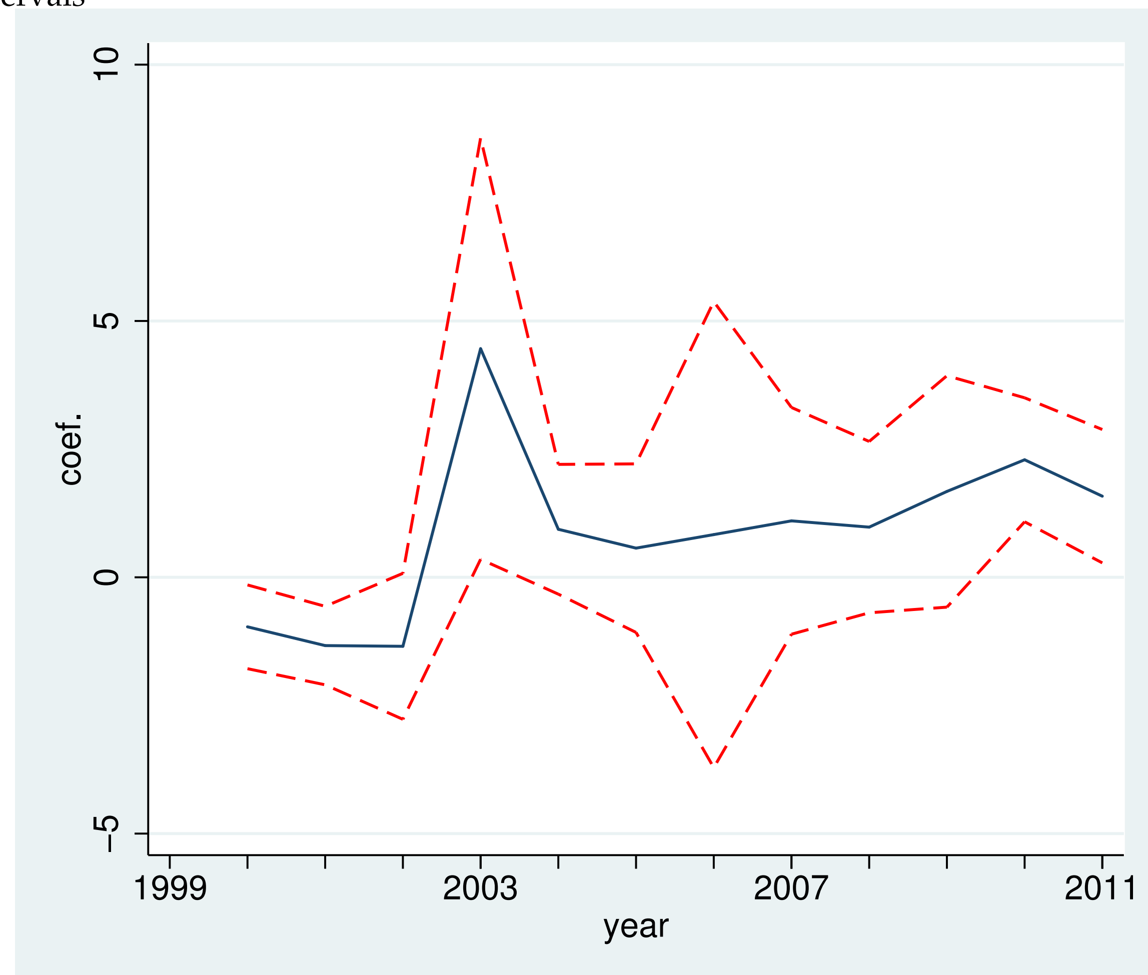


Figure 6: The difference-in-differences approach exploits changing incentives within the same zip code