UNIVERSITY OF CAMBRIDGE | Energy Policy Research Group

Social Effects in the Diffusion of Solar Photovoltaic Technology in the UK

EPRG Working Paper 1332

Cambridge Working Paper in Economics 1357

Laura-Lucia Richter

Abstract The main research question in this paper is whether the installation rate of solar PV technology is affected by social spillovers from spatially close households. The installed base, defined as the cumulative number of solar PV installations within a neighbourhood by the end of a particular month, serves as a measure for the social effects of interest. Motivated by the technology-specific time lag between the decision to adopt a solar PV panel and the completion of the installation, the third lag of the installed base serves as main regressor of interest in the panel data model employed. The results suggest small, but positive and significant social effects that can be exploited to promote adoption: at the average installation rate of 0.7 installations per 1,000 owner-occupied households, one more solar PV panel in the postcode district increases the installation rate three months later by one percent. At the average number of 6,629 owner-occupied households within a postcode district, this implies an increase in the number of new installations in the neighbourhood by 0.005. Projects involving a high number of installations could hence promote diffusion. A major limitation of the model is that social spillovers are assumed to spread within defined neighbourhoods, only. Spatial econometric methods could allow for social effects across these borders.

Keywords social effects, installed base, product adoption, diffusion, solar PV technology, micro-generation

JEL Classification C19, D12, D83, Q21, Q42

Contact llr23@cam.ac.uk
Publication December 2013
Financial Support DAAD – German Academic Exchange Service

www.eprg.group.cam.ac.uk

EPRG WORKING PAPER

Social Effects in the Diffusion of Solar Photovoltaic Technology in the UK*

Laura-Lucia Richter†
University of Cambridge

January 9, 2014

Abstract

The main research question in this paper is whether the installation rate of solar PV technology is affected by social spillovers from spatially close households. The installed base, defined as the cumulative number of solar PV installations within a neighbourhood by the end of a particular month, serves as a measure for the social effects of interest. Motivated by the technology-specific time lag between the decision to adopt a solar PV panel and the completion of the installation, the third lag of the installed base serves as main regressor of interest in the panel data model employed. The results suggest small, but positive and significant social effects that can be exploited to promote adoption: at the average installation rate of 7 installations per 10,000 owner-occupied households, one more solar PV panel in the postcode district increases the installation rate three months later by one percent. At the average number of 6,629 owner-occupied households within a postcode district, this implies an increase in the number of new installations in the neighbourhood by 0.005. Projects involving a high number of installations could hence promote diffusion. A major limitation of the model is that social spillovers are assumed to spread within defined neighbourhoods, only. Spatial econometric methods could allow for social effects across these borders.

*My most profound thanks to Dr. Michael Pollitt and Dr. Melvyn Weeks for their support and consult as supervisors of this paper. I am grateful for valuable suggestions of Professor Oliver Linton and Dr. Alexei Onatski in context of the PhD Econometrics Workshop. And I thank Alexander Ross, Julien Gagnon and seminar participants of the University of Cambridge for their helpful suggestions and comments.

†PhD Candidate, Faculty of Economics, University of Cambridge, UK. *Email*: llr23@cam.ac.uk.

Contents

List of Abbreviations

cdf	Cumulative distribution function
CMT	Continuous Mapping Theorem
DECC	Department of Energy and Climate Change
FD	First Difference
FE	Fixed Effect
FiT	Feed-in-Tariff
HH	Household
IEA	International Energy Agency
IMD	Index of Multiple Deprivation
LA	Local Authority
LSOA	Lower Level Super Output Area
GLS	Generalized Least Squares
NEED	National Energy Efficiency Database
OLS	Ordinary Least Squares
ONS	Office for National Statistics
OVB	Omitted Variable Bias
PCD	Postcode District
POLS	Pooled Ordinary Least Squares
PV	Photovoltaic
RE	Random Effects
ST	Slutsky Theorem
WG	Within-Group
WLLN	Weak Law of Large Numbers
zip	zip code (first four digits of a full UK postcode)

List of Figures

List of Tables

1 Introduction

The determinants of new technology adoption have been addressed in the economic literature for several decades. Among others, they matter for the design of policies to promote the expansion of the respective new markets. A recent example is the market for micro-generation technologies that can be installed by households, communities and small commercial sites. The installed capacity of those small-scale installations goes up to 50kW for electricity and 300kWth for heat generation (The Green Energy Act 2009). In a context of the EU target to increase the share of renewable electricity generation beyond 15 percent by 2020 and given the legally binding domestic energy policy goals to decrease national carbon emissions by 80 percent by 2050 compared to 1990, the UK government intends to encourage households to adopt micro-generation technologies and produce their own low-carbon electricity. But not only for policy design, also for economic and business reasons, the analysis of the diffusion of micro-generation technologies is particularly interesting: decentralised electricity generation has the potential to change the (energy) consumer - producer relationship, to alter the economic relations between customers and energy suppliers and to lead to new ownership and energy business models (Snape and Rynikiewicz, 2012, Watson and Devine-Wright, 2011).

So far, feed-in-tariffs (FiT) are the major instrument to promote adoption of small-scale electricity generation. In the UK they have been paid since April 2010 to mitigate the relatively high costs and uncertainties of solar PV, wind, hydro and anaerobic digestion technology. The FiT is paid for each kWh of electricity generated and the rate paid depends on the size of the system, the technology and the date of completion of the installation[1]. In addition, there is an export tariff that is paid if the micro-generated electricity is exported to the grid.

This paper focuses on solar PV technology. Firstly, because it a single technology rather than the choice between different technologies is of interest. Secondly, because solar PV technology is the most established micro-generation technology in the UK.[2] However, as the government's 2015 Micro-generation Strategy claims, financial incentives are not enough to guarantee sustained growth of micro-generation technologies. There are major non-financial barriers to be addressed (e.g. related to insurance and warranties or skills and knowledge). Besides those barriers, non-financial *drivers* of growth should be in focus and exploited in future policy and market strategy design towards a low-carbon decentralized electricity system. In particular, social effects from others might impact the adoption decision and hence drive diffusion. As suggested by Weber and Rode (2012), solar PV panel installations in a neighbourhood are visible for passers-by, reducing uncertainty surrounding this technology. Observational learning from spatially close households might thus lead to a correlation of adoption decisions within neighbourhoods. If the adoption behaviour of others significantly affects the adoption of solar PV technology, targeted interventions could serve as attention catching projects that, if placed in the key areas, could promote diffusion effectively and at lower cost than FiT.

The main research question in this paper therefore is, whether the installation rate of solar PV technology is affected by social effects as measured by the installed base in the immediate local environment. The installed base thereby refers to the cumulative number of

[1] The completion date is crucial for the determination of the FiT. The FiT valid at the installation date is paid for 20 years.

[2] By March 2013, 98.55 percent of all micro-generation installations were solar PV systems. Wind contributed only 1.22 percent and anaerobic digestion and hydro even smaller shares of 0.01 and 0.11 percent.

solar PV installations within a neighbourhood the end of a particular month and serves as a measure for social effects, such as observational learning, from spatially close households. The analysis is based on installation data that has been collected by Ofgem since the introduction of the FiT in April 2010. The main econometric panel data model specifies the postcode district-month as the smallest unit of observation. Identification of the suggested social effects within neighbourhoods is challenged by multiple endogeneity issues. To address these and to consistently estimate the installed base effect as a measure of social effects from spatially close households, the installation rate within a postcode district is modelled as being affected by the installed base three months prior. This is motivated by the technology-specific time lag between the *decision* to adopt a solar PV panel and the *completion* of the installation. Besides the lagged installed base and month dummies, the main panel equation includes time-varying fixed effects to account for potential self-selection into neighbourhoods and for correlated unobservable neighbourhood characteristics that are constant within a neighourhood but vary over time. A first difference estimation strategy then yields unbiased and consistent estimates. Further model specifications allow for a time-varying installed base effect and consider different lags of the installed base as well as different outcome variables and different geographical areas for robustness. In a last specification differences of the social effects across distinct groups of the population are analysed.

The paper contributes to previous literature in performing the first econometric analysis of the diffusion of solar PV technology within the UK. In particular, it delivers empirical evidence, in how far the adoption behaviour of others drives diffusion. The analysis is based on a remarkably recent and granular solar PV installation dataset of the UK. The results can be exploited for targeted marketing and resource allocations for the stimulation of future adoption. Nevertheless, the analysis has its limitations. Firstly, social effects are assumed to spread within neighbourhoods as defined by the neighbourhoods, only, while spillovers across neighbourhood borders are ignored. Spatial econometric methods, for example, could be employed to allow for more diverse spillover effects. However, it is a useful first highly disaggregated approach to explore the impact of social effects on solar PV adoption, that can be extended in future research. Another limitation is the aggregation to the neighbourhood level. Future research should make use of household level covariate data to further analyse the mechanisms underlying the adoption behaviour. In context of the National Energy Efficiency Database (NEED) the Department for Energy and Climate Change (DECC) is currently creating a database that matches the solar PV installation data with household characteristics.

This paper is organised as follows. Section 2 reviews literature on solar PV technology diffusion, on social effects in technology adoption and on social effects in the diffusion of solar PV technology more specifically. Moreover, the installed base approach and its previous applications and challenges are introduced. Section 3 specifies the econometric model and justifies the third lag of the installed base as measure for social effects in solar PV technology adoption. Section 4 presents endogeneity issues that can challenge the estimation of the social effects as measured by the installed base and highlights the identification strategy exploited in the paper. Section 5 introduces the estimation strategy and derives conditions for consistent estimation of the social effects. Section 6 presents the data. The main results and the results of several robustness checks are presented in section 7 and section 8, respectively. To get an idea to what extent social effects vary across postcode district characteristics, the impact of contextual factors on social effects is considered in section 9. Finally, section 10 points out the major limitations of the analysis and section 11 concludes.

2 Literature Review

2.1 Diffusion of Solar PV Technology

The determinants and dynamics of technology adoption and diffusion have been addressed in the economic literature for several decades. The diffusion of micro-generation technologies, solar photovoltaic (PV) technology more specifically, is of particular economic interest. Micro-generation technologies have the potential to change the energy consumer - producer relationship, to alter the economic relations between customers and energy suppliers and to lead to new ownership and energy business models (Snape and Rynikiewicz, 2012, Rynikiewicz, 2010, Watson and Devine-Wright, 2011).

So far, micro-generation technologies have been growing mainly in countries that provide financial incentives to encourage installation and to achieve the required economies of scale to compete with traditional electricity generation technologies. According to the International Energy Agency (IEA), in 2012 the five countries with the highest installed capacity of solar PV technology worldwide were Germany (32,411 MW), Italy (16,250 MW), USA (7,221 MW), Japan and China (both 7,000 MW). In all these countries, the governments provide financial incentives for adoption. As a matter of fact the existing literature on the diffusion of solar PV technology mainly examines the effect of subsidy policies and cost reductions and consistently finds a significant positive impact. Early studies tend to use aggregate data and to pursue cross-country comparisons. Beise (2004), for example, analyses whether PV technology diffuses across countries by exploiting a panel dataset covering 13 countries (1992-2002). He finds that government interventions have a significant positive impact on diffusion within countries. For sustainable international success of solar PV, however, he suggests that it would be essential to demonstrate that the adoption of PV systems is continuing without subsidies. Comparably, Guidolin and Mortarino (2010) provide insights on adoption dynamics of solar PV technology in various countries. They perform a cross-country analysis using panel data covering 11 countries over the period from 1992 through 2006. They find, too, that government policy incentives promote solar PV technology diffusion. Wüstenhagen (2006) addresses the impact of government policies on the renewable energy market in Germany and concludes that the high adoption rates of solar PV systems were primarily driven by the implementation of public policies, most significantly by the feed-in-tariff (FiT). Zhang *et al.* (2011) are among the first to use more highly disaggregated data: they analyse regional differences in solar PV technology diffusion using panel data on 47 Japanese prefectures from 1996 to 2006. Their results suggest that adoption rates are positively influenced by regional government policies and negatively influenced by installation costs. They particularly emphasize the importance of regional diffusion policies that reflect the residents′ degree of environmental awareness.

While cost reductions and subsidy policies have thus consistently been found to be critical for the diffusion of solar PV technology, research on the role of further, non-financial barriers (such as insurance and warranty issues) and non-financial drivers (such as social effects, supply-side customer service programmes or a structural awareness change) is still scarce. This paper focuses on the role of social effects such as observational learning from the adoption behaviour of others, as one potential further driver of adoption and diffusion of solar PV technology.

2.2 Social Effects in Technology Diffusion

The idea of social effects is consistent with classical models of technology diffusion. There are several reasons why agents might care about the adoption behaviour of others. Agents might have social preferences (i.e. others′ utility directly affects their utility), there may be word-of-mouth effects (i.e. feedback from others affects beliefs directly) or network effects (i.e. adoption by others affects the users′ valuations of the product) (e.g. Narayanan and Nair, 2011). Moreover, there might be observational learning from peers (i.e. learning upon the observation of others' decisions and outcomes as suggested by Snape and Rynikiewicz (2010)).

Due to limited data availability it is often hard to pin down the precise channel of social spillovers, but multiple economic models of social effects are consistent with the idea of observational learning. In particular, Rogers' traditional theory of 'Diffusion of Innovations' (1962) considers observability as one of five major influential factors of an individual′s decision to adopt. He argues that an innovation that is more visible will enhance communication among the individuals peers and personal networks and will in turn create positive or negative social multipliers (Rogers, 1962). According to this theory, segments of potential adopters have different attitudes towards the innovation and hence adopt at different points in time, which is consistent with the empirical fact that many diffusion processes over time can be described as an S-curve (increasing diffusion rate to start with, then decreasing diffusion rate due to satiation over time). In Rogers' model observational learning is of relatively minor importance for the rather risk-loving, early adopters, while it matters more for the technology adoption of the more risk-averse, late adopters.

Another example for a model of technology adoption that is consistent with observational learning is Rasul′s (2002) model of social learning. In contrast to previous models, in which social learning takes place only *after* the new technology has been adopted, Rasul′s model allows social learning to *lead* to a correlation of the initial adoption decisions within social networks. Rasul′s model is applicable to any situation in which a lack of information is a barrier to adoption and potential adopters can communicate with, or observe each other. In his empirical application, Rasul estimates farmers' propensity to adopt as a function of the number of adopters among family and friends. The results suggest that social network effects are increasing when the number of adopters is low, only to fall when this number reaches a certain level. On top of this, Rasul finds that social effects are heterogeneous. They depend on the farmers' characteristics and on the definition of the reference group. In particular, the social effects differ depending on whether the reference group is defined in a social, cultural or a geographical sense. Although the study does not distinguish between social learning and other causes of social effects (e.g. marketing externalities), it distinguishes between causal social effects and contextual effects as sources of spurious correlation. This distinction is important from a policy perspective: while social effects lead to a social multiplier, contextual effects do not.

In multiple models the existing stock of a technology within a reference group thus plays a crucial role for diffusion. The so-called 'installed base' of the technology in question is frequently used as a measure for social effects: the installed base refers the cumulative number of solar PV systems within a defined reference group at a specific point in time and arguably impacts the adoption behaviour of others within this group. This installed base approach, though frequently used in the existing literature, has recently been criticised by Narayanan and Nair (2011). They show that in the presence of endogenous group formation and correlated unobservables the traditional within estimator in combination with the first lag of the installed base leads to biased and inconsistent estimates of the installed base effect. They present two approaches to address the endogeneity: an instrumental variable (IV) approach and a bias correction approach.[3] They apply their theory to the adoption of the Toyota Prius Hybrid car in California using individual vehicle registrations from 2001 through 2007 and find that a one percent increase in the installed base of the Prius Hybrid in the zip code area of an individual increases the installation rate of a purchase significantly by 5.3 percent.

2.3 Social Effects in Solar PV Diffusion

As mentioned above, in context of micro-generation technologies, research addressing drivers of diffusion other than subsidies and cost-reductions is still scarce. This paper therefore focuses on social effects such as observational learning from the adoption behaviour of others, as one potential further driver of adoption and diffusion of solar PV technology. To the best of the author′s knowledge, two main studies have previously addressed this issue: Bollinger and Gillingham (2012) estimate peer effects in the diffusion of residential solar PV systems in California and Weber and Rode (2012) analyse information and knowledge spillovers in the diffusion of solar PV technology in Germany. Both papers use the installed base as a measure for social spillovers: Bollinger and Gillingham (2012) model the propensity of solar PV technology adoption in a Californian zip code as a function of the installed base in that zip code, i.e. the number of panels that have already been installed. They work with a sample of 80,046 residential solar PV installations that were requested in the period from 2001 to 2011 within three investor-owned utility regions in California. Within this framework, their identifying assumption is that social interactions do not have an effect until the solar PV installations have actually been completed. I.e. the time lag between adoption decision and the beginning of the social spillovers is crucial in their model. They find that peer effects are increasing in magnitude over time, are greater for larger installations and on a more localized level. They infer that visibility of the panels and word-of-mouth contribute to social interactions and lead to further adoptions. Their results suggest that at the average number of owner-occupied homes in a zip code, a marginal solar power installation increases the installation rate in that zip code by 0.78 percentage points. Weber and Rode (2011) also find support for the argument of social effects in the adoption of solar PV technology. They establish an epidemic diffusion model that includes a spatial dimension and test whether localized imitation drives the adoption of solar PV technology. They argue that solar PV systems are easily visible to passers-by and that learning is possible without direct social interaction. As such, they assume information flows between spatially close neighbours and find that imitative adoption behaviour is highly localized and plays a significant role in the diffusion of solar PV technology.

Even if previous studies suggest the existence of social effects in solar PV technology adoption, further empirical evidence is needed. In particular, analyses should be based on more granular data and should be pursued for different countries.

[3]They solve for the bias as a function of the variance of the errors and the variance of the installed base, but a disadvantage of their analysis is that the bias adjustment requires no autocorrelation in the errors.

3 Econometric Model

Consider the following linear dynamic panel model to analyse the impact of social effects on the adoption and diffusion of solar PV installation.

$$y_{zt} = \alpha_t + \beta \cdot b_{zt-3} + \underbrace{\alpha_{zq} + \epsilon_{zt}}_{u_{ztq}} \tag{1}$$

$$\text{where} \quad b_{zt-3} = \sum_{\tau=1}^{t-3} Y_{z\tau} \forall z, t \tag{2}$$

The model features three dimensions: firstly, there are the neighbourhoods z in the cross-section dimension. These are geographically defined neighbourhoods, as solar PV systems are visible to passers-by and learning is possible for anyone spatially close. By defining the reference groups in a geographical sense rather than on the basis of family or friends, the social effects identified are consistent with the idea of observing solar PV installations of spatially close households. Secondly, there are the months t in a first time-dimension and thirdly the quarters q in a second time-dimension. The months $t = 1, 2, 3$ pertain to quarter $q = 1$, the months $t = 4, 5, 6$ to $q = 2$ and so forth.

The outcome variable y_{zt} is an aggregation of individual choices. It is defined as the number of new solar PV installations Y_{zt} per owner-occupied household within a neighbourhood z in month t, $y_{zt} = \frac{Y_{zt}}{N_{zt}}$ and measures the degree of adoption and diffusion of solar PV technology. This installation rate as a measure of diffusion is preferred over a count variable Y_{zt}, as the number and the tenure type of households within a neighbourhood might vary considerably across neighbourhoods and a positive correlation between the number of owner-occupied households and the number of installations within a neighbourhood is likely. Dividing the count variable Y_{zt} by the respective number of owner-occupied households in the neighbourhood, N_{zt}, can control for this variation. y_{zt} is hence fractional, $y_{zt} \in [0, 1]$ and in the given application of solar PV technology likely to be very small.

For the normalisation the number of *owner-occupied* households is preferred over the total number of households in the neighbourhood, as the decision to install solar PV is likely to be made by households that own their property rather than by households that rent it. Even if tenants can benefit from the installed solar PV panels as well (e.g. they get some electricity for free), it is the landlord who receives the FiT and who decides whether to install a panel.[4]

Let Y_{zt} count the number of new installations in neighbourhood z in month t. Equation 2 defines the third lag of the installed base, b_{zt-3}, as the cumulative number of solar PV installations within a neighbourhood z by the end of time period $t-3$. The (lagged) installed base of domestic solar PV panels within a neighbourhood is used as a measure of social effects within this neighbourhood: installed solar PV systems are visible for passers-by and reduce uncertainty regarding the technology. Via the installed base, social effects such as observational learning can thus impact the adoption behaviour within the neighbourhood. The model uses the third lag of the installed base to capture the technology specific time lag between the *decision* to adopt a solar PV panel and the *completion* of the installation: according to quotas of the main solar PV suppliers in the UK, the lead time between the

[4]The number of tenants inducing their landlords to install solar PV is assumed negligible.

first contact with the supplier (which can be seen as a measure of a household′s decision to adopt) and the completion of the installation usually lies between two and three months. Hence, the inclusion of the third *lag* of the installed base, $b_{zt-3} = \sum_{\tau=1}^{t-3} Y_{zt}$, is justified. All adoptions of solar PV panels up to $t-3$, i.e. $Y_{zt-\tau} \forall \tau \geq 3$ thus enter the right hand side of the equation. The parameter of interest is β.

Apart from the installed base no further regressors are explicitly included in the equation. Instead, the right hand side of equation 1 includes three kinds of unobserved variables to capture the factors impacting the adoption rate within a neighbourhood, including in particular those characteristics that are correlated with the installed base. There are two types of fixed effects, α_t and α_{qz}, and there is an unobservable error term ϵ_{zt}:

The α_t are 35 month dummies[5], capturing month specific effects that are constant across neighbourhoods and could confound the estimation of the social effects. An example for such month-specific effects are nationwide policy announcements regarding the FiT. Historically, announcements to decrease the FiT have often been correlated with higher numbers of adoptions right before and with a lower number of adoptions after the effectivity date of the policy. Such events should be controlled for.

Compared to a standard panel model that includes time-constant fixed effects, α_z, to control for unobserved heterogeneity on the neighbourhood level, equation 1 is more general:[6] the model is specified with time-varying fixed effects. However, since the installed base b_{zt-3} is defined on the neighbourhood-month level, i.e. is by definition the same for all households within neighbourhood in a particular month, a specification with neighbourhood-month effects, α_{zt}, is not feasible - these would be perfectly collinear with the installed base and hence prohibit identification of the installed base effect β. Neighbourhood-quarter effects, α_{zq}, on the other hand, avoid this collinearity problem and allow to control for neighbourhood characteristics that are constant within a neighbourhood and a quarter, but vary over time, i.e. across quarters. Those fixed effects control for time-varying neighbourhood specific characteristics that are relevant for the installation rate in the neighbourhood. In particular they control for factors that are correlated with the lagged installed base in the neighbourhood *and* with the outcome variable y_{zt}. Those factors, if omitted from the equation, would lead to omitted variable bias of the installed base effect β. The neighbourhood-quarter effects could capture time and location-specific activities such as advertising or marketing campaigns. Changes of neighbourhood characteristics within a quarter are assumed to be negligible in the sense that existing variations do not significantly impact the decision to install solar PV. As an example, marginal changes in average income within a neighbourhood-quarter are assumed to be negligible for the installation rate, since the decision to invest in a solar PV panel is rather a question of accumulated capital than of marginally higher income.

Finally, ϵ_{zt} is an *i.i.d.* unobserved error term that captures random neighbourhood and month specific effects. In particular $E(b_{zt-3}\epsilon_{zt}) = 0$

[5]April 2010 being the reference category

[6]The model has also been specified with time-constant fixed-effects, α_z, only. The results are consistent with the theory of OVB and can be found in the Appendix.

4 Endogeneity and Identification

The econometric panel model relies on the third lag of the installed base, b_{zt-3}, to measure social effects within geographically defined neighbourhoods. The large number of fixed effects controls for observable and unobservable characteristics that could confound the estimation.

First of all, there might be self-selection into neighbourhoods. This problem has its origins on the household level, but implies non-randomness on the neighbourhood level as well: if factors driving households to live in a specific neighbourhood are also correlated with the adoption rate, y_{zt}, spurious correlation results. For example, energy conscious households might prefer to live in 'green', environmentally friendly neighbourhoods and this unobserved preference might also make them behave similarly regarding the adoption of solar PV technology. If so, an observed correlation of the installed base with the adoption decisions within the neighbourhood can be misleading. On the neighbourhood level this self-selection into neighbourhoods implies that c.p. a higher share of households with 'green' preferences lives in the same neighbourhood: pronounced 'green', environmental awareness within the neighbourhoods might increase the installed base *and* the adoption rate within the neighbourhood. Hence, some of the observed correlation between installed base and adoption rate is likely to be spurious, resulting from the correlation of unobserved tastes with the adoption rate rather than from social effects.

Secondly, unobservable neighbourhood characteristics, such as solar PV supplier activities or local advertising campaigns, can result in a correlation of adoption behaviour within the neighbourhoods and lead to spatial clustering, i.e. to spurious correlation of the installed-base with the adoption rate.

Manski′s (1993) well-known 'reflection problem' on the other hand is *not* a problem in the case of solar PV technology. The reflection problem refers to a phenomenon that frequently challenges the identification of social effects. It occurs if an individual′s adoption decision depends on the behaviour of others within her reference group and simultaneously impacts the adoption decisions of the others in that group. In case of solar PV technology adoption, this characteristic simultaneity is *not* a problem: the decision to adopt takes place on average three months *before* the completion of the solar PV system. Households installing solar PV panels in month t made their decision to adopt three months prior and it was the installed base in month $t-3$, b_{zt-3}, that affected their behaviour. The lag implies that a household′s decision to adopt solar PV technology might be impacted by the installed base in her neighbourhood, but social effects can only impact others once the panel is installed (i.e. approximately 3 months later).

Summing up, the technology specific time-lag between the *decision* to adopt a solar PV panel and the *completion* of the installation is crucial for identification and consistent estimation of the installed base effect. While the reflection problem is not an issue in the application of solar PV technology, self-selection and correlated unobservables are likely to be problematic and must be considered for identification and consistent estimation of the social effects.

5 Estimation Strategy

5.1 Pooled OLS Estimation

Consider equation 1. To consistently estimate the installed base effect, β, by pooled OLS (POLS), the following contemporaneous exogeneity assumption would need to hold:

$$E(b_{zt-3}u_{ztq}) = 0 \; \forall z, t, q \tag{3}$$

As ϵ is *i.i.d* by assumption, this implies in particular that:

$$E(b_{zt-3}\alpha_{zq}) = 0 \; \forall z, t, q \tag{4}$$

As equation 4 states, for consistent estimation all neighbourhood-quarter specific characteristics, α_{zq}, must be uncorrelated with the installed base. This condition is unlikely to hold: firstly, because the model omits all neighbourhood characteristics apart from the installed base and controls for them with the α_{zq}. There are almost surely neighbourhood characteristics that are correlated with both, the dependent variable and the included installed base - omitting them would lead to omitted variable bias (OVB). Secondly, there are diverse endogeneity concerns, such as self-selection and correlated unobservables, that would justify the inclusion of fixed effects even if all relevant observable neighbourhood characteristics were included. It is therefore likely that there exist observable and unobservable neighbourhood and neighbourhood time specific characteristics that are correlated with the installed base *and* the installation rate. This can be summarized as follows:

$$\exists \; \alpha_{zq} \; s.t. \; E(b_{zt-3}\alpha_{zq}) \neq 0 \cap E(y_{zt}\alpha_{zq}) \neq 0$$

If so, then the exogeneity condition in 3 breaks down and POLS results in biased and inconsistent estimates of the installed base effect, β.

5.2 Within-Group Estimator

Since equation 1 includes neighbourhood-quarter effects α_{z_iq}, mean differencing must be performed on the neighbourhood-quarter level to eliminate these effects and estimate the equation by a within-group estimator. More precisely, these are the means taken over all observations of α_t, y_{zt}, b_{zt-3} and ϵ_{zt} within reference group z and quarter q. Let $\overline{\alpha}_q$, $\overline{y}_{zq}$, $\overline{b}_{z3q}$ and $\overline{\epsilon}_{zq}$ be the neighbourhood-quarter means of the time-dummies, the installation rate, the lagged installed base and the idiosyncratic error, respectively. For example, $\overline{b}_{z3q}$ is the mean of the third lag of the installed base, taken over the three months that pertain to quarter q of month t. Mean differencing on the neighbourhood-quarter level then yields:

$$(y_{zt} - \overline{y}_{zq}) = (\alpha_t - \overline{\alpha}_q) + \beta \cdot (b_{zt-3} - \overline{b}_{z3q}) + (\epsilon_{zt} - \overline{\epsilon}_{zq}) \tag{5}$$

The resulting within-group estimator is given by:

$$\hat{\beta}_{WG} = \frac{\sum_z (b_{zt}\widetilde{_{-3} - \overline{b}}_{z3q})(y_{zt} - \overline{y}_{zq})}{\sum_z (b_{zt}\widetilde{_{-3} - \overline{b}}_{z3q})^2} \tag{6}$$

Where $\widetilde{(b_{zt-3} - \bar{b}_{z3q})}$ is the residual from a regression of the mean-differenced installed base on the mean-differenced time-dummies. Further, substituting in the equation for y_{zt} and rearranging yields:

$$(\hat{\beta}_{WG} - \beta) = \frac{\sum_z \widetilde{(b_{zt-3} - \bar{b}_{z3q})}(\epsilon_{zt} - \bar{\epsilon}_{zq})}{\sum_z \widetilde{(b_{zt-3} - \bar{b}_{z3q})}^2} \tag{7}$$

It follows by the Slutsky Theorem, the Continuous Mapping Theorem and the Weak Law of Large Numbers that:

$$\lim_{Z\to\infty} (\hat{\beta}_{WG} - \beta) = \frac{E[\widetilde{(b_{zt-3} - \bar{b}_{z3q})}(\epsilon_{zt} - \bar{\epsilon}_{zq})]}{E[\widetilde{(b_{zt-3} - \bar{b}_{z3q})}^2]} = \frac{A}{B} \tag{8}$$

Where Z refers to the number of neighbourhoods. Since the denominator B is non-zero by assumption, the nominator A must be considered carefully. In contrast to the case illustrated by Narayanan and Nair (2011) that is based on the first lag of the installed base, the neighbourhood-quarter mean of the installed base $\bar{b}_{z3q}$ includes future and past observations of adoptions within quarter $q-1$. On the contrary, the neighbourhood-quarter mean of the errors $\bar{\epsilon}_{zq}$ includes all future and past errors within quarter q.

$$\lim_{Z\to\infty} (\hat{\beta}_{WG} - \beta) = \frac{E[b_{zt-3}\epsilon_{zt}] - E[\bar{b}_{z3q}\epsilon_{zt}] - E[b_{zt-3}\bar{\epsilon}_{zq}] + E[\bar{b}_{z3q}\bar{\epsilon}_{zq}]}{E[(b_{zt-3} - \bar{b}_{z3q})^2]} = \frac{A}{B} = 0 \tag{9}$$

Multiplying out equation 8 yields equation 9, which allows to consider the terms in the nominator A separately. $\lim_{Z\to\infty} E[b_{zt-3}\epsilon_{zt}] = 0$ as b_{zt-3} is by construction correlated with all previous $\epsilon_{zt-l} \forall l \geq 3$: all previous installations $Y_{zt-l} \forall l \geq 3$ enter b_{zt-3}. However, b_{zt-3} is uncorrelated with all future errors, as ϵ_{zt} is *i.i.d* by assumption. Further, the second term, $\lim_{Z\to\infty} E[\bar{b}_{z3q}\epsilon_{zt}] = 0$. This is because the considered lag length is $l = 3$ and in all three months of quarter q the observations entering $\bar{b}_{z3q}$ thus lie in quarter $q-1$, while ϵ_{zt} lie in q. All observations of the installed base that enter $\bar{b}_{z3q}$ are uncorrelated with the error terms in quarter q. In month 4, for example (pertaining to the first month in the second quarter), $\bar{b}_{z3q}$ is calculated as average of the third lag of the installed base in month 4, 5 and 6. All these lags refer to the installed base in months 1, 2 and 3 respectively and hence lie in quarter 1. The third term converges to zero as well, i.e. $\lim_{Z\to\infty} E[b_{zt-3}\bar{\epsilon}_{zq}] = 0$. This term consists of the third lag of the installed base and the neighbourhood-quarter mean of the error. In all 3 months of a quarter q, b_{zt-3} lies in $q-1$ while $\bar{\epsilon}_{zq}$ is calculated based on errors in q and all these errors are assumed to be uncorrelated with prior observations of the installed base. Lastly, $\lim_{Z\to\infty} E[\bar{b}_{z3q}\bar{\epsilon}_{zq}] = 0$. This term refers to the correlation of the neighbourhood quarter mean of the lagged installed base with the neighbourhood quarter mean of the error term. In all 3 month of any quarter q, $\bar{b}_{z3q}$ is calculated based on observations in $q-1$.

Overall, if there is no autocorrelation across quarters, the mean differenced lagged installed base, $(b_{zt-3} - \bar{b}_{z3q})$, is uncorrelated with the mean differenced error, $(\epsilon_{zt} - \bar{\epsilon}_{zq})$, and A converges to zero.[7] More generally, as proven in the Appendix, for lags of a length larger than 2, i.e. exceeding the length of a quarter, within-estimators aiming to eliminate neighbourhood-quarter fixed effects, can yield consistent estimates.

[7] For the full proof of bias and inconsistency in case of the *first* lag of the installed base see Narayanan and Nair (2011). They assume no autocorrelation of the error terms and suggest that the bias is negative.

Proposition 1. *For lags of the installed base that exceed the length of a quarter, i.e. lags larger than 2, mean-differencing on the neighbourhood-quarter level eliminates the neighbourhood quarter effects and allows to consistently estimate the installed base effect by POLS on the mean-differenced equation.*

This contrasts with the model including the *first* lag of the installed base as considered by Narayanan and Nair (2011). They formalise that in the presence of self-selection and correlated unobservables, standard within-group estimators are biased and inconsistent, if there are installed base effects. In their model this is due to the correlation of the mean differenced installed base with the mean differenced errors, a problem relating closely to standard random or fixed effects estimators in dynamic panels. Due to a correlation of the transformed regressor(s) with the transformed errors, random effects GLS and fixed effects estimators (within-group estimators and first difference estimators) are biased and inconsistent. In dynamic panels, estimators as suggested by Anderson and Hsiao (1982) or by Arellano and Bond (1991) are required.

For the within-group estimator presented above, the time-lag of three months is essential for consistency. The estimator is not robust regarding the lag length of the installed base: for a lag length of two, $l = 2$, for example, the estimator is inconsistent due to a correlation of the mean-differenced installed base with the mean-differenced error. Hence, as consistency of the within-group estimator is based on relatively demanding conditions, such as a lag length of at least three months, this paper follows a first differencing estimation strategy as proposed by Bollinger and Gillingham (2012). For both estimation strategies, however, the natural time lag between adoption decision and installation is essential for identification and consistent estimation of the social effects. The following section illustrates that given the identification strategy above, the chosen first differencing strategy allows the consistent estimation of the installed base effect for any lag length as long as certain conditions regarding the autocorrelation ν of the error term are met.

5.3 First Difference Estimator

Comparable to the case of the within-group estimator, first differencing on the month level does not fully eliminate the neighbourhood-quarter effects. Rather, first differencing of the first month of a quarter yields a residual term of the size of the change of the neighbourhood-quarter effects from one quarter to the next. To eliminate this residual and hence the neighbourhood-quarter effects, the first month of each quarter is dropped after first differencing (see Bollinger and Gillingham, 2012). This leads to equation 10 and POLS on the differenced equation can follow.

$$(y_{zt} - y_{zt-1}) = (\alpha_t - \alpha_{t-1}) + \beta \cdot (b_{zt-3} - b_{zt-4}) + (\epsilon_{zt} - \epsilon_{zt-1}) \tag{10}$$

$$\Delta y_{zt} = \Delta\alpha_t + \beta\Delta b_{zt-3} + \Delta\epsilon_{zt} \tag{11}$$

The required exogeneity condition for consistent estimation of β is hence:

$$E(\Delta b_{t-3}\Delta\epsilon_{zt}) = 0 \tag{12}$$

As $\Delta b_{zt-3} = \Delta \sum_{\tau=1}^{t-3} Y_{zt} = (\sum_{\tau=1}^{t-3} Y_{zt} - \sum_{\tau=1}^{t-4} Y_{zt}) = Y_{zt-3}$, condition 12 can be written as:

$$E(Y_{zt-3}\Delta\epsilon_{zt}) = E(Y_{zt-3}\epsilon_{zt}) - E(Y_{zt-3}\epsilon_{zt-1}) = 0 \tag{13}$$

By construction of equation 1 Y_{zt-3} is not only correlated with its contemporaneous error ϵ_{zt-3}, but also with all previous errors, i.e. $E(Y_{zt-3}\epsilon_{zt-\tau}) \neq 0 \forall \tau \geq 3$. However, if ϵ_{zt-3} is uncorrelated with ϵ_{zt-1}, i.e. if the order of autocorrelation of ϵ_{zt}, is smaller than 2 ($\nu < 2$) the consistency condition given in 12, holds.

Taking this to the example of solar PV technology adoption, the suggested first differencing estimation strategy thus yields consistent estimates, if the lead time between adoption decision and installation is large enough. More specifically, given the natural time lag between adoption decision and installation of three months, the inclusion of the third lag of the installed base allows for a first differencing estimation strategy that yields a consistent estimate of the installed base effect, as long as the order of autocorrelation is smaller than 2, i.e. $\nu < 2$. This argument can be generalized: if the considered lag of the installed base is l, then $E(Y_{zt-l}\epsilon_{zt-\tau}) \neq 0 \forall \tau \geq l$ and consistent estimation is feasible as long as $E(\epsilon_{zt-l}\epsilon_{zt-1}) = 0$, i.e. the order of autocorrelation ν of the error ϵ must be smaller than $(t-1)-(t-l) = l-1$.

Proposition 1. *Let the lag of the installed base be l and let the order of autocorrelation of ϵ be ν. Then, if $\nu + 1 < l$, i.e. the order of autocorrelation ν is smaller than $l - 1$, consistent estimation of the installed base effect is feasible.*

5.3.1 Testing for Autocorrelation

The previous section established the condition for consistent estimation of the installed base effect: when including the third lag of the installed base, b_{t-3}, the order of serial correlation of ϵ must be smaller than 2 ($\nu < 2$). The tests for autocorrelation of ϵ are based on the first differenced equation, as the errors $u_{ztq} = \alpha_{zq} + \epsilon_{zt}$ in the level equation 1 are by construction serially correlated through the fixed-effects α_{zq}. However, the errors $\Delta\epsilon$ in the differenced equation 10 are by construction serially correlated as well (ϵ_{t-1} enters $\Delta\epsilon_t$ as well as $\Delta\epsilon_{t-1}$), which is why AR(2) of $\Delta\epsilon_t$ is usually tested to make inference on AR(1) of ϵ_{zt}. Equation 14 illustrates that AR(1) of ϵ_{zt} can lead to AR(2) in $\Delta\epsilon_{zt}$, if $E(\epsilon_{zt-1}\epsilon_{zt-2}) \neq 0$.

$$E(\Delta\epsilon_{zt}\Delta\epsilon_{zt-2}) = E(\epsilon_{zt}\epsilon_{zt-2} - \epsilon_{zt}\epsilon_{zt-3} - \underbrace{\epsilon_{zt-1}\epsilon_{zt-2}}_{\text{AR(1) of } \epsilon \text{ can lead to AR(2) in } \Delta\epsilon} + \epsilon_{zt-1}\epsilon_{zt-3}) \tag{14}$$

Hence, AR(2) of $\Delta\epsilon$ does not necessarily, result from AR(2) of ϵ. On the other hand, AR(3) in $\Delta\epsilon$ *must* result from serial correlation of at least order 2, i.e. $\nu \geq 2$.

$$E(\Delta\epsilon_{zt}\Delta\epsilon_{zt-3}) = E(\epsilon_{zt}\epsilon_{zt-3} - \epsilon_{zt}\epsilon_{zt-4} - \underbrace{\epsilon_{zt-1}\epsilon_{zt-3}}_{\text{AR(2) of } \epsilon \text{ can lead to AR(3) in } \Delta\epsilon} + \epsilon_{zt-1}\epsilon_{zt-4}) \tag{15}$$

Equation 15 illustrates that AR(2) of ϵ can lead to AR(3) in $\Delta\epsilon$, if $E(\epsilon_{zt-1}\epsilon_{zt-3}) \neq 0$. In the empirical application of this paper and motivated by the reasoning above, $\nu < 2$ of ϵ is hence tested based on the differenced equation 10. As expected, AR(1) of $\Delta\epsilon$ is high by construction of the differenced equation. Second, third and fourth order autocorrelation are found to be much lower. The test for AR(3) in the differenced model can reject the H_0 on a 1 percent significance level (see Appendix table A.3). It thus appears that the suggested identification and estimation strategy yields consistent estimates of the installed base effect.

6 Data

This paper focuses on social effects within geographically defined reference groups. Since solar PV panels are visible for passers-by, no social bonding in the sense of friendship, family, culture or religion is required for social effects such as observational learning. This motivates the spatial definition of the reference group.

6.1 Postcodes and LSOAs

To highlight the granularity of the analysis, consider the different geographic area definitions in the UK. A postcode district is identified by the first three to four digits of a full UK postcode, e.g. by CB3 in case of the postcode CB3 9DD. Each full UK postcode is divided by a space into two parts: the first part indicates the postcode area and the postcode district, e.g. CB and CB3 respectively. The second part begins with a single digit for the postcode sector within each district, e.g. 9 in case of CB3 9DD. In August 2012, there were 124 postcode areas, 2,987 postcode districts, approximately 11,000 postcode sectors and 1.78 million unit postcodes in the UK (including Channel Islands and Isle of Man)(Royal Mail, 2013). While postcodes are mainly used for postal services, the Office for National Statistics (ONS) defined so-called 'Output Areas'.'Lower Layer Super Output Areas'(LSOA), in particular, were designated to provide a geographical hierarchy with the intention of improving the reporting of small area statistics in England and Wales. Each LSOA has its unique code and can be assigned to a postcode in England and Wales.[8] There are 32,844 LSOAs in England and 1,909 in Wales (ONS, 2013). The analysis is performed on the postcode district rather than the LSOA level, as the number of solar PV installations within LSOAs is often small or zero, implying (too) little variation of the variables of interest for identification of the main effect. Moreover, a main feature of the model is the focus on social effects *within* neighbourhoods. No cross-border spillovers are taken into account.

6.2 Neighbourhood Statistics

For England and Wales the ONS provides Census statistics, such as on tenure for example, for multiple geographical aggregation levels, among others for the postcode district level. An LSOA has a size between a full postcode and a postcode district area: the LSOAs consist of 400 to 1,200 households and make up approximately 2,269 postcode districts in England and Wales. Each postcode district consists, on average, of 10,297 households, with just under 65 percent of these owner-occupied (ONS, 2013). The number of owner-occupied households is of particular relevance for the installation of solar PV panels, as the decision to install a solar PV panel is more likely to be made by households who own their property than by households who rent it.[9] Even if tenants can benefit from the panels as well (e.g. by being provided with some free electricity), it is the landlord who receives the FiT and who decides whether to install the panel. The number of tenants inducing their landlords to install solar PV are assumed negligible for this analysis.

[8] A pseudo code is available for Scotland, Northern Ireland, Channel Islands and the Isle of Man, but the definitions and the geographical size, in particular, are neither entirely consistent nor provided by the ONS.

[9] ONS tenure data provides information about whether a household rents or owns the accommodation that it occupies and, if rented, combines this with information about the type of landlord who owns or manages the accommodation.

Figure 1 shows the distribution of the number of households (top) and the number of owner-occupied households (bottom) within the postcode districts.

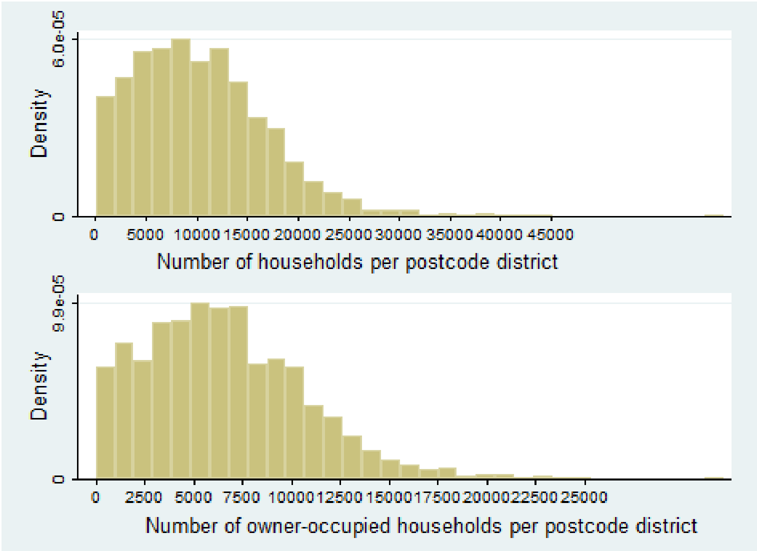

Figure 1: Distribution of the number of households (top) and of owner-occupied households (bottom) within postcode districts in England and Wales. The average number of households in a postcode district is 10,297. The average number of owner-occupied households in a postcode district is 6,546 (own calculation based on ONS, 2013).

Table A.2 in the Appendix lists the summary statistics, so called *contextual factors*, for the considered postcode districts as retrieved from the ONS. The table lists the characteristics for those 2,239 postcode districts that, by the end of March 2013, had at least one domestic solar PV installation. Tests of mean differences for adopting and non-adopting postcode districts do not show significant differences between the considered characteristics.

The ONS only provides neighbourhood statistics for England and Wales. Future research, however, could match the neighbourhood statistics of the Scottish Census with the postcode districts, using the pseudo LSOA code and the look up tables provided by the Scottish government. As the solar PV installation dataset does not include the LSOA pseudo code and in order to stick to a single definition of neighbourhoods[10], installations in Scotland were excluded from the main analysis. It is remarkable that even in Scotland around 93 percent of the postcode districts had at least one domestic solar PV installation installed by the end of March 2013, although the number of sunshine hours is 22 percent lower than in England (Met Office, 2007).

[10]The existing data zones in Scotland are smaller in population size than their LSOA counterparts in England and Wales. They have a minimum population of 500, compared to a minimum of 1000 in LSOAs.

6.3 Solar PV Installation Data

The econometric analysis in this paper is based on solar PV installation data from April 2010 through March 2013. The dataset is published by Ofgem and updated quarterly. It includes all domestic, commercial and communal micro-generation systems in the UK that have been registered to receive the FiT since its introduction in April 2010 (i.e. systems with up to 50kW capacity). By the end of March 2013, 379,531 installations were registered. Besides an individual identifier for each installation, the dataset includes geographic location of the installation down to the LSOA level, the corresponding postcode district (i.e. the first four digits of the postcode the LSOA code is associated with), the day at which the installation was commissioned, i.e. completed, as well as the installed capacity (in kW).

Of all 379,531 installations that were registered in April 2013, just under 83 percent were commissioned in England, 7.3 percent in Wales and 6.9 percent in Scotland. Solar PV technology dominates the market for micro-generated energy: 98.6 percent of all micro-generation installations are solar PV systems. Wind contributes only 1.2 percent and anaerobic digestion and hydro even smaller shares of 0.01 and 0.11 percent. Moreover, the share of domestic installations lies above 96 percent (96.5 of all micro-generation technologies and 96.9 percent of the solar PV installations are domestic installations). As this paper aims to analyse the diffusion of a single technology rather than the choice between different kinds of technologies and focuses on social effects within neighbourhoods, all non-solar and non-domestic PV technology installations are excluded from the analysis. 4,945 solar PV installations were not associated with any locational information on LSOA or postcode district and hence excluded from the analysis, too. Finally, for the reasons mentioned above, the analysis considers solar PV installations in England and Wales, only. The cleaned data set counts 332,216 domestic solar PV installations in England and Wales by the end of March 2013. These are associated with 2,239 postcode districts with an average of 10,427 households each, with on average 63.5 percent of this (6,629) owner-occupied. Only districts with at least one domestic solar PV system commissioned between March 2010 and 2013 are considered, which are striking 98.7 percent of all postcode districts in England and Wales. The 30 postcode districts that had not installed any domestic solar PV panels were mainly districts in the centre of big cities such as Manchester (M1, M17, M50, M60), Liverpool (L2), Birmingham (B1, B2, B3), Leeds (LS1) and London (20 postcode district areas around the Tower Bridge, Westminster, Leicester Square for example). Moreover, Heathrow Airport (TW6) is among the non-adopting postcode districts. It is thus not at all striking that there are no domestic solar PV installations in those areas as these non-adopting postcode districts obviously differ significantly in major neighbourhood characteristics. While the average number of households and owner-occupied households is significantly lower for example (1,412 compared to 10,417 households and 360 compared to 6,629 owner-occupied households), the share of households without car and the share of households using public transport to travel to work is significantly higher in non-adopting postcode districts.

The main variable of interest is the installed base of solar PV panels on a postcode district level. The installed base, as suggested above, serves as a measure of social effects and refers to the cumulative number of solar PV installations within a reference group z at a particular point in time t. For the remainder of this paper, z refers to the postcode district (interchangeably used with 'zip code') and t refers to the month. More precisely, the installed base, b_{zt} in postcode district z in month t is defined in equation 16. Where Y_{zt}

refers to the number of new installations Y in zip code z in month t.

$$b_{zt} = \sum_{\tau=1}^{t} Y_{zt} \tag{16}$$

These variables are derived exploiting the postcode district identifier and the commission date of each installation. The resulting postcode district level panel data set consists of 80,604 data points: for 2,239 zip codes in England and Wales and for each of the 36 months there is an observation for the installed base and the count of new installations .

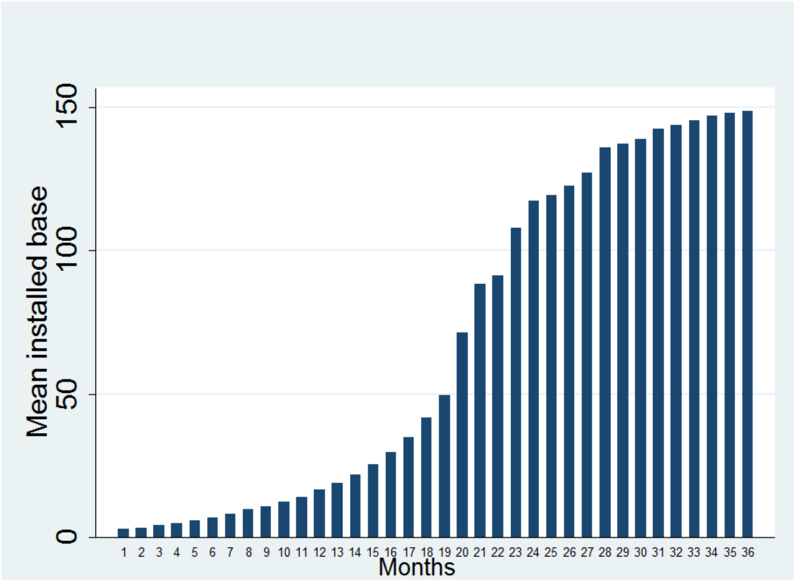

Figure 2: Average installed base in the postcode district (April 2010 - March 2013). $\frac{1}{2,239}\sum_{z}^{2,239} b_{zt} \forall\, t = 1 \cdots 36$. Source: Ofgem, own calculations.

Figure 2 illustrates the increase in the average installed base from 2.3 solar PV systems within a postcode district in April 2010 to 148.4 domestic solar PV installations by the end of March 2013. Figure 3 illustrates how the installed base varies across postcode districts. For each zip code the graph plots all 36 installed base observations.

Figure 4 plots the average number of monthly installations per owner-occupied household within the 2,239 considered postcode districts for the 36 considered months. The average adoption rate increased from 0.071 installations per 1000 households to 0.192, i.e. almost tripled within the considered two years. Moreover, the adoption rate has never fallen below the rate of March 2010 and peaked in month 20, i.e. in November 2011, with 4.065 installations per 1000 owner-occupied households.[11] The peaks of the adoption rate are most pronounced during the months before cuts of the subsidies came into effect. In December

[11] One could argue that these peaks in adoption rates lead to a decreasing pool of potential adopters within the zip code. However, as multiple installations of a household are possible, this paper assumes the pool of potential adopters to be unaffected by the number of installations. I.e. the number of owner-occupied households within a zip code as reported in the Census 2011 (ONS, 2013) is used for normalisation throughout.

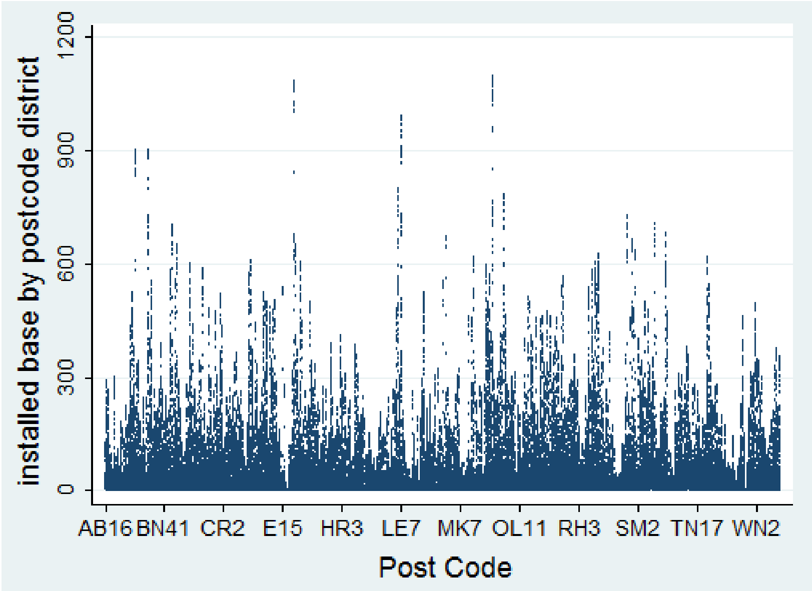

Figure 3: Installed base by postcode district (April 2010 - March 2013). For each zip code the graph plots the 36 installed base observations, i.e. $b_{zt} \forall\, t$ by postcode district. Source: Ofgem, own calculations.

2011, for example, the UK Government announced rigorous cuts of the FiT, which led to a visible demand response shortly before this change became effective. This reflects the fact that it is the commission date (i.e. the date of completion of the panel) that matters for the level of FiT paid. If a panel is installed before the official date of the subsidy change, the old FiT is paid throughout the 20 years of support. Such kinds of policy announcements must be controlled for, e.g. by the inclusion of time fixed effects in the econometric equation, as they impact the installed base as well as the adoption rate and can confound the estimation of the installed base effect.

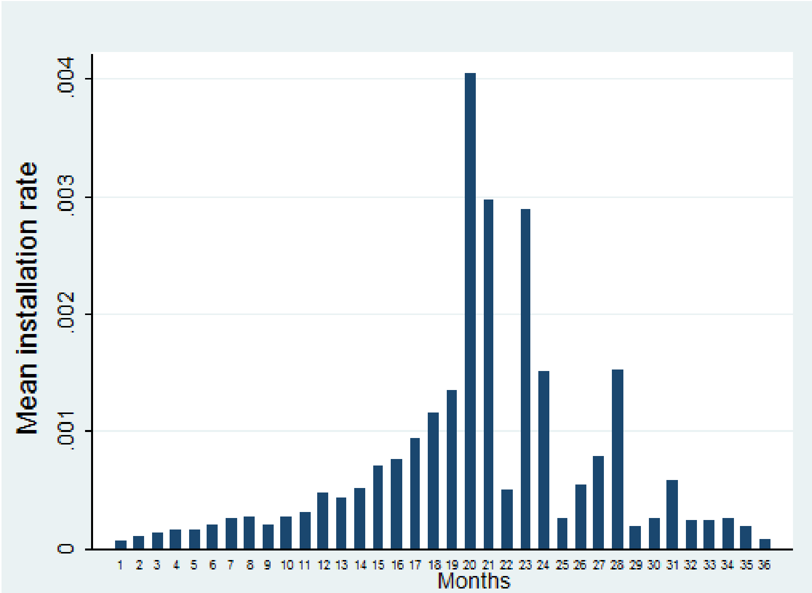

Figure 4: Average adoption rate (April 2010 - March 2013), i.e. $\frac{1}{2,239}\sum_{z=1}^{2,239} y_{zt} \forall t = 1 \ldots 36$. To control for the variation of owner-occupied households across postcode districts, the installation counts are normalised with the number of owner-occupied households in the postcode district. The resulting adoption rate serves as the main outcome variable of interest, y_{zt}. Source: Ofgem, ONS, own calculations.

7 Results

Table 1 lists the estimates of the social effects resulting from the suggested pooled OLS ($POLS_{cl}$), within-group (WG_{cl}) (de-meaned on the postcode district-quarter) and first difference estimations (FD_{cl}) based on equation 1. To account for serial correlation within a postcode district over time, the estimates allow for clustering on the postcode district level. The results suggest small, but positive and significant social effects.

The estimates are consistent with the theory: the POLS estimate ($POLS_{cl}$) is strongly downwards biased. This corresponds to the expected omitted variable bias, as the only regressor included in equation 1, apart from the month dummies, is the lagged installed base. The α_{zq} are unobserved, but as it is likely that there are several neighbourhood characteristics that are correlated with the installed base *and* the outcome variable, the omitted variable bias (OVB) results. The within-group estimator (WG_{cl}) is slightly lower than the first difference estimate (FD_{cl}). As illustrated above, for a lag of at least three, the within-group estimator of the installed base coefficient β can yield consistent estimates. For shorter lag lengths (i.e. shorter lead time between adoption decision and installation) the within-group estimator is inconsistent due to a correlation of the mean-differenced error with the mean-differenced installed base. As consistency of the first difference estimate (FD_{cl}) is conditional on weaker assumptions than those required for a consistent within-group estimate, the first difference estimate is the one of interest.

According to the first difference estimate (see FD_{cl} in table 1) one more solar PV panel in a postcode district increases the number of new installations per owner-occupied households three months later by $7.48e^{-06}$. At the average installation rate of 0.0007, this increase is equivalent to a one percent increase of the average installation rate. This is obviously and as expected a very small effect. At the average number of 6,629 owner-occupied households within a postcode district, it implies that one more solar PV panel in the neighbourhood increases the number of new installations in the neighbourhood by 0.005.[12] It would hence require around 200 additional solar PV panels in a postcode district, for social effects alone to be strong enough to cause one further installation within the neighbourhood three months later. The installed base elasticity at the average installed base of 68 installations within a neighbourhood and the average installation rate of 0.0007 is 0.71, implying a rather inelastic demand response. At the median installed base of 27 and the average installation rate of 0.0007, the installed base elasticity is with 0.29 even lower. These results illustrate that the social effects as measured by the installed base are very small, but exist and can promote adoption. Particularly larger scale projects might hence lead to observational learning from spatially close installations. It is tempting to compare the results to the study of Bollinger and Gillingham (2012), who find an increase of the installation rate of $1.567e^{-06}$ for any additional solar PV installation in a neighbourhood.[13] However, due to differences in the definition of neighbourhoods, the granularity of the data and the model assumptions, such comparisons should be treated with caution.

Table 1: The table shows the estimates of the installed base effect on the installation rate within a postcode district. Estimations are clustered on the postcode district level. The first difference estimate suggests small, but positive and significant social effects. One more solar PV panel in a postcode district increases the number of installations per owner occupied household (that is on average at 0.0007) by significant $7.48e^{-06}$.

Variable	$POLS_{cl}$	WG_{cl}	FD_{cl}
Installed Base (L.3)	1.93e-06***	6.59e-06***	7.48e-06***
	(1.20e-07)	(2.44e-06)	(2.66e-06)
Observations	73,887	73,887	49,258
R-squared	0.145	0.080	0.075

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

8 Robustness Checks

For robustness, the analysis is additionally performed for model specifications that allow for a non-constant installed base coefficient, for different lags of the installed base and for a distinct geographical level, namely the local authority level. Moreover, besides the installation rate, the installed capacity per owner-occupied household (in kW) is considered as outcome variable.

[12] The average third lag of the installed base was 60.93 and the average number of new adoptions within a neighbourhood was 4.07.

[13] Their neighbourhoods contain on average 4,959 households and are geographically larger.

8.1 Non-Constant Installed Base Effect

To allow for non-constant social-effects, the model in equation 1 is extended by the squared lagged installed base.

$$y_{zt} = \alpha_t + \beta_1 \cdot b_{zt-3} + \beta_2 \cdot (b_{zt-3})^2 + \underbrace{\alpha_{zq} + \epsilon_{zt}}_{u_{ztq}} \tag{17}$$

Estimating equation 17 by POLS, within-group and first difference strategies, the bias of the POLS and the within-group estimator remain their order and direction. All three estimates suggest, however, a positive installed base effect that decreases with the size of the installed base. The resulting estimates ($POLS_{cl}$, WG_{cl} and FD_{cl}) are listed in table 2. This positive but decreasing effect of the installed base on the installation rate is consistent with the idea of satiation within neighbourhoods. When the installed base is low, the solar PV panels might be particularly attention catching as they are perceived as especially innovative. If so, social spillovers are likely to be most pronounced in the early stages and decrease with the number of installations then.

Table 2: Testing for a heterogeneous installed base effect suggests positive and significant social effects that decreases with the size of the installed base.

Variable	$POLS_{cl}$	WG_{cl}	FD_{cl}
Installed Base (L.3)	4.65e-06***	6.78e-06***	1.45e-05***
	(3.34e-07)	(2.45e-06)	(3.88e-06)
Installed Base Squared (L.3)	-5.56e-09***	-4.83-11***	-1.17e-08***
	(6.73e-10)	(1.45-11)	(3.10e-09)
Observations	73,887	73,887	49,258
R-squared	0.148	0.080	0.075

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

$$y_{zt} = \alpha_t + \beta_1 \cdot b_{zt-3} + \beta_2 \cdot (D_q \cdot b_{zt-3}) + \underbrace{\alpha_{zq} + \epsilon_{zt}}_{u_{ztq}} \tag{18}$$

Another, comparable, specification allows for a variable installed base effect over time. Equation 18 specifies the model with interactions of 12 quarter dummies, D_q, with the lagged installed base. The results suggest a time-varying installed base effect that overall decreases over time. The coefficient varies throughout the year (see Figure **??** in the Appendix).[14] Table A.4 in the Appendix lists the respective coefficients. Interestingly, the social effects are particularly pronounced during times of policy announcements to decrease the FiT. In the seventh quarter (Q_7) for example, i.e. from October to December 2011, the social effects were significantly higher than in the initial months. This could be due to higher media presence that caused higher awareness of the technology and higher perception of the installed panels,

[14] Testing the coefficients of the first differenced equation for equality, the H_0 must be rejected, which is consistent with a time varying installed base effect. The test results in an F-statistic of 32.23 and a p-value of 0.000.

for example: in October 2011 the UK government announced a remarkable decrease of the FiT for December 2011. This policy change was then postponed until March 2011. Since all registrations until March 2011 were eligible to the higher FiT, strong and positive social spillovers before the effectivity date do make sense. Towards the end of the considered 12 quarters, the marginal impact of the installed decreased from the tenth quarter onwards. These are known to be months of relative policy uncertainty, during which adoptions in general were particularly low. Social effects not only decreased, but turned negative as well during this time.

These regressions show that the linear model is misspecified and a richer nonlinear model could improve the analysis. However, the main findings regarding the installed base effect are preserved. Future research could explore this further by testing the model against a dynamic specification that takes some kind of dynamic adjustment into account.

8.2 Testing Different Lags

The estimates of social effects are based on the presumption that the lead time between the adoption decision and the completion of the installation is three months. This is obviously just an average. On the one hand the lead may vary across suppliers, on the other hand the lead may vary over time. In particular, during times before the effectivity dates of the FiT cuts, the time between decision to adopt and actual installation might have varied. It makes thus sense to consider different time-lags of the installed base for robustness.

The coefficients of the contemporaneous and the first lag of the installed base show the expected downwards bias. This is due to the correlation of the first differenced installed base with the first differenced error (see section on identification). The second and third lag have a positive and significant impact on the installation rate. Given the quoted two to three months regarding the lead time between adoption decision and installation, sign and significance of the coefficient of the second and third lag make sense. The coefficient of the fourth lag is negative, but not significant. This could indicate a 'time threshold' for the installed base effect. It might be, that even if households already perceive and learn from the panels, they are not actually *deciding* that early. These findings might suggest that social effects exist, but are only effective in a rather narrow time window. Table 3 shows the first difference estimates (FD_{cl}) for the specifications with different lags of the installed base.

Table 3: First difference estimates for different lags of the installed base. The results are consistent with the quoted time lead between adoption decision and installation. The estimates suggest that social effects exist, but are only effective in a rather narrow time window.

Variable	FD_{cl}	$FD_{cl}L.1$	$FDb_{cl}L.2$	$FD_{cl}L.3$	$FD_{cl}L.4$
Installed Base	5.75e-05***	-3.63e-05***	3.81e-06**	7.48e-06***	-2.15e-06
	(3.18e-06)	(3.78e-06)	(1.78e-06)	(2.66e-06)	(1.69e-06)
Observations	51,497	51,497	49,258	49,258	47,019
R-squared	0.104	0.082	0.074	0.075	0.074

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

8.3 Installed Base Effect on the Local Authority (LA) Level

For robustness the main regressions are run at the Local Authority (LA) level. This approach tests for social effects within broader defined geographical neighbourhoods than in the case of the postcode districts. The average number of owner-occupied households in the considered local authorities is 42,602, the average installed base 409 solar panel installations. The analysis on the LA level allows for spillovers within larger geographical units and hence across the borders of the postcode districts. This can matter, for instance, if the mobility within a LA is rather high. The estimations based on equation 17, but with z referring to the local authority, yield the results listed in table 4. As expected, the installed base effect

Table 4: Installed base effect in local authoritie. The results suggest that the impact of social effects, as proxied by the installed base, are stronger on a more local level than in broader defined neighbourhoods.

	$POLS_{cl}$	WG_{cl}	FD_{cl}
Installed Base (L.3)	3.14e-07***	1.60e-06*	1.77e-06*
	(5.71e-08)	(8.24e-07)	(9.21e-07)
Observations	11,451	11,451	7,634
R-squared	0.396	0.341	0.346

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

on the LA level is smaller than on a postcode district level. The first difference estimates are $1.77^{e-06}*$ for the installed base effect within local authorities and $7.63^{e-06}***$ for the effect within postcode districts. This is consistent with the argument that social effects are stronger on a more localised level. The coefficient 1.77 implies that at the average number of owner-occupied households within a local authority (42,602), one more domestic solar PV installation increases the number of new installations in the local authority by 0.075.

8.4 Modelling the Installed Capacity

The main specification considered in equation 1 models the installation rate of owner-occupied households (defined as the number of installations per owner-occupied household) as the outcome variable. However, the average impact of the installed base on the installed capacity (in kW) per owner-occupied household can serve as robustness check of the results presented in section 7.

Table 5 lists the POLS, within-group and first-difference estimates for the installed base effect on the new installations when measuring both, installed base and the new installations in terms of capacity (i.e. in kW). According to the first difference estimate, a marginal increase in the installed capacity within a postcode district increases the installed capacity per 1000 owner-occupied households by significant 0.00739 kW (see Table 5). At the average number of 6,629 households within a postcode district, that implies that the installed capacity increases by 0.0049 kW. The installed base elasticity at the average installed base of 210kW and the average installation rate of 13kW per owner-occupied household is 0.67, which is close to the elasticity of 0.71 found at the average installed base and installation rate in the main specification that is based on installation counts. This holds also true for the installed base elasticity at the median installed base. When specifying the model in terms of capacity the elasticity is 0.23, while being 0.28 when specifying the model in terms of installation counts. These results are hence consistent with positive and significant social effects.

Table 5: Installed base effect on the installed capacity (kW) per owner-occupied household. In this specification the installed base and the outcome variable are measured in kW.

Variable	$POLS_{cap}$	WG_{cap}	FD_{cap}
Installed Base (L.3)	2.40e-06***	7.48e-06***	7.39e-06***
	(1.27e-07)	(1.40e-06)	(1.50e-06)
Observations	73,887	73,887	49,258
R-squared	0.156	0.090	0.081

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

9 Contextual Factors and the Installed Base Effect

For the UK, the Department for Energy and Climate Change (DECC) provided a descriptive analysis of the trends in the deployment of domestic solar PV technology under the FiT scheme (DECC, 2012). In this study the effect of contextual factors (e.g. of income, education, gas and electricity consumption within a neighbourhood) is considered and the report suggests that more affluent, higher energy consuming neighbourhoods are more likely to install solar panels. On top of this, areas with a high proportion of detached housing, a low proportion of social housing and a low proportion of low value housing tend to have a higher amount of solar PV installations. Moreover, rural areas have a higher density of solar PV installations than urban areas. Given that several neighbourhood characteristics crucially matter for the adoption rate, the relative importance of social effects, hence of the installed base effect, for different groups of the population should be explored further. Social spillovers in affluent neighbourhoods, for example, might differ from social spillovers within deprived neighbourhoods. As suggested by DECC, higher income neighbourhoods tend to be more likely to adopt than lower income neighbourhoods. Social effects, on the other hand, might be ceteris paribus of more importance in less educated or less affluent neighbourhoods, as higher income households tend to adopt early and observational learning might therefore play a less important role (Rogers, 1962). An argument that could support negative social effects, for example, could be that of snob effects within rich neighbourhoods.

9.1 The Use of ONS Neighbourhood Statistics

To estimate the installed base effect for different types of neighbourhoods, interactions of contextual factors with the installed base are incorporated into the model. However, while the installation data is available in panel form, the postcode district neighbourhood characteristics are only available for the cross-section of March 2011, the date of the last Census in England and Wales. It is hence assumed that any changes of the postcode district characteristics over the considered time frame (2010-2013) are negligible for the installed base effect. This is clearly a strong assumption and simplification, but allows for first insights into heterogenous installed base effects.

Furthermore, most of the ONS neighbourhood characteristics are continuous variables. To allow for more meaningful interpretation of the coefficients of the interaction terms (i.e. the interactions of contextual factors with the installed base), the postcode districts are grouped into quintiles, i.e. categories. Dummy variables indicate whether the respective postcode district characteristic lies above the 60th percentile or not. As an example, the dummy *Share Social Class AB* indicates whether the share of households in social class AB within the postcode district lies above the 60th percentile. If so, the remainder of this paper classifies such a neighbourhood as 'high social class' neighbourhood. Likewise, the dummy *Population Size* indicates whether the population in the neighbourhood is particularly high compared to other neighbourhoods, i.e. lies in the upper 40th percentiles. Equation 19 specifies the model that includes the installed base, b_{zt-3}, the dummies indicating the contextual factors, X_z[15] and their interactions with the installed base.

$$y_{zt} = \alpha_t + \beta_1 \cdot b_{zt-3} + \beta_2 \cdot X_z + \beta_3 \cdot (X_z \cdot b_{zt-3}) + \underbrace{\alpha_{zq} + \epsilon_{zt}}_{u_{zt}} \tag{19}$$

First differencing equation 19 results in equation 20. POLS on this differenced equation (clustered on the postcode district level and dropping, again, the first month of each quarter s.t. the neighbourhood-quarter effects drop out) is performed. The main effects of the contextual factors are not identified as all time-constant variables drop out when first differencing. However, the coefficients of the interaction effects can indicate differences of the installed base effect across neighbourhoods with different contextual characteristics.

$$\Delta y_{zt} = \Delta\alpha_t + \beta_1 \cdot \Delta b_{zt-3} + \beta_3 \cdot \Delta(X_z \cdot b_{zt-3}) + \Delta\epsilon_{zt} \tag{20}$$

9.2 Results for Contextual Factors and The Installed Base

The coefficients of interest are the ones of the interaction terms resulting from the first difference estimation. The results suggest significantly stronger peer effects in neighbourhoods with a high population (in the sense of a population size above the 60th percentile as explained above). In neighbourhoods with a high share of unshared houses social effects are less pronounced, which might result from the higher likelihood to interact with neighbours in the immediate living environment (e.g. in the case of attached houses). Interestingly, relatively affluent (non-deprived) neighbourhoods and neighbourhoods with a high share of high social class households, show a less pronounced installed base effect. As Rogers (1962) suggests, this might result from the fact that those households are early adopters and hence the learning from others is less important. Higher educated neighbourhoods on the other hand show stronger installed base effects than neighbourhoods with on average lower educated populations. In neighbourhoods with a high ratio of economically active people, social effects are relatively low, which can result from the fact that those people spend less time in their neighbourhood. Finally, social effects in neighbourhoods with a high share of people travelling to work by bike or foot are less strong while private transport in a car increases the installed base effect.

[15] As contextual factors are assumed to be constant over the considered time frame, the dummies only have a z index and no time index.

Table 6: Coefficients of the First Difference Estimation with Covariate-Installed Base Interactions (Table only shows the coefficients relevant for interpretation. Month dummies and installed installed base main effect are ommitted.)

Variable	FD_covXb fd_new_norm
Installed Base (L.3)	-4.91e-06
	(6.73e-06)
Population Size X L.3 Base	6.69e-06**
	(2.92e-06)
Unshared Houses X L.3 Base	-6.10e-06**
	(2.73e-06)
Share Male X L.3 Base	8.65e-06***
	(2.40e-06)
Share White X L.3 Base	3.03e-06
	(3.81e-06)
Share 20 to 45 X L.3 Base	8.07e-06**
	(3.86e-06)
Share 65 plus X L.3 Base	4.06e-06*
	(2.43e-06)
Share Deprived in 0 Dim X L.3 Base	-4.06e-06
	(3.71e-06)
Share No Car X L.3 Base	4.31e-06
	(4.70e-06)
Share Social Class AB X L.3 Base	-5.20e-06*
	(3.10e-06)
Share University Degree X L.3 Base	1.14e-06
	(2.57e-06)
Share Economically Active X L.3 Base	-4.23e-06
	(3.68e-06)
Share No Travel to Work X L.3 Base	9.41e-06***
	(2.24e-06)
Share Public Travel X L.3 Base	-4.32e-06
	(2.93e-06)
Share Private Travel X L.3 Base	4.68e-06
	(3.69e-06)
Share Bike or Foot X L.3 Base	-2.09e-06
	(2.30e-06)
Constant	2.29e-05**
	(9.02e-06)
Observations	49,258
R-squared	0.076

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

10 Limitations and Suggestions for Further Research

The presented analysis, which is at a relatively disaggregated level, has several limitations. Still, the results serve as a useful first indicator for the impact of social effects on solar PV technology adoption in England and Wales. The model could be extended in future research.

The first major constraint of the model employed in this paper is that social effects are assumed to spread within predefined neighbourhoods only, while spillovers across the neighbourhood borders are ignored. Spatial econometric methods, for example, could be employed to allow for more diverse spatial effects. Another related limitation is the aggregation of the installations to the neighbourhood level. Household level data could improve the analysis and, in particular, shed light on the exact mechanisms of social interactions. Unfortunately, such data is currently not available.

Moreover, the approach to measure social effects with the lagged installed base has some drawbacks: as has been emphasised throughout this paper, the quarter lagged specification is important for the estimation method. The lag reflects the assumption that social spillovers such as observational learning only come into effect once the solar PV panel is installed. This is a reasonable and justifiable assumption, but one can imagine some social effects coming into effect before the completion already. For example, there might already be social effects once the solar PV panel supplier van shows up at the neighbour′s house to discuss the further procedure, or there might be word-of-mouth through the neighbours once they have set-up the contract with the supplier. These kind of spillovers are not captured by the lagged installed base. Additionally, friends, family and colleagues might contribute to social spillovers other than those captured by the lagged installed base. The consideration of different lags and the wider definition of neighbourhoods are an attempt to get an idea of the relevance of such effects, but further methods such as IV-strategies or a bias-correction approach, as suggested by Narayanan and Nair (2011), could be considered for comparison of the results and could help to pin down the mechanisms underlying the social effects more precisely.

A further concern could be that the model obscures any possible partial adjustment processes resulting from inertia in the decision process. An increase in the lagged installed base b_{zt-3} might have an effect on the installation rate in t, but due to adjustment costs, a part of the intended increase might be postponed to the next period. If so, the causal interpretation of the installed base effect comes into question. As an example, there might be households who decide to install in a particular month, but are not able to do so in time due to lacking financial resources. If this is the case, then the observed installation rate, y_{zt}, might be a compromise between the previous installation rate, y_{zt-1}, and the lagged installed base, b_{zt-1}, (and other exogenous variables). This idea can be captured in a partial adjustment model. Let $y^{\star}_{zt}$ be the true, but unobserved installation rate in t and assume the true relationship between the installed base and this 'intended' installation rate is given as follows:

$$y^{\star}_{zt} = \gamma_t + \delta b_{zt-3} + v_{ztq} \tag{21}$$

γ_t are month dummies, δ is the causal installed base effect of interest and v_{ztq} is an error term equivalently defined as in the main specification given in equation 1. Now, assume that there is inertia in the system such that the econometrician does not observe the 'intended' installation rate, $y^{\star}_{zt}$, but due to the inertia observes the realised installation rate, y_{zt}, instead.

$$(y_{zt} - y_{zt-1}) = \lambda(y^{\star}_{zt} - y_{zt-1}) \tag{22}$$

Equation 22 reflects the idea of partial adjustment. In this framework the actual (observed) change in the outcome variable, $(y_{zt} - y_{zt-1})$, is assumed proportional to the 'intended' change $(y^{\star}_{zt} - y_{zt-1})$. Rearranging equation 22 yields an equation that suggests that the observed adoption rate y_{zt} is a weighted average of the 'intended' installation rate $y^{\star}_{zt}$ and the lagged observed installation rate y_{zt-1}. λ measures the speed of adjustment:

$$y_{zt} = \lambda y^{\star}_{zt} + (1 - \lambda) y_{zt-1} \tag{23}$$

Substituting equation 21 into equation 23 yields an equation in terms of the observable variables included in equation 23.

$$y_{zt} = \lambda\gamma_t + \underbrace{\lambda\delta}_{\tilde{\beta}} b_{zt-3} + (1 - \lambda) y_{zt-1} + \lambda v_{ztq} \tag{24}$$

Equation 24 exemplifies how a partial adjustment process could challenge the identification of the causal installed base effect. Firstly, if there is indeed partial adjustment, the model given in equation 1 above, would be misspecified. Even when estimating equation 24, the installed base coefficient is $\tilde{\beta} = \lambda\delta$. The actual causal installed base effect, δ, is thus not separately identified from λ. Rather, $\delta = \dfrac{\tilde{\beta}}{\lambda}$ where λ can be derived from the estimated coefficient of the lagged dependent variable y_{zt-1}.

To get an idea how relevant the concern regarding partial adjustment is, one could compare the diffusion of solar PV technology with the diffusion process of a technology with a comparable time-lag between adoption decision and installations but of which observability of the installation per se is not a main feature. An example could be the installation of domestic internet connections in the early days of this technology. In summary, if there is partial adjustment in the application of solar PV technology adoption, the estimated coefficients of the lagged installed base might be biased and likely to indicate correlation rather than causation.

A final concern to mention here is the specification as a linear probability model. The regression results show that although the main findings regarding the installed base effect are preserved, the linear model is misspecified and a richer non-linear model could improve the analysis. Future research could test the model against a dynamic specification that takes some kind of dynamic adjustment into account. Also, it could be considered to use transform the outcome variable to $F^{-1}(y_{zt}) \in \mathbb{R}$ where F is a cdf that potentially could have heavy tails like a Cauchy distribution.

11 Conclusion

The main research question in this paper is, whether the installation rate of solar PV technology is affected by social effects as measured by the installed base in the immediate local environment. The installed base thereby refers to the cumulative number of solar PV installations within a neighbourhood by the end of a particular month and serves as a measure of social effects from spatially close households. The analysis is based on installation data that has been collected by Ofgem since the introduction of the FiT in April 2010. The econometric panel data model specifies the postcode district-month as the smallest unit of observation. Besides the lagged installed base and month dummies, the main panel equation includes time-varying fixed effects to account for potential self-selection into neighbourhoods and for correlated unobservables that are constant within neighbourhood, but vary over time (i.e. across quarters). Exploiting the time lag between adoption decision and installation, a first difference estimate yields unbiased and consistent estimates of the social effects of interest. Further model specifications allow for a time-varying installed base effect and consider different lags of the installed base as well as different outcome variables and different geographical areas for robustness. In a last specification differences of the social effects across distinct groups of the population are analysed.

The results suggest small, but positive and significant social effects: one more solar PV panel in a postcode district increases the number of new adoptions per owner occupied households in a given month by $7.48e^{-06}$. At the average installation rate within the neighbourhoods, this is equivalent to a one percent increase in the rate to install a solar PV panel. At the average number of 6,629 owner-occupied households within a postcode district, it implies that one more solar PV panel in the neighbourhood increases the number of new installations in the neighbourhood by 0.005. This is obviously and as expected a very small effect. It would require around 200 additional solar panels in a postcode district, for social effects alone to be strong enough to cause one further installation within the neighbourhood. The installed base elasticity at the average installed base of 68 installations within a neighbourhood and the average installation rate of 0.0007 is 0.71. These results illustrate that the social effects as measured by the installed base are very small, but exist and can promote adoption. Especially community projects that involve a high number of installations could hence promote diffusion.

The social effects vary across months and overall diminish over time. Moreover, social spillovers on the postcode district level are found to be stronger than on a higher geographical level, the local authority level. Remarkably, relatively affluent (non-deprived) neighbourhoods show a less pronounced installed base effect. This might result from the fact that those households are early adopters and hence the learning from others is less important. Higher educated neighbourhoods on the other hand show stronger installed base effects than neighbourhoods with on average lower educated populations.

This paper contributes to previous literature in performing the first econometric analysis of the diffusion of solar PV technology within the UK. In particular, it delivers empirical evidence, in how far the adoption behaviour of others drives diffusion. The analysis is based on a remarkably recent and granular solar PV installation dataset of the UK. The results can be exploited for targeted marketing and resource allocations for the stimulation of future adoption. For example, increasing the visibility of the panels could increase the rate of adoption and the use of demonstration sites could have positive effects on the adoption of green technologies.

Nevertheless, the analysis has its limitations. Firstly, social effects are assumed to spread within neighbourhoods as defined by the postcode districts or local authorities, only, while spillovers across the neighbourhood borders are ignored. Spatial econometric methods, for example, could be employed to allow for more diverse spillover effects. Another limitation is the aggregation to the neighbourhood level. Future research should make use of household level covariate data to further analyse the mechanisms underlying the adoption behaviour and thus the installed base effect. In context of the National Energy Efficiency Database (NEED) the Department for Energy and Climate Change (DECC) is currently creating a database that matches the solar PV installation data with household characteristics. Lastly, if there is inertia in the decision process, the consideration of a partial adjustment process in the model might be useful. However, overall, this paper delivers a useful first highly disaggregated analysis of the impact of social effects on solar PV adoption, that can be extended in future research.

References

BANDIERA, O. AND I. RASUL (2002): "Social Networks and Technology Adoption in Northern Mozambique," STICERD - Development Economics Papers 35, Suntory and Toyota International Centres for Economics and Related Disciplines, LSE.

BASS, F. M. (1969): "A New Product Growth for Model Consumer Durables," *Management Science*, 15, 215–227.

BEISE, M. (2004): "The International Adoption of Photovoltaic Energy Conversion Is Japan a Lead Market?" Discussion Paper Series 153, Research Institute for Economics & Business Administration, Kobe University.

BOLLINGER, B. AND K. GILLINGHAM (2012): "Peer Effects in the Diffusion of Solar Photovoltaic Panels." *Marketing Science*, 31, 900–912.

DUFLO, E. AND E. SAEZ (2002): "The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment," NBER Working Papers 8885, National Bureau of Economic Research, Inc.

GUIDOLIN, M. AND C. MORTARINO (2010): "Cross-country diffusion of photovoltaic systems: Modelling choices and forecasts for national adoption patterns," *Technological Forecasting and Social Change*, 77, 279–296.

MANCHANDA, P., Y. XIE, AND N. YOUN (2008): "The Role of Targeted Communication and Contagion in Product Adoption." *Marketing Science*, 27, 961–976.

MANSKI, C. F. (1993): "Identification of Endogenous Social Effects: The Reflection Problem," *The Review of Economic Studies*, 60, pp. 531–542.

NARAYANAN, S. AND H. S. NAIR (2011): "Estimating Causal Installed-Base Effects: A Bias-Correction Approach," Research papers, Stanford University, Graduate School of Business.

RODE, J. AND A. WEBER (2011): "Knowledge Does Not Fall Far from the Tree - A Case Study on the Diffusion of Solar Cells in Germany," ERSA conference papers ersa11p497, European Regional Science Association.

ROGERS, E. M. (1962): *Diffusion of Innovations.*, Glencoe: Free Press.

RYNIKIEWICZ, C. (2012): "Gouvernance participative et choix energetiques : initiatives emergentes et institutions necessaires pour des territoires viables," *CNRS*.

SACERDOTE, B. (2001): "Peer Effects With Random Assignment: Results For Dartmouth Roommates," *The Quarterly Journal of Economics*, 116, 681–704.

SNAPE, R. AND C. RYNIKIEWICZ (2012): "Peer effect and social learning in micro-generation adoption and urban smarter grids development?" *Network Industries Quarterly, 14 (2&3).*, ISSN ISSN 1662-6176, pp. 24–27.

Watson, J. and P. Devine-Wright (2011): "Centralisation, decentralisation and the scales in between." *The Future of Electricity Demand: Customers, Citizens and Loads.*, Cambridge University Press, pp. 542–577.

Wstenhagen, R. and M. Bilharz (2006): "Green energy market development in Germany: effective public policy and emerging customer demand," *Energy Policy*, 34, 1681 – 1696.

Zhang, Y., J. Song, and S. Hamori (2011): "Impact of subsidy policies on diffusion of photovoltaic power generation," *Energy Policy*, 39, 1958 – 1964.

A Appendix

Model with Time-Constant Unobserved Heterogeneity α_z

For robustness, random effects GLS and fixed effects within-group estimations are performed based on a standard panel model with time-constant fixed effects as specified in equation 25. As the analysis is performed on the postcode district level, postcode district effects, α_z, are included. A Hausmann test leads to the rejection of the random effects assumption with a χ^2 statistic of 515.56 and p-value of 0.000.

$$y_{zt} = \alpha_t + \beta \cdot b_{zt-3} + \alpha_z + \epsilon_{zt} \tag{25}$$

Table A.1: Results withtime-constant unobservables. As before, POLS and the within-gropu/fixed-effects estimator seem downwards biased. However, the first difference estimate seems downwards biased now, too, which supports the argument, that controlling for time-constant unobservables is not enough to address the confoundedness resulting from correlated unobservables.

Variable	POLS	RE	WG	FD
Installed Base (L.3)	1.93e-06***	5.64e-07***	-1.43e-06***	3.96e-06*
	(1.20e-07)	(8.98e-08)	(1.24e-07)	(2.35e-06)
Observations	73,887	73,887	73,887	71,648
R-squared	0.145		0.154	0.087
Number of zips		2,239	2,239	

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

As before, POLS and the within-group/fixed-effects estimator seem downwards biased. The bias of the within-group estimator is now more pronounced, as the mean-differenced third lag of installed base is now correlated with the mean differenced errors (the average of all variables is now calculated across all time periods, not only across the observations within one quarter). This is similar to the standard problem of within-group estimators in dynamic panels.

The first difference estimate is lower than in the specification with postcode district quarter fixed effects. This supports the argument, that controlling for time-constant unobservables is not enough to address the confoundedness resulting from correlated unobservables. Therefore, the specification with quarter-zip code effects seems to be more suitable.

Consistency of the Within-Group Estimator

Proposition 1. *The within-group estimator is consistent, if the time lag l between adoption decision and the installation of the solar PV panel is larger than 2, i.e. $l > 2$.*

Consider the general form of the within-group estimator after rearranging.

$$\lim_{N\to\infty} (\hat{\beta}_{WG} - \beta) = \frac{E[(\widetilde{b_{zt-l} - \overline{b}_{zlq}})(\epsilon_{zt} - \overline{\epsilon}_{zq})]}{E[(\widetilde{b_{zt-l} - \overline{b}_{zlq}})^2]} \tag{26}$$

Where $(\widetilde{b_{zt-3} - \overline{b}_{z3q}})$ is the residual from a regression of the mean-differenced installed base on the mean-differenced time-dummies. Multiplying out and omitting the *tilde* for simplicity, yields:

$$\lim_{N\to\infty} (\hat{\beta}_{WG} - \beta) = \frac{E[b_{zt-l}\epsilon_{zt}] - E[\overline{b}_{zlq}\epsilon_{zt}] - E[b_{zt-l}\overline{\epsilon}_{zq}] + E[\overline{b}_{zlq}\overline{\epsilon}_{zq}]}{E[(b_{zt-l} - \overline{b}_{zlq})^2]} = \frac{A}{B} \tag{27}$$

Now consider the terms in the nominator separately.Firstly, $\lim_{N\to\infty} E[b_{zt-l}\epsilon_{zt}] = 0 \forall l \geq 1$ as b_{zt-l} is by construction correlated with all previous errors, $\epsilon_{zt-L} \forall L \geq l$: all previous installations $Y_{zt-L} \forall L \geq l$ enter b_{zt-l}. However, b_{zt-l} is uncorrelated with all future errors, as ϵ_{zt} is *i.i.d* by assumption. Therefore, as long as the lag length $l \geq 1$, the first term converges to zero. Further, the second term, $\lim_{N\to\infty} E[\overline{b}_{zlq}\epsilon_{zt}] \neq 0 \forall l < 3$. If the considered lag length is $l < 3$, not all the observations entering $\overline{b}_{zlq}$ lie in quarter $q - 1$. If $l = 2$, for example, the second lag of the installed base in month 3 of any quarter q falls into q. All ϵ_{zt} lie in q. In month 4, for example (pertaining to the first month in the second quarter), $\overline{b}_{z2q}$ is calculated as average of the second lag of the installed base in month 4, 5 and 6. All these lags refer to the installed base in months 2, 3 and 4 respectively and hence lie in quarter 1 and 2. The neighbourhood quarter mean hence includes the installed base observation of the first month of the quarter and $\overline{b}_{z2q}$ is correlated with the error in the first month of q. For comparable reasons the third term does not converge to zero either, if the lag length is smaller than 3. $\lim_{N\to\infty} E[b_{zt-3}\overline{\epsilon}_{zq}] \neq 0 \forall l < 3$. This term consists of the third lag of the installed base and the neighbourhood-quarter mean of the error. If $l < 3$, not all lags of the installed base, b_{zt-l}, lie in $q - 1$. However, $\overline{\epsilon}_{zq}$ is calculated based on the errors in q. If $l = 2$ for example, in any third month of a quarter q, b_{zt-l2} pertains to the first month of q and is hence correlated with $\overline{\epsilon}_{zq}$ that includes the errors of all past and future errors within the quarter q. Lastly, $\lim_{N\to\infty} E[\overline{b}_{z3q}\overline{\epsilon}_{zq}] \neq 0 \forall l < 3$. This term refers to the correlation of the neighbourhood quarter mean of the lagged installed base with the neighbourhood quarter mean of the error term. For $l < 3$ $\overline{b}_{z3q}$ is calculated based on observations in $q - 1$ and q, while $\overline{\epsilon}_{zq}$ is based on ϵ_{zt} within q. As an example, consider month three of any quarter. The postcode-quarter mean includes all error observations of this quarter, i.e. also the error of the first month of this quarter. Consider the postcode district mean of the lagged installed base. If $l = 2$, for example, in month 3 of the quarter the mean includes hence an installed base observation of month 1 of the quarter, which is correlated with the contemporaneous error in this month.

Average Neighbourhood Statistics

Table A.2: Summary statistics for the 2,239 considered postcode districts. Percentages refer to the percentage of *households* in case of number of cars, deprivation, education, social class and to *people* otherwise.

Variable	Obs	Mean	Std. Dev.	Min	Max	60th Percentile
Population Size	2239	25006.94	16331.41	136	154233	27992
Area in Hectars	2239	6743.75	9798.42	2.33	85165.67	4247.69
Population Density	2239	18.91	27.29	0.1	233.9	13.2
# HH	2239	10417.01	6676.33	54	62010	11498
# owner-occupied HH	2239	6629.24	4176.46	12	32132	7158
Ratio active/inactive	2239	0.81	0.05	0.49	0.99	0.82
Mean age	2239	40.64	4.22	22.4	55.2	41.7
Median age	2239	41.08	6.21	20	62	43
% aged 24 to 45	2239	0.32	0.08	15.03%	80.39%	32.28%
% aged 65 plus	2239	0.18	0.06	0.43%	43.07%	19.14%
% 0 car	2239	22.93%	0.14	2.50%	81.73%	22.73%
% 1 car	2239	41.84%	0.05	14.42%	54.61%	43.39%
% 2 cars	2239	26.61%	0.09	1.46%	48.99%	30.22%
% 3 cars	2239	6.27%	0.03	0.00%	15.67%	7.01%
% 4 cars	2239	2.36%	0.01	0.00%	8.96%	2.54%
% Deprived in 0 Dim	2239	43.29%	0.09	17.66%	64.66%	46.11%
% Deprived in 1 Dim	2239	32.81%	0.03	24.86%	57.27%	33.12%
% Deprived in 2 Dim	2239	18.70%	0.05	7.61%	34.93%	19.74%
% Deprived in 3 Dim	2239	4.74%	0.03	0.00%	18.29%	4.78%
% Deprived in 4 Dim	2239	0.46%	0.00	0.00%	3.05%	0.43%
% Education 0	2239	22.38%	0.07	1.39%	48.64%	23.63%
% Education 1	2239	13.14%	0.03	2.76%	21.16%	14.07%
% Education 2	2239	15.49%	0.02	3.33%	24.77%	16.41%
% Education 3	2239	3.71%	0.01	0.00%	9.52%	4.09%
% Education 4	2239	27.83%	0.10	7.20%	77.03%	29.22%
% Apprenticeship	2239	12.24%	0.03	5.85%	53.91%	12.07%
% No travel to work	2239	4.30%	0.03	0.36%	25.58%	4.42%
% Public Transport	2239	8.52%	0.09	0.00%	53.84%	6.98%
% Private Transport	2239	42.64%	0.11	2.61%	63.33%	47.48%
% Travel with Bike/Foot	2239	8.70%	0.05	0.63%	50.00%	8.37%
% Males	2239	49.18%	0.01	44.71%	64.24%	49.15%
% Social Class AB	2239	23.57%	0.10	3.22%	71.50%	24.90%
% Social Class C1	2239	30.57%	0.04	12.86%	65.62%	31.43%
% Social Class C2	2239	21.74%	0.06	1.05%	43.24%	23.44%
% Social Class DE	2239	24.12%	0.09	1.05%	67.14%	25.13%
# Unshared Dwellings	2239	10381.51	6639.54	54	61589	11451
# Detached Houses	2239	2371.65	1821.74	0	11279	2491
# Semi-Detached Houses	2239	3262.12	2545.94	0	17843	3373
# Flats	2239	2139.19	2978.58	0	34053	1687
# Shared Dwellings	2239	35.49	95.22	0	1293	9

Table A.3: Testing autocorrelation in the First Difference Model. The results suggest that the order of autocorrelation in equation 17 is low enough for consistent estimation of the installed base effect. * indicates significance, *no* indicates the autocorrelation is not significant.

Tested lag of $\Delta\hat{\epsilon}$	Autocorrelation	α= 0.05	α= 0.01
L.	-0.7523	*	*
L2.	0.0295	*	*
L3.	-0.004	no	no
L4.	-0.0200	*	no

Table A.4: Time-Varying Installed Base Effect (interactions with quarter dummies). The results suggest a time-varying installed base effect. While social effects appear to increase initially, they decrease towards the end of the considered period and eventually even get negative

Variable	FD_{cl}
Installed Base (L.3)	-2.32e-05**
	(9.12e-06)
Q3 x Installed Base (L.3)	-1.46e-06
	(9.99e-06)
Q4 x Installed Base (L.3)	2.42e-05*
	(1.28e-05)
Q5 x Installed Base (L.3)	3.21e-05***
	(9.04e-06)
Q6 x Installed Base (L.3)	4.96e-05***
	(9.55e-06)
Q7 x Installed Base (L.3)	4.81e-05***
	(1.09e-05)
Q8 x Installed Base (L.3)	4.87e-05***
	(9.11e-06)
Q9 x Installed Base (L.3)	2.78e-05***
	(9.09e-06)
Q10 x Installed Base (L.3)	-3.39e-05***
	(1.02e-05)
Q11 x Installed Base (L.3)	-1.01e-05
	(1.12e-05)
Q12 x Installed Base (L.3)	-1.53e-05
	(1.12e-05)
Constant	6.19e-05***
	(1.20e-05)
Observations	49,258
R-squared	0.007

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1