Projection Bias in the Decision to Go Solar: Evidence on Costly Cancellations

Yanjun (Penny) Liao*

December 15, 2016

Abstract

This paper studies the effect of short-run weather fluctuations on solar panel adoption in California. I find that customers whose sign-up for solar panels is followed by bad weather are more likely to cancel their contract. In contrast, non-residential customers appear unaffected by short-term weather conditions. These results are consistent with residential customers suffering from projection bias – that is, they rely too heavily on transient conditions when predicting long-run utility. I show further evidence that this effect is mainly driven by mistaken predictions about future solar panel performance rather than electricity demand. There is also suggestive evidence that having more neighbors with solar panels does not mitigate projection bias, and that marginal customers are more susceptible to it. The results add to the emerging field evidence on projection bias, particularly by carefully considering the dynamics of the entire purchase-return cycle. The results also have efficiency implications for current subsidy programs on solar panel adoption.

Keywords: projection bias, solar adoption, weather.

JEL Classification: D03, D12, D81.

*University of California, San Diego. Email: yal005@ucsd.edu. I am especially grateful to Mark Jacobsen, Joshua Graff Zivin, Julie Cullen, and Charles Sprenger for their invaluable advice and support. I also thank Richard Carson, Tom Chang, Gordon McCord, Sally Sadoff, David Victor, and seminar participants at UCSD for helpful comments. I am also grateful to the California Solar Initiative (CSI) for making their data publicly available. All errors are my own.

1 Introduction

Many important economic decisions require the consumer to estimate future utility. When purchasing a durable good, for example, the consumer needs to trade off her future utility against the present cost of purchasing it. Standard economic theory assumes that these estimates are made accurately, but emerging empirical evidence suggests otherwise.[1] For instance, people tend to be overly influenced by the current state when they make predictions of future utility. This effect is called "projection bias" (Loewenstein et al., 2003). It not only affects small decisions such as shopping for grocery (de Magistris and Gracia, 2016), but could even affect high-stake decisions such as college choice and car purchase (Busse et al., 2014; Simonsohn, 2010).

In this paper, I investigate whether residential customers exhibit projection bias in solar photovoltaic (PV) adoption. If current weather is unusually good for real-time return from solar panels, a consumer with projection bias would overestimate the lifetime return, while bad weather would lead to the opposite effect. In particular, the change in valuation due to weather fluctuation is best captured by cancellations of signed contracts. Therefore, I test for projection bias by examining how weather fluctuations after signing up for solar affect the probability of the customer canceling the contract.

There are several advantages to studying projection bias using cancellations of solar contracts. First, solar panels have a long lifespan. Given that their lifetime typically lasts more than 25 years, it is extremely difficult for short-run weather fluctuations to affect the return to such investments and lead to cancellation by a rational customer. Second, since the cancellation decisions are made prior to installation, cancellations are not likely driven by consumers rationally learning about the profitability of solar panels. Third, as an outcome of interest, *cancellation* reflects a real change in demand that entails substantial economic consequences. Projection bias suggests that on sunny days, consumers are more likely to purchase solar panels. It is tempting, therefore, to estimate whether there are more *purchases* on sunny days. However, such a pattern is also consistent with a particularly sunny period "harvesting" solar purchases that would have occurred rationally, which implies no real behavioral change and hence minimal welfare impacts. In contrast, cancellations

[1] See DellaVigna (2009) for a review.

reflect permanent behavioral change. It is costly – previous site visits and assessments by the solar company are wasted; and the household has incurred a hassle cost and a penalty for breaching the contract. Due to this costly nature, a household is not likely to enter into another solar contract anytime soon after they cancel the last one.

Using transaction-level administrative data from the California Solar Initiative (CSI), I find that residential customers are more likely to cancel their solar contract when they experience worse weather after signing up. Specifically, a one-standard-deviation decrease in solar radiation is associated with 8% increase in cancellations, which in turn account for 12.06% of all CSI applications. This effect is identified using variations within season by controlling for quarterly fixed effects and other potentially confounding economic factors at the monthly level. It is also robust across a variety of specifications. This result is not consistent with the behavior of fully rational or even present-biased economic agents, unless discount rates are implausibly large and weather shocks are extremely persistent. Moreover, I also find evidence that non-residential customers are much less affected by current weather. This suggests that they are less susceptible to the psychological mechanism that affects residential customers, possibly because they usually go through formal investment calculations when deciding whether to go solar.

To understand how weather information is incorporated into the household's valuation of solar panels, I model this decision when the household is subject to projection bias. The model shows that the effects of weather can be broken down into a solar production channel and an energy demand channel. I construct an index to represent each channel using engineering models and weather variables, and estimate the effects of these indices jointly. The result suggests that the changes in cancellations are primarily driven by concerns over solar panel performance. I also show evidence that negative update in weather conditions relative to the pre-contract period plays an important role, and the responses are particularly strong for households who sign up following a period of above-average good weather. All of these lend further support to the mechanism of projection bias.

Lastly, I explore heterogeneous responses along various dimensions. I find evidence of peer effects, namely that customers in areas with high market penetration are less likely to cancel their contracts. This is consistent with the mechanism found in Bollinger and Gillingham (2012). Such peer effects, however, do not appear to alleviate responses to short-run weather. In addition, I find that some types of customers are more susceptible to projection bias than others. The effects are stronger for customers with third-party ownership, those who might be rushed into the contract by imminent incentive rate decrease, and who live in more urban neighborhoods.

These findings not only advance our understanding of projection bias, but also open up new avenue for cost-effective public policy on the solar industry. Currently, there are strong policy interests in promoting solar power, given its clear environmental advantage over traditional electricity generation. In the U.S., there are various tax credits and subsidies offered at the federal, state, and local levels. However, prior studies on such programs have found that they are not cost-effective.[2] Exploiting behavioral patterns such as projection bias could be an easier and cheaper way to increase solar panel uptake.

This paper is closely related to a growing literature in behavioral economics that documents "smoking gun" evidence of projection bias in the field (Busse et al., 2014; Chang et al., 2016; Conlin et al., 2007; Simonsohn, 2010). Most of them focus on how consumers are induced into purchase by favorable conditions on the purchase date. This paper instead highlights the importance of weather updates, especially negative ones, in changing the customers' minds. Furthermore, most previous studies do not distinguish between *mistaken utility* and *mistaken beliefs*. That is, it is not clear in their contexts whether the consumer is mistaken about her utility in a future state, or how likely today's weather will occur in the future. This paper provides a clear interpretation that the latter is the case, which allows interventions to be better targeted.

This paper also contributes to a broader literature that focuses on understanding consumer behaviors regarding energy efficiency investments.[3] As the real-time profitability of these investments also depends on current weather conditions, this paper suggests that their customers might be subject to similar behavioral mechanisms.

[2]Hughes and Podolefsky (2015) find program-induced behavioral change does not justify the cost of CSI because many claimants are inframarginal. Bollinger and Gillingham (2014) find that non-appropriable learning-by-doing in the solar industry similarly do not justify the cost. Furthermore, Borenstein and Davis (2015) notes the distribution of benefits of such programs are highly regressive.

[3]See Allcott and Wozny (2014) for a recent review.

2 Weather in the Decision to Go Solar

2.1 Projection Bias

Projection bias refers to the tendency for people to exaggerate the degree to which the future state will resemble current state. A simple theory and suggestive evidence is first laid out by Loewenstein et al. (2003). The theory is later adapted to the purchase of durable good by Conlin et al. (2007) and Busse et al. (2014). This framework distinguishes between actual utility $u(G, s_1)$ from using good G in period 1 (with state s_1) and the projection of the same utility from current period (with state s_0), denoted $\tilde{u}(G, s_1|s_0)$. A simple representation of over-influence by current state is

$$\tilde{u}(G, s_1|s_0) = (1-\alpha)u(G, s_1) + \alpha u(G, s_0), \tag{1}$$

where α is between 0 and 1. When α is nonzero, the projected utility is different from the actual utility and biased towards the period 0 utility. At period 0, the projected total utility from purchasing a good in period 1 and using it through period T is given by

$$\begin{aligned}\tilde{U}(G|s_0) &\equiv \sum_{\tau=1}^{T} \delta^{\tau} \tilde{u}(G, s_\tau|s_0) \\ &= (1-\alpha)U(G) + \alpha\delta\Big(\frac{1-\delta^T}{1-\delta}\Big)u(G, s_0),\end{aligned} \tag{2}$$

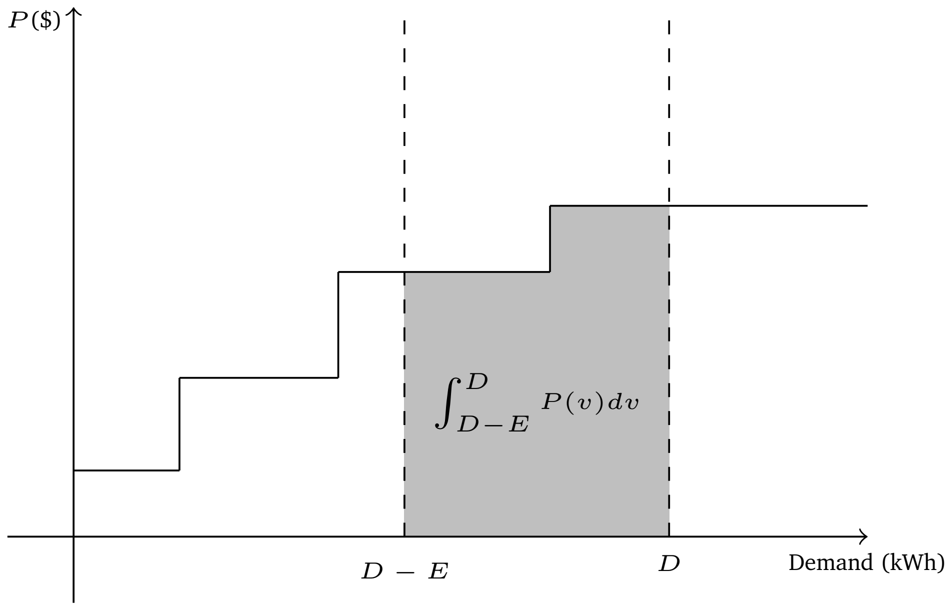
where the second equality follows from Equation (1). The total projected utility is also biased by current utility with some multipliers due to discounting.

In the context of solar adoption, I consider states as different weather conditions. Households decide whether to go solar by weighing the projected total utility against a fixed cost. Therefore, although current weather does not affect the true utility $U(G)$ at all, it affects the projected utility and hence the decision of marginal customers.

2.2 Channels: Solar Production vs. Energy Demand

The benefit from a solar PV system in each period depend on the total electricity bill savings in that period, which in turn depends on how much power it generates as well as how expensive the

Figure 1: Benefit from Solar PV System



displaced electricity is.[4] I assume that the current period utility equals to the bill savings under net metering, which takes the following form:

$$u(G, s_0) = \int_{D(s_0)-E(s_0)}^{D(s_0)} P(v)dv. \tag{3}$$

$D(s_0)$ is gross electricity demand, $E(s_0)$ is solar power production, and $D(s_0) - E(s_0)$ is the net demand that eventually enters into the price schedule. $P(\cdot)$ is the marginal price schedule of electricity and is weakly increasing in electricity usage due to increasing-block pricing scheme in California.[5] As illustrated in Figure 1, Equation (3) is essentially measuring the shaded area for given electricity demand D and solar power production E.

For simplicity, assume the scaler s_0 is a sufficient statistic that summarizes the state-of-the-world in period 0. It is straightforward to see that s_0 affects projected total utility through both

[4]While some solar customers have green preference in reality, it is ignored here for simplicity. The model assumes that projection bias mainly takes the form of mistaken beliefs about future weather patterns, rather than mistaken utility.

[5]See Ito (2014) for more details of electricity pricing in California. Note that we can easily use $P(\cdot)$ to denote any perceived price schedule if it differs from the actual marginal price schedule. It does not affect the result below.

$D(s_0)$ and $E(s_0)$. More formally, we can derive

$$\frac{d\tilde{U}(G|s_0)}{ds_0} = \kappa\big[\underbrace{P|_{D(s_0)-E(s_0)} \cdot E'(s_0)}_{\textit{Solar Production Channel}} + \underbrace{\big(P|_{D(s_0)} - P|_{D(s_0)-E(s_0)}\big) \cdot D'(s_0)}_{\textit{Energy Demand Channel}}\big]. \tag{4}$$

Equation (4) highlight the two channels through which weather might affect the decision to go solar. The first term represent the solar production channel. That is, when the weather does not favor solar power production, customers might estimate worse overall performance of solar panel. The second term represent the energy demand channel. During periods of greater energy demand (particularly hot or cold days), customers might think they have greater future energy demand and hence higher utility from the solar PV system. These two channels do not necessarily go in the same direction. For example, heat reduces solar panel productivity, but increases electricity demand for cooling.

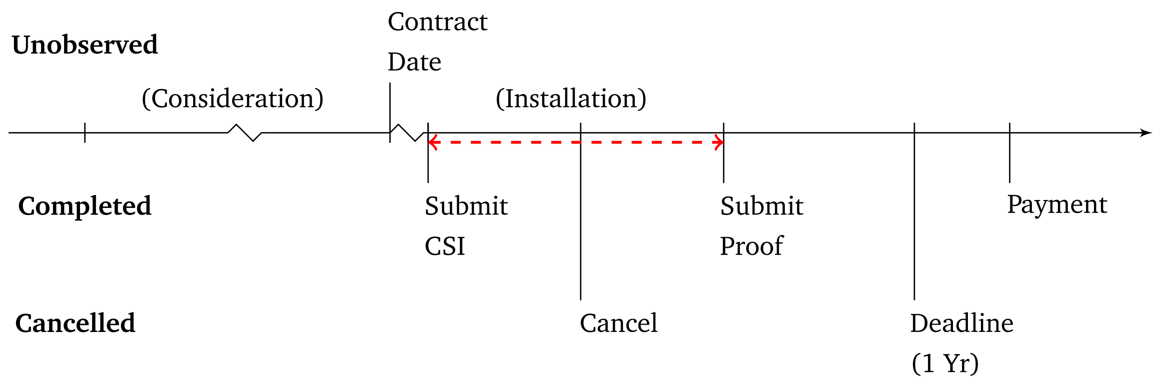
Equation (4) also reveals that the production channel will shut down when the solar panels generate sufficient power to cover all electricity consumption; and the demand channel will stop working when solar generation is not enough to move to a lower price tier. Both conditions are rare in reality. And because customers often respond to average rather than marginal price (Ito, 2014), both channels are likely to be at work.

3 Policy Background and Data

3.1 The California Solar Initiative (CSI)

The CSI is a solar rebate program in California for the customers of the three investor-owned utilities (IOUs), covering most of California. It funds solar on existing homes, commercial, agricultural, government and non-profit buildings. Overseen by the California Public Utilities Commission (CPUC), the program has a total budget of $2.167 billion between 2007 and 2016 and a goal to install approximately 1,940 MW of new solar generation capacity. Over these years, California continues to be the leading solar market in the United States, accounting for about half of the nationwide installed capacity.

Figure 2: Timeline of a CSI application



An important feature of the CSI program is that the rebate rate is designed to decrease automatically based on the total volume of solar megawatts (MWs) with confirmed project reservations (see Figure A1 in the Appendix). This design is intended "to transform the market for solar energy by reducing the cost of solar" through learning-by-doing. Indeed, average cost has been rapidly decreasing, which was partially attributed to learning-by-doing by prior studies (Bollinger and Gillingham, 2014).

In this paper, I use the administrative dataset from CSI as a main data source, which contains the universe of CSI applications, regardless of whether they are eventually completed. They represent about 80% of solar projects in California before 2012, and much less afterwards as the incentive weakens. For each rebate application, the dataset provides detailed information on system characteristics such as the CSI rating (size), total cost, incentive amount, whether the system is third-party-owned (TPO), and the sectors for the host and owner.[6] More importantly, it identifies the zip code and relevant dates for each application, which can be matched to meteorological data at a fine geographical scale.

Figure 2 shows the typical time line for a CSI application, where unobservable activities are plotted above the line and observed ones below. The process begins with the households spending an unknown amount of time considering about solar investment. Once they are in contact with

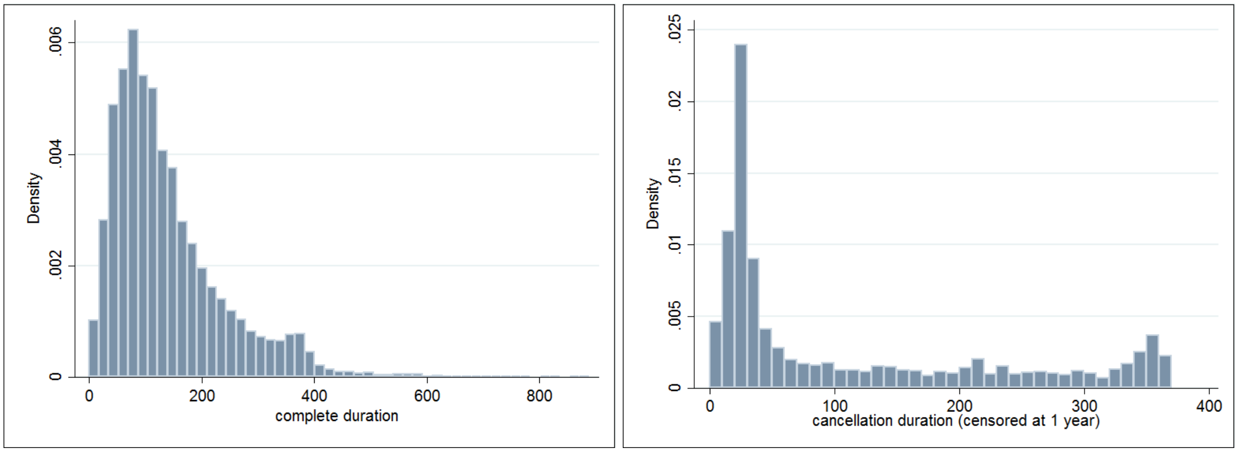
[6]The host refers to where the system is physically located, or the customer. For systems with third-party ownership (TPO), the host is different from the owner, the latter being the solar company.

a solar company, the latter typically conducts a site inspection and provides detailed information on estimated bill savings, design options and quotes, etc. These procedures lengthen the process, especially when households shop around for the best offer. When they make up their minds, the household signs a contract with the solar company, the date of which is unobserved. However, the CSI dataset provides the date when a rebate application is submitted to CSI, at which point a signed contract is required by program rules. The time between the contract and the submission date is likely to be short, because customers have an incentive to submit as soon as possible due to uncertainties in when the incentive rate jumps to a lower step. It is also confirmed by CSI staff that the submission date is the best proxy for the contract date.

After submission, a rebate is reserved for the household at the current incentive level. The household is given a year to complete the installation and submit proof to claim the rebate. While installation date is unobserved, the submission of proof is. The period between application and proof submission is referred to as "completion duration" henceforth. About 12.06% of the applicants in the dataset, however, eventually cancel their projects. For these applications we observe a cancellation date, but with a major complication: about two thirds of the cancellations are recorded beyond the one-year deadline. This might be because they canceled the contract without reporting to the CSI administrator and went undiscovered until the deadline. In the left panel of Figure 3, I plot the distribution of complete duration, which is what we would expect if most dates are reported accurately. There is mild bunching around one year, which might be driven by administrator checking with applicants that have not reported back. In the right panel, I plot the distribution of cancellation duration, censored at one year, because dates beyond that are uninformative and entirely driven by administrative procedure. The density of reported cancellations have an earlier peak than completed ones, and quickly flattens out to a low level after a month or two. It is, however, impossible to know what the counter-factual distribution would look like, had all canceled projects been reported without delay.

In the data, we observe two outcomes that might be affected by short-run weather fluctuations: the purchase itself and return/cancellation conditional on purchase. I choose the latter as the outcome of interest for two main reasons. First, the consideration period for purchase is unob-

Figure 3: Distribution of Completion and Cancellation Duration



served and may vary widely, but under projection bias the estimated utility from solar PV system can change according to weather in each new day. This distinguishes solar purchase from car or college choices examined in prior studies, where mainly weather on the day of test drive or college visit matters. Second and more importantly, cancellation of signed contract represent a real change in demand that entails substantial economic consequences. California state law requires a three-day cool-down period for customers to freely cancel their solar contracts, but cancellations afterwards can legally be subject to penalty. A costly cancellation thus reflects a change in valuation for the solar panels, and it also implies that a household is not likely to enter into another solar contract anytime soon afterwards. In contrast, a pattern of more purchase on sunny days might simply reflect short-run postponement with minimal welfare implication.

I assume that the cancellation decision is made within the same time frame as completion: as discussed above, the recorded cancellation duration is uninformative. This assumption is reasonable given that solar companies should follow the same schedule routine with all customers. Specifically, I consider weather in a post-contract period of 113 days as relevant for every project, as 113 days is the median duration for completed projects.

However, this assumption of a fixed period length is almost surely inaccurate, and measurement errors are inevitable. This has three implications. First, the estimates in this paper are likely to be lower bounds of the true effect of weather on people's decision to cancel solar contracts. Second, we would expect the estimates based on any period length within a reasonable range to be similar because moving within this range reduces measurement errors for some applications while exacerbates those for others. Third, the estimate should get seriously attenuated when it is based on a period that is either too long or too short, and be even more so when we move further away. In short, due to unknown and potentially heterogeneous lengths of the relevant period, we expect the estimates to display an inverse U-shape with respect the the length of periods that they are based on, and the peak estimates might still be lower bounds of the true effects. In the empirical analysis, I provide evidence consistent with these testable implications in a series of robustness checks.

Another concern of using cancellations conditional on purchase is that some unnecessary contracts might be selected into the sample as the result of good weather before sign-up. In that case, mean-reversing weather trends might introduce mechanical correlation between the probability of cancellation and post-contract weather. Note that this is still a phenomenon consistent with projection bias, but the interpretation is more complicated. This is difficult to test due to greater uncertainties in the pre-period length. In the empirical section, I shed light on this issue by analyzing whether favorable pre-period weather induces customers into solar contracts, or is associated with larger system size. I also explore how the dynamics of pre- and post-period weather affect cancellations.

3.2 Solar and Weather Data

The weather data in this paper comes from two main sources. Station-level data on daily maximum temperature, wind speed, and precipitation are taken from the Global Surface Summary of the Day (GSOD) dataset at NOAA, a common data source used in numerous other studies. I exclude weather stations with elevation above 1,500 meters, which leaves 191 stations operating in the sample period to be matched with 1,428 zip codes. Based on the center coordinates of each zip code, I find all the weather stations within 100 kilometers (62 miles), and calculate the mean of observations from these stations, weighted by inverse distance. Since there are a much larger number of zip codes than weather stations, a natural concern is that the measures are not very precise for each application, adding to the attenuation bias. Among the main variables, temperature is more smoothly distributed across space and less problematic, while precipitation and wind speed can be highly localized and hence suffer from more severe measurement errors.

Another main variable is surface solar radiation, which is the feature commonly associated with solar panels. I extract daily solar insolation by zip code coordinates from the Prediction of Worldwide Energy Resource (POWER) Project at NASA.[7] This measure is constructed based on satellite data at a 1° latitude by 1° longitude resolution.

I then aggregate the daily observations over a 113-day post-contract period. A simple-minded measure would be the mean for all variables. However, for the decision to cancel a solar project, the distribution of certain weather features might be more important than the average level. For example, consecutive days of drizzle is more damaging than one day of heavy rain that consumes all the clouds in the sky. Moreover, human response to temperature is found to be nonlinear in other settings (Baylis, 2015; Graff Zivin and Neidell, 2014). Here, both the energy production and demand channels predict a nonlinear relationship between temperature and the probability of cancellation especially at the two ends, translated from nonlinearities in solar panel productivity or air conditioner use. Due to these concerns, I represent solar insolation and wind speed in means, precipitation in number of rainy days, and daily maximum temperature in both mean and a set of variables representing number of days in a given temperature bin (e.g. below 40°F, 40-60°F, etc.).

3.3 Summary Measures for Solar Production and Energy Demand

To understand whether the response to current weather is due to projection bias over solar production or energy demand, we need summary measures for each channel.

The estimated energy demand is relatively straightforward to represent. Weather mainly affects energy demand through the need to turn on heat/air conditoner in cold/hot days. Such weather-driven energy demand are conventionally represented by heating and cooling degree days (HDD and CDD), defined as

$$HDD = 1\{Temp < 65\} \cdot (65 - Temp) \quad \text{and} \quad CDD = 1\{Temp > 65\} \cdot (Temp - 65).$$

[7] Solar insolation is the amount of electromagnetic energy (solar radiation) incident on the surface of the earth, measured in MJ/m^2 per day. See Stackhouse Jr et al. (2015) for methodology and accuracy of this measure.

I also define a total degree days (TDD) measure:

$$TDD = CDD + HDD = |Temp - 65|.$$

These measures are constructed at the daily level and aggregated into means over the relevant period.

CDD is a reliable measure of hot weather on electricity use because air conditioners and fans run on electricity, but it provides no information on cold seasons. Another common measure is HDD, but in California it is not a useful representation of electricity use. This is because electricity only accounts for 8.5% of energy use for heating in California homes (EIA, 2016), and winter heating is not needed in many parts of Southern California. This is clearly illustrated by year-round dynamics of the monthly peak demand in the California ISO system (see Figure A2 in the Appendix). The peak demand is much higher in the summer months than the rest of the year, and although there is a small spike around January in some years, the scale is trivial compared to the summer. Nevertheless, I also show results using TDD to account for the small spike in winter.

Solar power production, on the other hand, is a complicated process. The primary factor is intensity of solar radiation, but solar panel efficiency is also affected by heat negatively and by wind positively in a nonlinear relationship. To capture this process, I employ a power conversion model developed by Kleissl (2013), a project funded by CSI RD&D. This model generates predictions of solar panel output for given inputs of weather and system characteristics, and its performance is comparable to the PVWATTS model, the standard calculator of PV performance in the industry. For each CSI application, I generate a series of daily indices by plugging in daily solar insolation, maximum temperature, and wind speed, and then aggregate them over the entire relevant period. Since I keep all other characteristics the same across applications, the variations in the index are entirely driven by how favorable the particular combination of weather variables is to solar production.

3.4 Economic Controls and Demographic Characteristics

Variations in weather are usually considered to be orthogonal to most other variables, but it is less certain in this study since summary of weather in a period is used. Therefore, I also control for economic factors that might drive cancellations, which might or might not be correlated with the weather variables. For this purpose, I collected the following monthly economic variables for the entire California from the FRED database: leading index,[8] unemployment rate, and prime interest rate. Furthermore, I have also obtained the monthly index of consumer sentiment and buying conditions for the West region from the Survey of Consumers.

I have also collected zip-code-level demographics. These variables include race, education, income, housing cost, household size, etc. They are either from the 2010 Decennial Census or 5-year estimates from the 2011 American Community Survey (ACS). I match them to the applications to analyze whether responses to weather are heterogeneous along these dimensions.

4 Econometric Framework

As discussed in Section 2.2, the main implication we test in this paper is whether weather fluctuations in the post-contract period affect the probability of cancellation. Let $Cancel_{izt}$ be an indicator of whether the project is eventually canceled. The main estimating equation is essentially a linear probability model relating the probability of cancellation with weather:

$$Cancel_{izt} = \beta Solar_{zt} + \gamma Weather_{zt} + \eta X_{it} + \delta_z + \lambda_{c\tau} + \epsilon_{izt}. \tag{5}$$

i denotes individual application and z zip code. There are two notations of time: t denotes the application date, or the post-contract period uniquely identified by the application date; and τ is a generic notation for time units. The main variable of interest is $Solar_{zt}$, solar insolation at zip code z in the corresponding post-contract period. $Weather_{zt}$ is a vector of weather variables that households might also respond to, which might be correlated with solar insolation. The coefficients

[8]The leading index for each state predicts the six-month growth rate of the state's coincident index, which in turn measures the current state of economic activity.

on these variables are of interest as well. X_{it} is a vector of controls including system characteristics and economic conditions.[9] $\lambda_{c\tau}$ is a set of county-by-time and time fixed effects (FEs), which will be discussed in more details below. δ_z denotes zip code fixed effects. X_t and δ_z are included in all regressions.

There are a few major choices regarding specification. One is related to the weather variables. With aggregation over time, the correlation between certain weather variables becomes too strong to be included in the regression simultaneously. Solar insolation, for instance, is highly correlated with number of rainy days and temperature. Therefore, for the main regressions on solar insolation, I only control for wind speed and extreme temperature bins. A related and common concern is that correlations between the weather variables might lead to spurious significance. To address these issues, I provide robustness checks with alternative specifications, as detailed in Section 5.

The choice of $\lambda_{c\tau}$ critically determines the nature of variations used for identification. The main purpose of the FEs is to control for non-weather drivers of cancellation rate that might be correlated with weather. In particular, two concerns of such confounding factors are particularly plausible. One is seasonal employment cycles – suppose people are more likely to be fired at the end of the year, leading them to cancel their solar contracts, it will be incorrectly attributed to lower solar insolation level in winter. Another would be business cycle interacting with climatic ones, such as the El Niño Southern Oscillation. In light of these concerns, the preferred specifications in this study use within-year and within-quarter variations.

It should be noted, however, that the benefit of ruling out confounders comes at a cost of leaving out meaningful variations. This trade-off is important because it is unclear as to which dimension of weather fluctuations the households might respond. If they are generally more hesitant about their contracts in the winter but does not distinguish between winters that are more or less sunny, then seasonal fixed effects would eliminate most of the meaningful variations. Moreover, another difficulty arises because this is a decision made over a period and the weather variables are summary statistics of the period. Therefore, a larger portion of variations in the weather variables would be absorbed by the fixed effects than in other contexts. Lastly, it should also be noted that

[9] System characteristics include CSI rating (size), third-party ownership, average out-of-pocket cost. Economic controls are as discussed in Section 3.4.

the fixed effects might exacerbate the attenuation bias due to measurement errors in the weather variables.

To examine this trade-off in a transparent manner, I use six different specifications for $\lambda_{c\tau}$ in the main regressions, going from zip code FEs only to including year FEs, to also controlling for quarters. In two specifications I allow for quarter or year FEs to differ across counties. It should be noted that, when comparing across these specifications, we would expect the estimates to vary although the qualitative pattern should be similar. This is because they use different sources of variations to capture a possibly nonlinear relationship.

In my preferred specification, seasonal economic dynamics specific to the county are controlled for by quarter-by-county FEs, and any trends in solar panel installations or common economic and policy shocks to the entire state are controlled for by year FEs. In another specification, I control for shocks to specific counties in a specific year using county-by-year FEs, in addition to controlling for seasonal factors using quarter FEs. Since it would be too demanding to control for county-by-quarter-of-sample FEs, a potential confounder would be differential policy or economic shocks to specific counties in a specific quarter-year. But in order for that to work, these shocks must also be correlated with the weather deviations from local climate normal in a certain way, which is highly unlikely.

The later sections of the empirical analysis focus on discerning the solar production channel and the energy demand channel, examining more closely selection into contracts due to pre-period weather, as well as exploring whether there are any interesting heterogeneous responses. These analyses are based on different outcome variables and regressors from above, but the threats to identifications are more or less the same. Therefore, similar FE specifications are used in these analyses.

Lastly, since there are overlaps in the weather stations matched with different zip codes, I show standard errors clustered at two levels, zip code and county, for the main results on weather. The number of clusters are 1,428 and 54, respectively. The rest of the results are shown with standard errors clustered at the county level.

5 Results and Discussions

Table 1 presents the summary statistics. For the main analysis, I exclude applications whose host (customer) sector is non-residential, whose CSI rating is above 100, and whose application submission starts after September 1, 2015. After cleaning, there are a total of 154,655 residential applications, of which 18,653 (12.6%) were eventually canceled. The average system cost is $35,377.04, which is substantial compared to most home appliances. The average incentive amount is about one tenth of system cost.

Canceled projects are larger and more expensive than completed ones, but the difference is small. They are slightly more likely to be third-party-owned (TPO). That is, the system is hosted by the customer but owned by the solar company. The mean of contract year is similar, which suggests that the majority of cancellations are not concentrated in a short period or driven by, for example, a special event or policy change at a certain point in time. The simple mean of weather variables are quite close for the two groups as well, but the formal analysis makes use of finer variations than this simple comparison. Canceled projects also appear to be located in zip codes with slightly smaller fraction of whites and college graduates, and with lower median income. But the differences are very small and insignificant.

In the rest of this section, I first establish that residential customers indeed respond to weather fluctuations in the post-contract period, by showing a set of results and robustness checks on various weather variables. Second, I show results that the solar production channel is the main mechanism while energy demand is less of a concern. Next, making further use of the solar production index, I examine in more details the dynamics of purchase and cancellation in response to weather shocks in the pre- and post-contract periods. I also use this index to explore potential heterogeneous responses from residential vs. non-residential systems, as well as across a variety of different system and neighborhood characteristics. Last but not least, I discuss implications of the empirical evidence.

Table 1: Summary Statistics

	All				Completed	Canceled
Variables	Mean	SD	Min	Max	Mean	Mean
A. Application Characteristics						
N		154,655			136,002	18,653
CSI rating	4.86	3.02	.14	99.44	4.82	5.22
cost (USD)	35,377.04	23,829.66	0	1,726,706	35,090.3	37,467.7
incentive (USD)	3,551.24	4,838.65	51	176,442	3,526.86	3,729.01
TPO	.500	.500	0	1	.495	.541
year	2011.11	1.83	2007	2015	2011.09	2011.27
B. Weather						
solar insolation	18.96	5.85	7.05	29.36	18.98	18.81
max temp (F)	74.50	9.40	48.63	107.68	74.53	74.24
#(max ≤ 40)	.0160	.214	0	25	.0157	.0177
#(max $\in (40, 60]$)	13.87	19.90	0	111	13.79	14.48
#(max $\in (60, 80]$)	63.39	27.62	0	113	63.39	63.43
#(max $\in (80, 100]$)	32.47	28.34	0	112	32.55	31.93
#(max > 100)	3.24	9.50	0	112	3.26	3.14
wind speed (knots)	4.99	1.48	1.47	13.86	5.00	4.93
#(precipitation> 0)	23.43	17.46	0	112	23.43	23.45
production index	98.51	31.45	13.81	153.89	98.63	97.66
CDD	2.81	3.94	0	28.70	2.82	2.76
HDD	5.92	4.74	0	28.60	5.91	5.97
TDD	8.73	3.79	1.48	28.70	8.73	8.73
C. Economic Controls						
leading index	1.41	1.42	–3.69	3.04	1.39	1.55
unemployment	10.02	1.78	5	12.2	10.01	10.15
prime interest rate	3.56	1.04	3.25	8.25	3.57	3.46
buying condition	128.07	13.12	88	160	128.07	128.17
D. NEM Interconnection Data						
(zip code level)						
current installed base	435.66	491.10	1	3241	–	–
installed penetration	.049	.040	.000	.314	–	–
E. Demographics (Zipcode Aggregate)						
white (%)	67.30	15.88	5.5	100	67.53	65.65
bachelor degree (%)	36.15	17.87	0	100	36.29	35.12
median income	79,414.95	27,052.83	2500	240,833	79,617.02	77,940.58
monthly housing cost	2,170.04	655.63	131	4,000	2,172.93	2,148.93
mean household size	2.86	.460	0	5.18	2.85	2.88
urban (%)	.915	.187	0	1	.915	.917

5.1 Do Cancellations Respond to Weather?

In this subsection, I examine whether solar customers respond to any weather pattern by regressing the probability of contract cancellation on a set of weather variables. Among them, the primary variable of interest is solar insolation, or the amount of electromagnetic energy (solar radiation) incident on the surface of the earth. Solar insolation is the most important factor for solar power generation from an engineering perspective. Moreover, because it can be directly experienced by people, it is more likely that they will respond to it. The other weather variables in the analysis are: wind speed, number of days with maximum temperature below $40°$F and the same for above $100°$F. These other variables are included to control for confounding weather patterns (Auffhammer et al., 2013), as they are also theoretically relevant for solar power generation and electricity demand. Nevertheless, I also assess whether the results on solar insolation is robust with alternative sets of weather controls.

Table 2 shows the main results based on separate weather variables, six fixed effects specifications, and two level of clustered standard errors. The coefficients on solar insolation are negative and statistically significant across the board. This implies that a residential customer exposed to less sunshine in the post-contract period is more likely to cancel the contract. Wind speed is generally positively correlated with cancellation, but the effects are not stable across specifications. It might be because marginal effect of wind is different across seasons and other weather patterns. Extremely cold days also appear to be positively associated with cancellations, with large and statistically significant scales. However, this result should be interpreted with the caveat that it is identified from a small number of observations with nonzero values. Extremely hot days, on the other hand, does not appear to affect cancellation significantly, although the estimate is negative throughout.

Examining the scales more closely, column (1) has the smallest estimate on solar insolation, where none of the temporal fixed effects are controlled for and the effect is identified using deviations from the year-round mean solar insolation at each zip code. This is also the specification most prone to confounding factors and misspecification. In column (2), using within-year variations increases the scale slightly, which is more likely due to controlling for statewide confounding factors

Table 2: Cancellation and Solar Insolation within 113 Days

Cancel = 1	(1)	(2)	(3)	(4)	(5)	(6)
solar insolation	–.092	–.106	–.146	–.112	–.203	–.110
	(.027)***	(.029)***	(.043)***	(.044)**	(.045)***	(.044)**
	[.029]***	[.037]***	[.055]**	[.060]*	[.045]***	[.051]**
wind speed	.464	.411	.262	.225	.799	–.065
	(.153)***	(.155)***	(.167)	(.172)	(.206)***	(.172)
	[.242]*	[.247]	[.241]	[.254]	[.271]***	[.205]
#days(tmax $<$ 40)	1.034	.998	1.055	1.133	1.406	.658
	(.480)**	(.477)**	(.476)**	(.473)**	(.495)***	(.518)
	[.369]***	[.374]**	[.383]***	[.422]***	[.521]***	[.349]*
#days(tmax$\geq$100)	–.036	–.032	–.020	–.018	–.024	–.019
	(.014)***	(.014)**	(.014)	(.014)	(.016)	(.015)
	[.021]*	[.022]	[.023]	[.023]	[.021]	[.024]
cancellation rate	12.06	12.06	12.06	12.06	12.06	12.06
R^2 (within zip code)	.033	.034	.034	.035	.036	.043
N	154,655	154,655	154,655	154,655	154,655	154,655
Fixed Effects						
Quarter-of-year			Yes			Yes
Quarter-by-year				Yes		
County-by-quarter					Yes	
County-by-year						Yes
Year		Yes	Yes		Yes	
Zip Code	Yes	Yes	Yes	Yes	Yes	Yes

Notes: All coefficients are multiplied by 100 for legibility. Standard errors in parenthesis are clustered by zip code, and the ones in squared brackets by county. System characteristics and county-level economic conditions are always controlled for. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

that vary over the years. Column (3) further controls for quarterly fixed effects, which gets rid of seasonal economic fluctuations while alleviating misspecification. This results in a substantially larger point estimate. While column (1) and (2) provide useful baselines for comparison, column (3) presents a much more reliable estimate by using sufficiently fine variations. Columns (4)-(6) are further refinements of specification (3) in different directions. Column (4) allows the quarterly fixed effects to differ across years, (5) allows the quarterly fixed effects to differ across counties, and (6) allows for county-specific year fixed effects while controlling for common quarterly fixed effects. The results moves around the estimate in column (3), but are all negative and statistically significant. This suggests that confounders are unlikely and the estimated negative relationship represents real responses to weather. The estimates in column (3)-(6) suggest that one standard deviation decrease in solar insolation is associate with 5.2-9.3% increase in cancellations, while the

Table 3: Robustness Checks: Varying Controls

Cancel = 1	(1)	(2)	(3)	(4)	(5)	(6)
solar insolation	–.129	–.129	–.160	–.219	–.115	–.115
	[.045]***	[.041]***	[.051]***	[.044]***	[.050]**	[.045]**
wind speed			.262	.797		
			[.227]	[.259]***		
#days(tmax $<$ 40)					1.086	1.460
					[.392]***	[.551]**
#days(tmax $\geq$ 100)					–.019	–.021
					[.023]	[.020]
cancellation rate	12.06	12.06	12.06	12.06	12.6	12.06
R^2 (within zip code)	.034	.036	.034	.036	.034	.036
N	154,655	154,655	154,655	154,655	154,655	154,655
Fixed Effects						
Quarter-of-year	Yes		Yes		Yes	
County-by-quarter		Yes		Yes		Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Zip Code	Yes	Yes	Yes	Yes	Yes	Yes

Notes: All coefficients are multiplied by 100 for legibility. Standard errors are clustered by county. System characteristics and county-level economic conditions are always controlled for.
* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

baseline cancellation rate is 12.6% of all applications.

When multiple weather variables are included in a regression, a common concern is that the estimates are driven by specific correlations between these variables. To investigate this issue, I run a variety of robustness checks. First, Table 3 shows results excluding either wind speed or temperature bins from specifications (3) and (5) in Table 2. All the estimates are comparable to before. Second, I run a placebo test by assigning randomly drawn contract dates to the applications and replicate Table 2 (see Table A1 in the Appendix). The coefficients for all variables are much smaller and statistically insignificant, suggesting that correlations among the regressors are not a serious concern.

As a further support, I run a similar analysis based on a different set of weather variables, where solar insolation is dropped, while adding more maximum temperature bins and number of days with positive precipitation (see A2 in the Appendix). The estimates on all the overlapping variables are similar as before. While not included due to multicollinearity problem previously, precipitation is associated with a higher probability of cancellation in a statistically significant and robust pattern. This is consistent with the intuition that customers tend to have lower valuations of solar panels when they experience less favorable weather shocks.

I also check that the estimates on solar insolation and wind speed are robust when the period length is changed to 70 or 90 days (see Table A3 in Appendix). This is consistent with the discussions in Section 3.1, that the estimates are expected to be similar for a range of reasonable period lengths and attenuated for much shorter or longer periods due to increasing measurement errors. In Subsection 5.2 when weather is summarized by a single index, I further test whether the evolution of the estimate along period length indeed follows an inverse U-shape.

To briefly sum up, the results in this subsection show that customers respond to short-run weather fluctuations when deciding whether to cancel their solar contracts. Specifically, intensity of solar radiation in the short-run is significantly associated with the probability of cancellation: the weaker the solar insolation, the more likely the cancellation. This relationship is robust to different sets of fixed effects, controls, and post-period lengths. I also find suggestive evidence that increase in other weather variables, such as wind speed, precipitation, and extreme cold, is also associated with more cancellations.

5.2 Production or Demand Channel?

The above results show that households are responding to the intensity of solar radiation, which is consistent with households being concerned about solar panel profitability. However, it is not straightforward to interpret the estimated coefficients, due to three reasons. First, electricity generation from solar power is a nonlinear and complicated function of several weather variables, which is not likely to be correctly specified in a linear model as above. Second, the marginal effect of some variables, such as wind speed, might depend on other variables and varies a lot across seasons. Third and most importantly, the simple theory in Section 2 highlights two potential channels that projection bias might be at work. In reality, the households might have other concerns such as the possibility for strong wind to do damage to the solar panel and the roof. The estimates above capture the combined effects of the two potential channels and these other concerns.

This section is devoted to disentangle the two main channels outlined in the theory in Section

Table 4: Channel: Solar Production vs. Energy Demand

Cancel = 1	(1)	(2)	(3)	(4)	(5)	(6)
$ProductionIndex$	–.761	–.792	–.650	–.687	–.830	–.793
	[.264]***	[.259]***	[.323]**	[.362]*	[.247]***	[.252]***
CDD			–.260	–.223		
			[.297]	[.341]		
$TDD(= CDD + HDD)$					–.126	–.003
					[.279]	[.314]
cancellation rate	12.06	12.06	12.06	12.06	12.06	12.06
R^2 (within zip code)	.034	.036	.034	.036	.034	.036
N	154,655	154,655	154,655	154,655	154,655	154,655
Fixed Effects						
Quarter-of-year	Yes		Yes		Yes	
County-by-quarter		Yes		Yes		Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Zip Code	Yes	Yes	Yes	Yes	Yes	Yes

Notes: All indices are normalized. All coefficients are multiplied by 100 for legibility. Standard errors are clustered by county. System characteristics and county-level economic conditions are always controlled for. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

2.2, *assuming* that the other unobservable concerns are orthogonal to them. Specifically, I estimate the responses to summary indices for the two channels simultaneously, following the logic of an encompassing test Davidson et al. (1993). Construction of the two indices are detailed in Section 3.3. The index for solar production channel is generated from an engineering model, and the energy demand channel is represented by "degree days" measures: CDD or TDD (sum of HDD and CDD). Both indices are normalized in the results below so that the interpretation of the estimates become "percentage point change in cancellation associated with one standard deviation increase in the index".

Table 4 reports the results on the two channels, using FE specifications (3) and (5) in Table 2. In the first two columns, I regress the cancellation dummy on the production index only, and the estimates are again negative and statistically significant. The scale translates into around 6.4% increase in cancellation probability for one standard deviation decrease in the production index. CDD was added into the regressions in columns (3)-(4). The results are slightly smaller, but as the standard errors increase, the significance level decreases. The increase in standard error is likely due to multicollinearity problem, because the variations in production index and CDD are strongly driven by solar insolation and temperature, respectively, and the latter two are highly correlated. The coefficient on CDD is negative as expected, but highly insignificant and economically small. The last two columns switch to TDD to represent energy demand. The coefficient on TDD is much smaller, while the scale and statistical significance of the production index is at similar levels as before. Therefore, these estimates suggest that the solar production aspect plays a greater role than the energy demand channel in the households' decisions.

There are several potential explanations of why people are not responding to energy demand as strongly as to solar production. The theory predicts that the energy demand channel would be at work only if the marginal electricity price is quite different with and without the solar generation. For some households, it might be the case that their solar system is not large enough to move the marginal price to a lower tier. For some others, it might be because they were not at the highest tier to begin with, and hence the difference between two lower tiers is small. Moreover, since households appear to respond to average rather than marginal price, this also tends to attenuate the value of electricity displaced (Ito, 2014). Lastly, it could be that households derive large non-pecuniary utility from solar panels on sunny days, which amplifies the solar production channel. For example, some people might feel more proud of going solar when the weather is more favorable to production. Note that although this in itself can be perfectly rational preference, it is also a form of projection bias if the household uses the current weather to estimate the total non-pecuniary utility they would derive from the solar system.

One important takeaway from the above results is that the solar production index alone is a useful summary of relevant characteristics of the current period weather. In Figure A3 in the Appendix, I show the non-parametric relationship between the probability of cancellation and this index using two methods, and both strongly support a linear relationship. Therefore, I can use this index to perform further analysis without getting into complications of interpreting the reduced-form estimates on multiple weather variables.

For example, I examine whether current period weather can predict average weather in the near future (See Table A4). For each application, I calculate the average production index in an one-year period starting right after the end of the 113-day post-contract period, as a measure of production

Table 5: Pre-period Weather and Purchase

(Zip Code-by-Month)	Number of Applications			log(Number of Applications)		
	(1)	(2)	(3)	(4)	(5)	(6)
Pre-period Length	1Mo	2Mo	3Mo	1Mo	2Mo	3Mo
$Index_{pre}$	.569	.714	.570	.102	.132	.107
	[.085]***	[.106]***	[.096]***	[.008]***	[.009]***	[.009]***
R^2 (within zip code)	.407	.409	.408	.480	.481	.480
N	46,018	46,001	45,911	46,018	46,001	45,911
Fixed Effects						
Quarter-of-year	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Zip Code	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The dependent variable for columns (1)-(3) is the number of applications by zip code by month, and for (4)-(6) is its log level. The third row indicates the length of the pre-period used to calculate $Index_{pre}$. Standard errors are clustered by county. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

potential. Regressing this measure on the post-period index using the preferred specifications show small and insignificant coefficients.

For another example, I examine the first half of the conjectured inverted U-shape based on the index (See Table A5 in the Appendix). Six estimates are presented, which are based on 10, 30, 50, 70, 90, and 113-day post-contract periods respectively. The scale of the estimates are increasing in the length of period, and all other estimates are statistically different from the main 113-day estimate except for the 90-day one. These are all clearly consistent with the previous conjecture.

5.3 Selection into Contract

In this subsection, I test another implication of projection bias, that purchase itself might also be affected by short-run weather conditions prior to it. Conlin et al. (2007) find that cold weather (favorable in their case) on the purchase day is associated with higher probability of cancellation later. Their interpretation is that some unnecessary purchases are induced by weather. Busse et al. (2014) and Simonsohn (2010) find direct evidence on purchase, but are less able to infer change of valuation because returns in their contexts are rare and unobservable. The first goal of this subsection is to provide evidence on purchase in a similar spirit to these two papers.

Most previous tests of projection bias, including my analysis above, are based on weather in just one period. However, this "single-period" approach might have only painted a partial picture, because the customers might not be responding to just weather conditions in the pre-contract, or the post-period, but the comparison of the two. Suppose one customer experiences above-average weather in the pre-period, and even better weather afterwards, theory of projection bias predicts that she should have lower probability of cancellation than those who experience the average weather throughout. Suppose another customer experiences the same weather in the pre-period, and worse weather afterwards, she should have higher probability to cancel instead. The single-period approach essentially pulls these two types together. However, distinguishing between them can better shed light on how customers are processing information about weather at different points in time. This is the second goal of this subsection.

For solar panels, the relevant pre-periods for considering purchase are more heterogeneous and harder to infer than the relevant post-period. Therefore, I show results assuming different pre-period lengths. Still, it is imperfect, which is a major caveat for interpreting the results in this subsection.

Table 5 presents results of whether solar production index in the pre-period is correlated with the number of applications, after controlling for quarterly, year and zip code FEs.[10] Columns (1)-(3) use number of applications at a given zipcode in a given month as the dependent variable, while (4)-(6) use the log of number of applications. Three estimates are shown for each dependent variable, assuming the relevant pre-period as 1, 2, and 3 months, respectively. The estimates are all positive and statistically significant for both level and log. The level estimates range from 0.57 to 0.71 additional application for one standard deviation increase in the index, and the log estimate range from 10.2% to 13.2% increase. Therefore, the evidence generally suggest that the number of applications are responding to the solar radiation in the prior period. Note, however, that part of this might be attributable to short-run shifting of purchase timing by one or two months. In contrast to the extensive margin, there does not appear to be systematic selection on the intensive margin, namely the system size (see A7 in the Appendix).

In Table 6, I examine whether the customers are responding to the change in post-period weather relative to the pre-period ("weather update" henceforth). I assume a 30-day pre-period, and use

[10]Results under alternative FE specification are very similar and available upon request.

Table 6: Responses to Weather Updates

	Updates		Interactions		
Cancel = 1	(1)	(2)	(3)	(4)	(5)
$Index_{pre}$	.253	–.472	–.055		
	[.235]	[.313]	[.288]		
$Update$		–.802	–.124	–.257	.368
		[.269]***	[.601]	[.203]	[.444]
$Dummy =$					
$1(Update < 0)$			–.020		
			[.004]***		
$1(Index_{pre} > \overline{Index}_z)$				–.004	
				[.007]	
$1(Index_{post} < \overline{Index}_z)$					.004
					[.004]
$Update \times Dummy$			–1.746	–.556	–.976
			[.556]***	[.314]*	[.335]***
cancellation rate	12.06	12.06	12.06	12.06	12.06
N	154,655	154,655	154,655	154,655	154,655
Fixed Effects					
Quarter-of-year	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes
Zip Code	Yes	Yes	Yes	Yes	Yes

Notes: $Index_{pre}$ is the normalized average production index for a 30-day pre-contract period. $Update$ is the average post-period index minus the pre-period one. $1(Update < 0)$ indicates whether $Update$ is negative. $1(Index_{pre} > \overline{Index}_z)$ is an indicator of whether the pre-period index is higher than zipcode average, and $1(Index_{post} < \overline{Index}_z)$ is whether the post-period one is lower. All coefficients are multiplied by 100 for legibility. $Dummy$ in the interaction term in the last rows refers to the dummy variable that is included in the same regression. Standard errors are clustered by county. System characteristics and county-level economic conditions are always controlled for.
* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

the same FE specifications as in Table 5.[11] Column (1) estimates the relationship between cancellation and the pre-period production index. The coefficient is positive but insignificant. When weather update is added to the regression (Col (2)), the coefficient on it is negative and statistically significant, while the coefficient on pre-period index switches signs but is still insignificant.

Next, I test for differential effects of the update by including an interaction term of Update and a category indicator. In Column (3), I test whether negative updates have a larger effect on cancellations than positive one. The idea is, a positive update might not be informative given that these customers have already self-selected into a contract. Indeed, the interaction term, which sep-

[11] Using a 60-day pre-period yields qualitatively similar results (see A6 in the Appendix). Results under alternative FE specifications are also very similar and available upon request.

arates out the negative updates, completely absorbs the effect of the $Update$ variable and becomes twice as large. This provides strong support for different information contents between positive and negative updates. Column (4) shows that the update has a larger impact on customers who were exposed to above-average weather before signing up, and (5) shows even greater impacts on those exposed to below-average weather in the post-period. [12] These effects suggest that the customers fail to recognize that deviations from the mean weather are transient. Instead, they make a long-lasting decision based on these deviations.

5.4 Heterogeneous Effects

This subsection explores whether responses are heterogeneous across various dimensions. For ease of interpretation, I again focus on the production index and use Col (1)-(2) in Table 4 as the baseline estimates for comparison. I interact the index with different characteristics that might affect the cancellation decision as well as the extent to which people rely on current weather in their decisions. The interaction term is the variables of interest throughout because it captures the differential responses by customers with the specific characteristic.

For testing behavioral bias, one informative dimension of heterogeneity is residential vs. non-residential customers. The latter group, including small businesses, government agencies, and NGOs, are also eligible for rebate under CSI. There are much fewer of them – the clean data contains only 8,730 non-residential applications. The decision process of these customers is fundamentally different from residential ones, as they are more likely to employ formal calculations of discounted investment returns to guide the decision. The program rules are also different. They follow a three-step procedure and are not required to have a signed contract in order to reserve the rebate. Hence, it is unclear to what extent the cancellations of non-residential systems are costly. Nevertheless, I view the analysis below as useful evidence that sheds light on the behavioral mechanism.

Table 7 reports results on the cancellations of these non-residential applications. The first two columns are exact counterparts of the columns (1)-(2) in Table 4. The estimates are still negative, but the scales are substantially smaller and insignificant. However, the lack of statistical power

[12] "Above-average" here means having an index level greater than the year-round average at the given zip code.

Table 7: Cancellations of Non-residential Applications

Cancel = 1	(1)	(2)	(3)	(4)
	Non-residential Only		Pooled	
$Index_{post}$	−.427	−.427	−.802	−.843
	[1.004]	[.992]	[.260]***	[.257]***
$1(non\text{-}residential)$			−.114	−.120
			[.255]	[.253]
$Index_{post} \times 1(non\text{-}residential)$			.893	1.050
			[.138]***	[.109]***
F-stat: $\beta_1 + \beta_3 = 0$			0.08	0.94
N	8,730	8,730	163,385	163,385
Fixed Effects				
Quarter-of-year	Yes		Yes	
County-by-quarter		Yes		Yes
Year	Yes	Yes	Yes	Yes
Zip Code	Yes	Yes	Yes	Yes

Notes: The production index is normalized, and all coefficients are multiplied by 100 for legibility. Standard errors are clustered by county. System characteristics and county-level economic conditions are always controlled for. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

constrains our ability to conclude that this represents a null effect.

To enhance power, I pool the non-residential applications with the large residential sample used above, and add an indicator for non-residential system as well as an interaction term of this indicator with the production index. The results are shown in Columns (3)-(4) in Table 7, and display a striking reversal pattern. The coefficient on $Index$ captures the effect of weather on residential customers, and is similar to the baseline (Table 4, Col(1)-(2)). On the other hand, the one on the interaction term goes the opposite direction and more than compensates the negative scale by a small margin. Captured by the sum of the two coefficients, the effect of weather on non-residential customers is very small in scale. The F-statistic of a formal test is insignificant.

However, there is no guarantee that the fixed effects are similar for the two types of customers. It is unclear whether the seasonal and location specific economic confounders are similar for households and these non-residential entities. The coefficient on the interaction term will be biased when the omitted differences are correlated with weather. For example, a problem would arise if 2009 has particularly unfavorable weather for solar production, while the financial crisis in that year makes businesses more time-inconsistent than households in terms of their solar purchase decisions. If

Table 8: Peer Effects in Cancellations and Response to Weather

Cancel = 1	(1)	(2)	(3)	(4)
	Current Installed Base		Installed Base $>$ 1 Year	
$Index_{post}$	–.702	–.716	–.758	–.751
	[.374]*	[.368]*	[.351]**	[.352]**
$Penetration$	–.859	–.863	–1.110	–1.093
	[.285]***	[.273]***	[.471]**	[.460]**
$Index_{post} \times Penetration$	1.829	.702	6.839	3.978
	[12.643]	[13.155]	[15.437]	[17.418]
N	151,043	151,043	151,043	151,043
Fixed Effects				
Quarter-of-year	Yes		Yes	
County-by-quarter		Yes		Yes
Year	Yes	Yes	Yes	Yes
Zip Code	Yes	Yes	Yes	Yes

Notes: The production index is normalized. The first and third coefficients are multiplied by 100 for better display. Standard errors clustered by county. System characteristics and county-level economic conditions are always controlled for. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

these are true, the latter effect would be incorrectly attributed to the former. Such confounders are not likely to be a main driver of the reversal result we are seeing, though, especially when the outcome is cancellation conditional on applying. Therefore, I view these results as suggestive that non-residential customers do not respond to weather, potentially because they have formal decision-making process and thus are less vulnerable to behavioral bias in their decisions.

Next, I explore heterogeneous responses among residential applications in locations with different solar market penetration. Peer effects are found to play a role in the adoption of solar panels (Bollinger and Gillingham, 2012), and it might also affect cancellations and responses to weather. For instance, households with more neighbors having solar PV systems might be more determined and less likely to cancel. If the neighbor has the system for over a year, he might be able to share his experience with the year-round fluctuation of solar power production. Therefore, households living near neighbors with older systems might be less vulnerable to projection bias. These conjectures are tested below.

Identification of peer effects in (Bollinger and Gillingham, 2012) uses the fact that installation typically occur with some delay after the purchase decision. This means the increase in current installed base is predetermined and uncorrelated with other contemporaneous shocks to the peer group. The logic of the following analysis is similar, that the increase in interconnected systems is predetermined and would only affect current decisions through peer effects. I construct two measures of solar market penetration in a zip code using the net energy metering (NEM) interconnection dataset. This dataset contains all interconnected project with information on zip code and relevant dates, which allows me to calculate the current installed base in a zip code for a given month. I also calculate the number of systems that are at least one year old, because these neighbors might be more helpful in alleviating projection using current weather. Each variable is divided by the total number of households in the zip code to obtain a measure of market penetration.

Table 8 presents results of regressing cancellation on the production index, penetration, and their interaction. The penetration measure in Columns (1)-(2) is based on the current installed base, while (3)-(4) are for those older than one year. The coefficient on the production index is similar as before. Higher market penetration is associated with less cancellation, and the relationship is significant. This is consistent with the previous studies – peer effects not only make the households more likely to sign up, but also prevent them from canceling the contract. However, there is no support for the conjecture that neighbors with older systems might provide useful information against the over-reliance on current weather. The estimates are indeed of opposite sign of the main effect, larger for the older base, but statistically insignificant throughout.

I also explore whether responses differ with individual system and application characteristics. Table 9 reports these results. Columns (1) and (2) shows that systems with larger sizes or lower average cost are more likely to be canceled, but their response to weather does not differ systematically.

Column (3) examines how third-party ownership (TPO) affects reponses to weather. TPO is a common option to finance solar system, which allows the household to acquire the system without paying a large up-front cost. Under this scheme, the solar company owns the system, and either signs a power-purchase agreement (PPA) or a lease with the household.[13] In the earlier years, most systems are self-owned. Later, TPO was introduced to tap into a broader customer base and has since become the most popular option. In the dataset, about half of all applications are under TPO.

[13] PPA allows the household to pay a low fixed rate for the electricity generated from the system, while the lease is similar to a loan which the household need to pay back.

Table 9: Heterogeneous Effects by System Characteristics

Cancel = 1	(1) System Size Above Median	(2) Average Cost Above Median	(3) Third-Party Ownership	(4) Near Incentive Step Change
$Index_{post}$	–.753	–.787	–.449	–.595
	[.309]**	[.275]***	[.251]*	[.280]**
$1(Characteristic)$	1.723	–1.221	.837	1.863
	[.383]***	[.615]*	[.522]	[.580]***
$Index_{post} \times 1(Characteristic)$	.007	.081	–.638	–.251
	[.270]	[.210]	[.292]**	[.334]
N	154,655	154,657	154,655	154,655
Fixed Effects				
Quarter-of-year	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
Zip Code	Yes	Yes	Yes	Yes

Notes: The production index is normalized. All coefficients are multiplied by 100 for better display. Standard errors are clustered by county. Other system characteristics and county-level economic conditions are always controlled for. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

It is unclear whether customers with TPO systems should be more or less sensitive to weather. On one hand, TPO typically reduces the risk from weather by compensating the customer when production is particularly low. On the other hand, it is well known that TPO create restrictions for selling the house within the 20-year term. The customer might therefore place more value on whether the system "pays off" sooner and hence make their decision more reliant on projected performance. Furthermore, selection into TPO is not random – the option is created to appeal to customers who are more liquidity-constrained or risk averse. The results suggest that they are much more sensitive to weather. The effect of weather on their cancellations is more than two times larger than that for self-owned systems. A likely reason is that many of them are marginal customers who can easily change their minds.

The last application characteristic I investigate is whether it is submitted near the next incentive step change. Step changes occur at times when the total solar capacity reserved in the program exceeds certain thresholds. Although the exact timing of the change is unknown, applicants still respond to imminent step change.[14] One possibility is that more marginal customers might be rushed into a contract for the option value of getting the higher rebates. Therefore, they not only

[14] According to a CSI staff, "... there was always a flurry of reservations and overall activity around step changes".

have a higher propensity to cancel, but also are more likely to cancel due to weather fluctuations.

Based on the realized schedule for incentive step change,[15] I calculate the duration to the next step change for each application. The sample is split by whether this duration is less than 47 days, the 25th percentile of this variable (see Figure A4 in the Appendix for the distribution of days to next step change).[16] Column (4) suggests that these customers indeed have a substantially higher probability of cancellation, and they are slightly more sensitive to weather.

Lastly, I examine whether there are heterogeneous responses based on zip code characteristics (see Table A8 in the Appendix). These characteristics include median income, fraction of white people, fraction with a bachelor degree, average household size, fraction of urban households, housing cost, whether the zip code is in Norther California, and whether it is in non-coastal counties. The only significant finding is that the responses to weather are entirely driven by zip codes that have more urban households than rural ones. The rest of the zip code characteristics do not seem to split the sample in meaningful ways. It might be because solar customers in any zip code are a highly selected group (Harding and Rapson, 2013; Borenstein and Davis, 2015), and hence the aggregate characteristics are not very informative.

5.5 Interpretation and Alternative Mechanisms

The empirical analysis shows that better weather is associated with more solar panel contracts, and worse weather in the post-contract period is associated with higher probability of cancellation. The evidence is also consistent with customers changing their valuation of solar panels when they experience a change in weather patterns after signing up relative to before. In contrast, there is suggestive evidence that non-residential customers are not affected.

I interpret these findings as consequences of consumers suffering from projection bias. That is, they rely too much on current conditions to predict future utility, and are making long-lasting decisions based on transient conditions. Here I use the term "projection bias" in a broad sense, to

[15] Source: https://www.californiasolarstatistics.ca.gov/reports/budget_forecast/#Step.

[16] There is no special reason to use the 25th percentile. The results are robust to other thresholds in a similar range. However, if the threshold is too low, the result will be driven by a small number of observations. If the threshold is too high, it fails to capture the notion of being close to the imminent change.

include both *mistaken utility* and *mistaken beliefs* about weather.[17] Mistaken utility means that the consumer incorrectly predicts that her future utility will be more like today, and mistaken belief means that she instead thinks that today's weather condition is more likely to arise in the future. Both are psychological biases, predict similar behaviors, and are usually hard to tell apart. However, because solar panel are highly instrumental and not something the customers can directly enjoy, I interpret the effects as mainly driven by mistaken beliefs.

There are several competing mechanisms that can explain some of the findings. The most plausible one is rational learning. That is, the customers are in fact extracting useful information from weather fluctuations to guide their choice. There are a few things that the customers can be learning about.

One possibility is that the current weather can help predict future weather. It is certainly true that short-run weather has predictive power cross-sectionally: while Los Angeles is hotter than San Francisco in December, it is also hotter around the year. However, zip code fixed effects are included in all specifications, absorbing most of the cross-sectional correlations. Within the same location, weather conditions are not persistent in the relatively short run. As has been discussed before, Table A4 shows that current weather is not predictive of that in the near future (1 year). It is more unlikely that current weather can predict weather into the distant future.

Does current weather help the consumer infer something about climate change? Researchers find that recent outdoor temperatures might indeed affect belief in climate change (Li et al., 2011). However, this kind of learning, if present, is also not likely to be rational here given the nature of weather variations used for identification. The preferred specifications include year fixed effects and another specification has county-specific year fixed effects. Under these specifications, gradual changes in climate are largely absorbed, but I still find significant effects of solar insolation on cancellation.

Yet another possibility is that the consumer is learning something about the profitability to go solar. Learning can be rational if the consumer is previously inattentive. For example, hot and sunny weather might lead an inattentive consumer to be correctly aware about (i) her electricity

[17]Some might instead prefer to use the term narrowly for mistaken utility only.

needs, (ii) solar panel productivity. This explains why more contracts are signed following good weather. However, (i) appears not to be important according to Table 4. For (ii), it is unclear what can be meaningfully learned before the solar panels are installed. Furthermore, both stories are hard to reconcile with the results on cancellation: if the learning is complete, why would a customer be more likely to cancel a contract when the weather turns bad later?

There might be other types of learning that can explain some of the findings. However, it is hard to think of anything useful that the household can learn from short-run weather, that has additional value over the estimates provided by the solar company. Furthermore, the results on cancellations also pose a general challenge to stories of rational learning.

Another alternative explanation is that current weather affects productivity of the solar companies, which then affect purchase and cancellation. This can arise because: first, solar salesmen and installers primarily work outdoor, and therefore are more productive in good weather; second, some solar companies are aware that it is easier to make a successful sales pitch in the summer than winter, as suggested by anecdotal evidence. However, the second is evidence of solar companies taking advantage of customers' projection bias, and hence need not be separated out. As for the first, we would expect the effect to be mainly driven by extreme weather and is not consistent with the linear relationship in Figure A3. Additionally, heat is found to lower human productivity, but it is in fact associated with lower probability of cancellation.

To sum up, none of the proposed alternative mechanisms can satisfactorily explain all the main findings. The most plausible mechanism is a psychological bias that leads the consumers to rely too much on transient conditions when making a long-lasting decision.

6 Conclusion

This paper studies how weather fluctuations affect the decision to go solar and also the decision to cancel a signed solar contract. The main result shows that customers whose sign-up for solar panels is followed by bad weather are more likely to cancel their contract. Furthermore, consumers are more likely to cancel when they experience worse weather after signing up relative to before, suggesting that the negative update changes their valuation of solar panels. This effect is mainly driven by their concerns over future solar panel productivity. I also find suggestive evidence that non-residential customers do not respond to weather, and marginal residential customers respond more. The most plausible mechanism consistent with all these results is projection bias driven by mistaken beliefs. That is, the customers rely too heavily on current weather to predict future weather.

These results suggest that projection bias can substantially affect demand for a big-ticket item like solar panel, and might also be at work for other energy-related investments. Such investments are different from other goods studied before, in that they are under-adopted socially and often heavily subsidized. Therefore, policymakers can benefit from learning and using these behavioral patterns to better target subsidies to marginal customers and reduce wasteful cancellations.

Finally, there are several limitations in this paper, each representing a potential extension for future work. The first limitation is not knowing the exact pre- and post-period, leading to under-estimation due to attenuation bias. It would be helpful to study such intertemporal biases in other contexts where projections are also continually updated, but with more clearly defined pre- and post-periods. Second, without individual characteristics of the solar customers, this paper cannot reach a solid conclusion on what type of consumers are more susceptible to such bias. Third, this paper finds that information provided by the solar companies or peer effects by neighbors do not eliminate such bias. Future studies can explore better strategies to "de-bias" the customers. Answering these questions are important for both economic modeling and policy design.

References

Allcott, H. and Wozny, N. (2014). Gasoline prices, fuel economy, and the energy paradox. *Review of Economics and Statistics*, 96(5):779–795.

Auffhammer, M., Hsiang, S. M., Schlenker, W., and Sobel, A. (2013). Using weather data and climate model output in economic analyses of climate change. *Review of Environmental Economics and Policy*, page ret016.

Baylis, P. (2015). Temperature and temperament: Evidence from a billion tweets. *Working Paper*.

Bollinger, B. and Gillingham, K. (2012). Peer effects in the diffusion of solar photovoltaic panels. *Marketing Science*, 31(6):900–912.

Bollinger, B. K. and Gillingham, K. (2014). Learning-by-doing in solar photovoltaic installations. *Available at SSRN 2342406*.

Borenstein, S. and Davis, L. W. (2015). The distributional effects of US clean energy tax credits. In *Tax Policy and the Economy, Volume 30*. University of Chicago Press.

Busse, M. R., Pope, D. G., Pope, J. C., and Silva-Risso, J. (2014). The psychological effect of weather on car purchases. *The Quarterly Journal of Economics*, page qju033.

California ISO (2012). 2012 summer loads and resource assessment. Technical report.

Chang, T. Y., Huang, W., and Wang, Y. (2016). Something in the air: Projection bias and the demand for health insurance. *Working Paper*.

Conlin, M., O'Donoghue, T., and Vogelsang, T. J. (2007). Projection bias in catalog orders. *The American Economic Review*, pages 1217–1249.

Davidson, R., MacKinnon, J. G., et al. (1993). Estimation and inference in econometrics.

de Magistris, T. and Gracia, A. (2016). Assessing projection bias in consumers' food preferences. *PloS one*, 11(2):e0146308.

DellaVigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic literature*, 47(2):315–372.

EIA (2016). Sources of heating for california homes.

Graff Zivin, J. and Neidell, M. (2014). Temperature and the allocation of time: Implications for climate change. *Journal of Labor Economics*, 32(1):1–26.

Harding, M. and Rapson, D. (2013). Do voluntary carbon offsets induce energy rebound? a conservationist's dilemma. *Working Paper*.

Hughes, J. E. and Podolefsky, M. (2015). Getting green with solar subsidies: Evidence from the california solar initiative. *Journal of the Association of Environmental and Resource Economists*, 2(2):235–275.

Ito, K. (2014). Do consumers respond to marginal or average price? evidence from nonlinear electricity pricing. *The American Economic Review*, 104(2):537–563.

Kleissl, J. (2013). Improving economics of solar power through resource analysis, forecasting, and dynamic system modeling. Technical report.

Li, Y., Johnson, E. J., and Zaval, L. (2011). Local warming: daily variation in temperature affects beliefs and concern about climate change. *Psychol Sci*, 22:454–459.

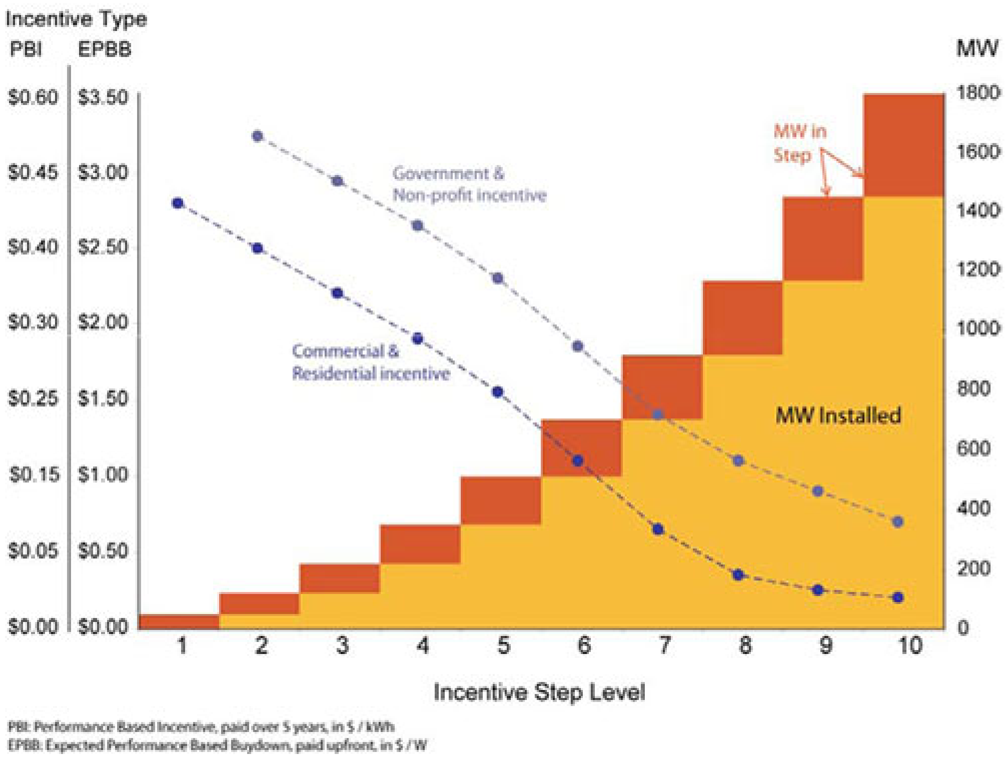
Loewenstein, G., O'Donoghue, T., and Rabin, M. (2003). Projection bias in predicting future utility. *The Quarterly Journal of Economics*, pages 1209–1248.

Simonsohn, U. (2010). Weather to go to college. *The Economic Journal*, 120(543):270–280.

Stackhouse Jr, P. W., Westberg, D., Hoell, J. M., Chandler, W. S., and Zhang, T. (2015). Prediction of worldwide energy resource (power) agroclimatology methodology. Technical report, NASA Langley Research Center and SSAI/NASA Langley Research Center.

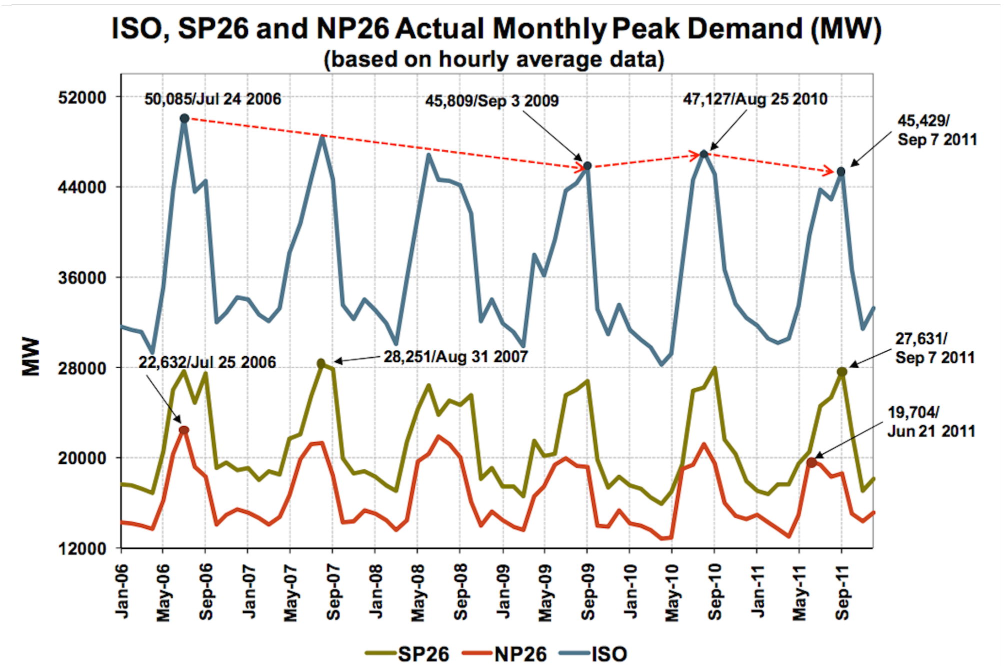
Appendix

Figure A1: CSI Incentive Step Design



Source: http://www.gosolarcalifornia.ca.gov/csi/rebates.php

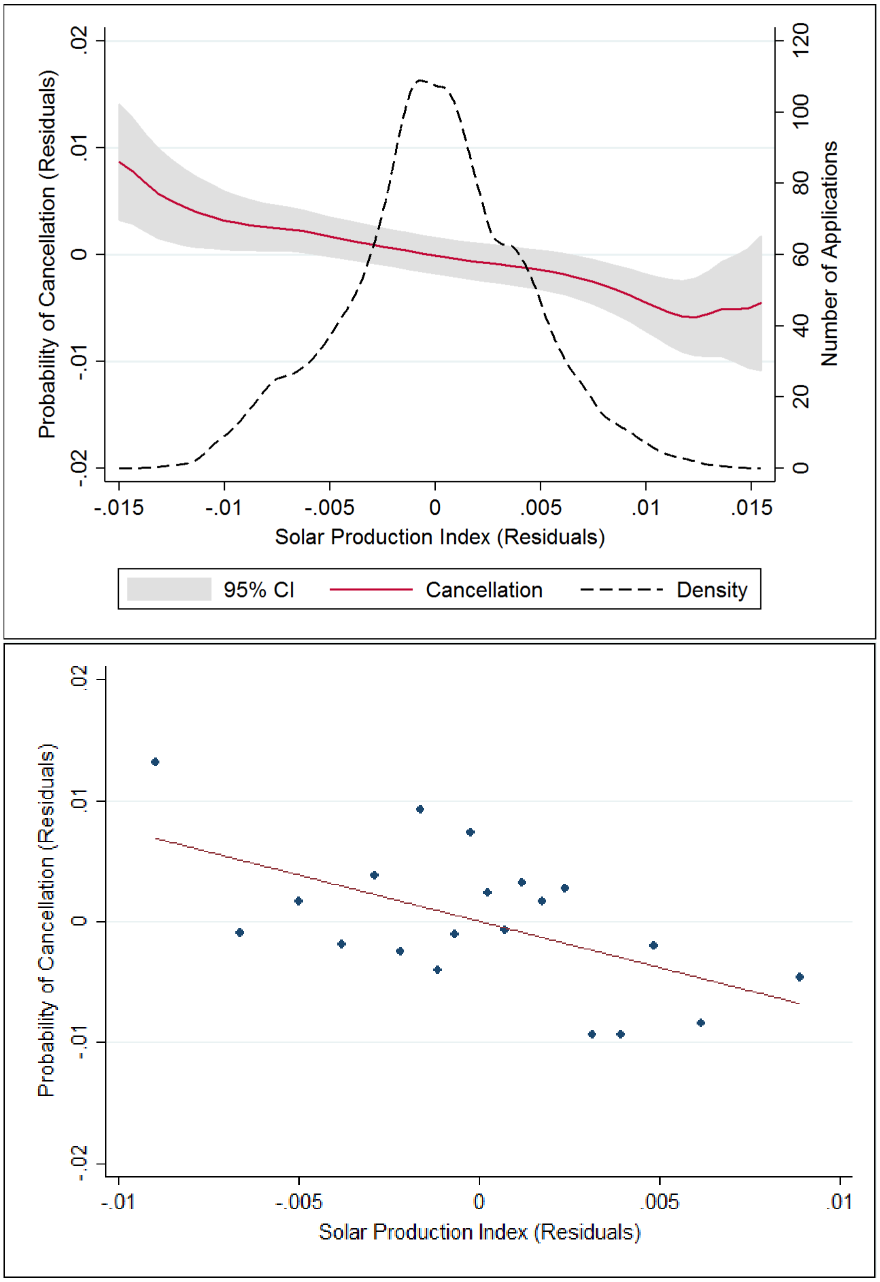
Figure A2: Monthly Peak Demand in California, 2006-2011



Source: California ISO (2012)

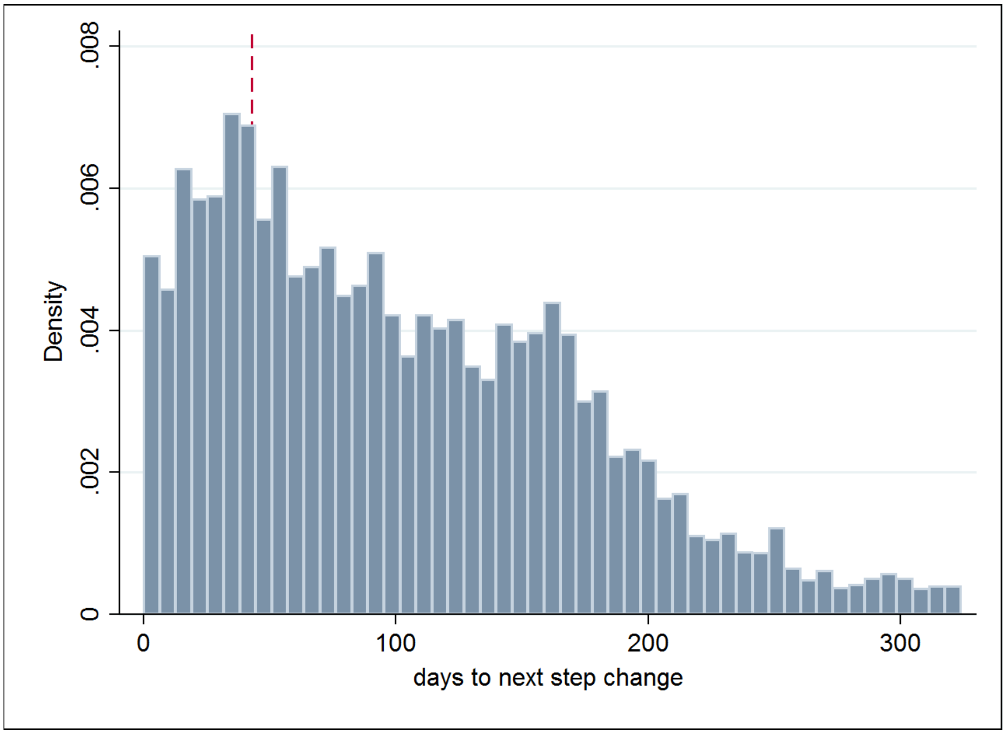
Notes: This graph shows the actual monthly peak demand for southern California (SP26), northern California (NP26), and the entire California ISO system over the years 2006-2011.

Figure A3: Support for Linear Specification



Notes: These graphs verify the linear specification between the probability of cancellation and solar production index using two approaches. The upper panel fits the relationship with a flexible local polynomial, and the lower one uses a binned scatter plot by dividing the observations into twenty bins.

Figure A4: Density of Days to Next Incentive Step Change



Notes: This graph plots the distribution of days to next incentive step change. The vertical dashed red line represents 43 days, the cutoff used in the heterogeneity analysis.

Table A1: Placebo Replication of Main Results on Solar Insolation

Cancel = 1	(1)	(2)	(3)	(4)	(5)	(6)
solar insolation	–.030	–.028	–.020	–.027	–.004	–.026
	(.024)	(.025)	(.037)	(.037)	(.039)	(.037)
	[.017]	[.019]	[.032]	[.029]	[.034]	[.032]
wind speed	.206	.196	.174	.231	.240	.136
	(.135)	(.136)	(.142)	(.143)	(.169)	(.153)
	[.126]	[.124]	[.126]	[.127]*	[.165]	[.129]
#days(tmax $<$ 40)	.065	.079	.070	.078	.100	.368
	(.312)	(.315)	(.314)	(.318)	(.395)	(.504)
	[.214]	[.219]	[.216]	[.221]	[.227]	[.381]
#days(tmax $\geq$ 100)	–.002	–.002	–.002	–.003	–.008	.000
	(.012)	(.012)	(.013)	(.012)	(.014)	(.013)
	[.011]	[.011]	[.011]	[.011]	[.012]	[.011]
cancellation rate	12.06	12.06	12.06	12.06	12.06	12.06
R^2	.032	.032	.032	.032	.033	.034
N	154,684	154,684	154,684	154,684	154,684	154,684
Fixed Effects						
Quarter-of-year			Yes			Yes
Quarter-by-year				Yes		
County-by-quarter					Yes	
County-by-year						Yes
Year		Yes	Yes		Yes	
Zip Code	Yes	Yes	Yes	Yes	Yes	Yes

Notes: All coefficients are multiplied by 100 for legibility. Standard errors in parenthesis are clustered by zip code, and the ones in squared brackets by county. System characteristics and county-level economic conditions are always controlled for. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table A2: Cancellation and Other Weather Variables

Cancel = 1	(1)	(2)	(3)	(4)	(5)	(6)
#days(tmax $\leq$ 40)	.971	.999	1.023	1.158	1.361	.490
	(.480)**	(.478)**	(.482)**	(.481)**	(.503)***	(.526)
	[.378]**	[.395]**	[.407]**	[.442]**	[.508]***	[.333]
#days($40 <$ tmax ≤ 60)	.006	–.001	.005	.004	.018	.018
	(.009)	(.010)	(.010)	(.011)	(.013)	(.010)*
	[.011]	[.013]	[.013]	[.016]	[.021]	[.012]
#days($80 <$ tmax ≤ 100)	–.010	–.014	–.015	–.006	–.016	–.013
	(.006)*	(.006)**	(.007)**	(.007)	(.007)**	(.007)**
	[.006]	[.006]**	[.008]*	[.010]	[.009]*	[.007]*
#days(tmax $>$ 100)	–.030	–.028	–.026	–.023	–.035	–.022
	(.014)**	(.014)*	(.015)*	(.015)	(.016)**	(.015)
	[.026]	[.027]	[.026]	[.024]	[.022]	[.027]
wind speed	.393	.239	.195	.188	.580	–.089
	(.115)***	(.122)**	(.149)	(.158)	(.176)***	(.150)
	[.177]**	[.187]	[.228]	[.252]	[.261]**	[.186]
#days(precipitation$>$0)	.030	.024	.024	.032	.022	–.000
	(.010)***	(.010)**	(.010)**	(.011)***	(.011)**	(.010)
	[.011]***	[.011]**	[.013]*	[.012]**	[.016]	[.014]
cancellation rate	12.06	12.06	12.06	12.06	12.06	12.06
R^2 (within zip code)	.033	.034	.034	.035	.036	.043
N	154,655	154,655	154,655	154,655	154,655	154,655
Fixed Effects						
Quarter-of-year FE			Yes			Yes
Quarter-by-year FE				Yes		
County-by-quarter FE					Yes	
County-by-year FE						Yes
Year FE		Yes	Yes	Yes	Yes	
Zip Code FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: All coefficients are multiplied by 100 for legibility. Standard errors in parenthesis are clustered by zip code, and the ones in squared brackets by county. System characteristics and county-level economic conditions are always controlled for. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table A3: Robustness Checks: Varying Period Length

Cancel = 1 Period length	(1) 70 days	(2) 70 days	(3) 90 days	(4) 90 days	(5) 113 days	(6) 113 days
solar insolation	–.147	–.185	–.157	–.206	–.146	–.203
	[.045]***	[.043]***	[.048]***	[.043]***	[.055]**	[.045]***
wind speed	.319	.652	.317	.751	.262	.799
	[.210]	[.248]**	[.220]	[.249]***	[.241]	[.271]***
#days(tmax $<$ 40)	.358	.511	.379	.569	1.055	1.406
	[.351]	[.400]	[.352]	[.421]	[.383]***	[.521]***
#days(tmax $\geq$ 100)	–.042	–.052	–.030	–.036	–.020	–.024
	[.024]*	[.022]**	[.025]	[.023]	[.023]	[.021]
cancellation rate	12.06	12.06	12.06	12.06	12.06	12.06
R^2 (within zip code)	.032	.033	.032	.033	.032	.033
N	154,655	154,655	154,655	154,655	154,655	154,655
Fixed Effects						
Quarter-of-year	Yes		Yes		Yes	
County-by-quarter		Yes		Yes		Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Zip Code	Yes	Yes	Yes	Yes	Yes	Yes

Notes: All coefficients are multiplied by 100 for better display. Standard errors are clustered by county. System characteristics and county-level economic conditions are always controlled for.
* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table A4: Does Current Weather Predict the Near Future?

Future Index	(1)	(2)	(3)
$ProductionIndex$	–.001	–.008	.033
	[.026]	[.025]	[.027]
N	154,242	154,242	154,242
Fixed Effects			
Quarter-of-year		Yes	
County-by-quarter			Yes
Year		Yes	Yes
Zip Code	Yes	Yes	Yes

Notes: The dependent variable is mean daily production index in a one-year period starting right after the 113-day post-period.
* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table A5: Attenuation Bias with Different Period Lengths

System Size	(1) 10 days	(2) 30 days	(3) 50 days	(4) 70 days	(5) 90 days	(6) 113 days
$ProductionIndex$	.057	–.249	–.443	–.542	–.692	–.761
	[.141]	[.141]*	[.144]***	[.154]***	[.172]***	[.191]***
Test: Equality of Coefficients						
$\chi^2 : \beta_{(i)} = \beta_{(6)}$	25.38***	13.70***	6.90***	5.00**	1.89	–
N	154,655	154,655	154,655	154,655	154,655	154,655
Fixed Effects						
Quarter-of-year	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Zip Code	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The second row indicates the length of the post-contract period used to calculate the production index. Standard errors are clustered at county level. System characteristics and county-level economic conditions are always controlled for. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table A6: Responses to Weather Updates

	Updates		Interactions		
Cancel = 1	(1)	(2)	(3)	(4)	(5)
$Index_{pre}$	–.039	–.854	–.591		
	[.246]	[.374]**	[.354]		
$Update$		–.769	1.170	.472	.588
		[.264]***	[.653]*	[.252]*	[.352]
$Dummy =$					
$1(Update < 0)$			–.000		
			[.005]		
$1(Index_{pre} > \overline{Index_z})$				–.014	
				[.004]***	
$1(Index_{post} < \overline{Index_z})$					.006
					[.004]
$Update \times Dummy$			–2.779	–1.705	–1.053
			[.600]***	[.332]***	[.280]***
cancellation rate	12.06	12.06	12.06	12.06	12.06
N	154,655	154,655	154,655	154,655	154,655
Fixed Effects					
Quarter-of-year	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes
Zip Code	Yes	Yes	Yes	Yes	Yes

Notes: $Index_{pre}$ is the normalized average production index for a 30-day pre-contract period. $Update$ is the average post-period index minus the pre-period one. $1(Update < 0)$ indicates whether $Update$ is negative. $1(Index_{pre} > \overline{Index_z})$ is an indicator of whether the pre-period index is higher than zipcode average, and $1(Index_{post} < \overline{Index_z})$ is whether the post-period one is lower. All coefficients are multiplied by 100 for legibility. $Dummy$ in the interaction term in the last rows refers to the dummy variable that is included in the same regression. Standard errors are clustered by county. System characteristics and county-level economic conditions are always controlled for.
* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table A7: Potential Selection by Pre-contract Weather (System Size)

System Size	(1) 1 month	(2) 1 month	(3) 2 months	(4) 2 months
$Index_{pre}$	.024	.020	.006	.004
	[.014]*	[.014]	[.013]	[.012]
Mean	4.69	4.69	4.69	4.69
R^2 (within zip code)	.204	.206	.204	.206
N	153,108	153,108	153,108	153,108
Fixed Effects				
Quarter-of-year	Yes		Yes	
County-by-quarter		Yes		Yes
Year	Yes	Yes	Yes	Yes
Zip Code	Yes	Yes	Yes	Yes

Notes: Standard errors are clustered by county. Other system characteristics and county-level economic conditions are always controlled for.
* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table A8: Heterogeneous Effects by Area Characteristics

	Panel A. Local Demographics			
Cancel = 1	(1) Median Income Above Median	(2) % White Above Median	(3) % College Above Median	(4) Household Size Above Median
$Index_{post}$	–.664	–.802	–.715	–.731
	[.386]*	[.292]***	[.321]**	[.337]**
$Index_{post} \times 1(Characteristic)$	–.125	.079	–.064	–.045
	[.283]	[.231]	[.239]	[.247]
	Panel B. Housing and Geographic Characteristics			
Cancel = 1	(5) % Urban > 0.5	(6) Housing Cost Above Median	(7) Northern California	(8) Non-coastal Counties
$Index_{post}$	.381	–.575	–.875	–.840
	[.442]	[.335]*	[.313]***	[.261]***
$Index_{post} \times 1(Characteristic)$	–1.211	–.249	.177	.314
	[.362]***	[.233]	[.211]	[.260]
Fixed Effects				
Quarter-of-year	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
Zip Code	Yes	Yes	Yes	Yes

Notes: The production index is normalized. All coefficients are multiplied by 100 for better display. Standard errors are clustered by county. Other system characteristics and county-level economic conditions are always controlled for. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.