Mauritzen, J. (2017). Cost, contractors and scale: An empirical analysis of the California solar market. *The Energy Journal, 38*(6), 177-197 DOI:
http://dx.doi.org/10.5547/01956574.38.6.jmau

Cost, Contractors and Scale: An Empirical Analysis of the California Solar Market

Johannes Mauritzen
BI Norwegian Business School
Department of Accounting, Auditing and Business Analytics
Trondheim Campus
Trondheim, Norway
johannes.mauritzen@bi.no
jmaurit.github.io

November 8, 2016

Abstract

I present an empirical analysis of the rapidly growing California rooftop solar photovoltaic market using detailed data of over 100,000 solar installations between 2007 and 2014. The rapid fall in the cost of solar panels stand central in the expansion of this market. I use a semi-parametric regression model to aid identification of cost factors by decomposing time-varying and cross-sectional components. I find that the use of Chinese manufactured panels are associated with costs that are 6% lower. Economies of scale at the local level (number of yearly installations in a zip code) and at the installation level (size of the installation) are also associated with lower costs. Higher subsidies, and higher contractor market-share are associated with higher costs. I use an exploratory analysis of the dominant contractor, SolarCity, to discuss non-cost factors in the expansion of the solar photovoltaic market.

1 Introduction

The market for rooftop solar has expanded rapidly over the last decade as costs have fallen. Solar photovoltaic power has become a viable energy alternative. The growth of rooftop solar has implications for electricity market structure and stability, grid infrastructure and operations, and climate change policy.

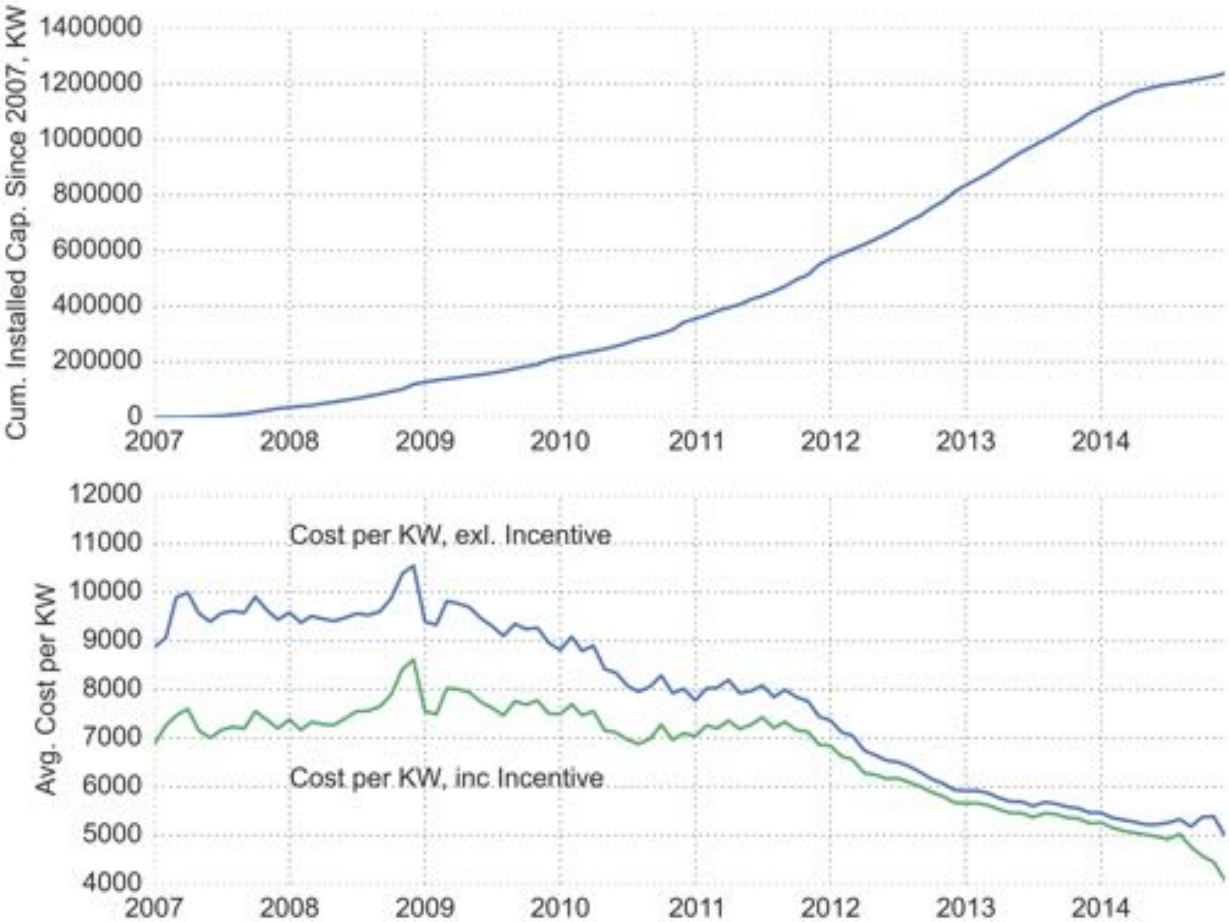

Figure 1: The top panel shows cumulative added installed capacity of rooftop solar photovoltaic generation in California since the beginning of 2007. The bottom panel shows the fall in average prices over time, with and without subsidies.

The rapid fall in solar panel costs has made rooftop solar panel systems profitable for many homeowners and businesses in California and elsewhere. The top panel in figure 1 shows the added cumulative capacity in roof-top solar in California since the beginning of 2007. The bottom panel shows the falling average price, both with and without California state subsidies in the same period.

Identifying the sources of cost variance is important in understanding the expanding rooftop solar panel market. For example, the extent to which the use of Chinese manufactured panels helps lower costs is important in judging the effects of trade tariffs that have been put in place following the time window of this study. Other relevant questions include whether subsidies have an indirect inflationary effect on costs[1], and the role of local economies of scale.

However, identifying and decomposing the variation in solar costs can be problematic. Over time, much of the price fall is explained by unobserved and non-linear variables of technological change and global economies of scale in solar panel manufacturing. At the same time, local observable variables of interest – like the effect of using Chinese panels, local economies of scale, and cost-inflationary effects of subsidies – can be expected to be correlated with the cost trends over time. A failure to properly control for the non-linear trend will then bias the estimated coefficients on the included observed variables.

To decompose the variation in costs and help identify the role of observable local factors, I use a semi-parametric model within the Generalized Additive Model frameworks of Hastie and Tibshirani [1990]. I use a cubic regression spline to account for the general non-parametric shape of the production profile, while allowing the observable variables to enter the equation linearly.

The results show that installations that used Chinese manufactured panels tended to be significantly less costly per kilowatt (kW) of installed capacity. Local economies of scale – defined by the number of installations per year in a given zip code – and economies of scale at the installation level – represented by the nameplate capacity of a given installation – were also associated with lower costs per kW. On the other hand, higher subsidies are associated with higher costs, all else equal, providing evidence for a cost-inflationary effect of subsidies.

In the period studied, increasing market concentration among the solar contractor firms is evident. However, contractors with large market shares cannot on average be shown to provide lower cost systems, all else equal. I therefore provide an extended discussion of potential non-cost explanations for the expanding solar photovoltaic market and increasing market concentration among the contractors. One rooftop solar company in particular stands out in California – SolarCity. SolarCity made up a total of 20 percent of the California market

[1] For example, for early studies of the role of subsidies on inflating costs in the transport and health care industries see [Pucher et al., 1983] and [Feldstein and Friedman, 1977]

by 2014 and changes in the company's strategy and business model has a substantial effect on the market as a whole. I use SolarCity as a case study, both as a tool for understanding the company's success in the midst of a crowded field of competitors, but more importantly also for understanding the drivers of the market as a whole.

2 Literature on Investments in Solar Power

The economics of solar photovoltaic power are unique within power generation. Unlike traditional power plants, most solar photovoltaic is non-dispatchable, with the level of solar irradiance and deployed solar capacity largely determining output at any given time. Baker et al. [2013] and Borenstein [2008] provide overviews of the economics of solar power with a focus on its intermittency and the short- and long-term implications for power markets of increased solar penetration.

Solar power is also different from most other generation because investment decisions are made by individual consumers and small businesses.[2] Large energy companies have considerable expertise in generation technologies and engineering, investment risk, electricity market structure and other specialized knowledge and competencies involved in generating electricity. A consumer or small business, on the other hand, can be expected to have much more limited knowledge and expertise.

Informational and behavioral issues therefore become important factors in analyzing investment decisions. For example, Dastrup et al. [2012] argue that solar panels cannot be considered a pure investment good, but are also bundled as a type of green conspicuous consumption. The authors support this argument by showing how the installation of solar panels affects home prices in the San Diego area and finds evidence for a "solar price premium", which is positively correlated with a measure of a given neighborhood's environmental awareness. Bollinger and Gillingham [2012] study the the role of peer effects in solar

[2]Due to the widespread popularity of leasing arrangements, ownership of the solar assets is often in the hands of the contractors.

photovoltaic adoption. They find evidence that the adoption of solar panels by homeowners in a certain zip-code will increase the probability that other households in that zip-code will install solar panels.

Recently, several articles focusing on the policy impact and efficiency of solar subsidies in general and the California Solar Initiative (CSI) in particular have appeared. Hughes and Podolefsky [2015] use variation in rebates across electric utilities to find that CSI subsidies have a large effect, and that installations would have been more than halved without incentives. In a working paper, Burr [2014] estimates a dynamic structural model using CSI data. She finds that upfront subsidies and performance-based subsidies provide roughly equivalent effectiveness in promoting solar photovoltaic installations. However, performance-based subsidies will tend to encourage better siting of panels. The author also argues that most of the solar photovoltaic investments in the time period studied would not have been made without subsidies.

While the economics literature on solar policy is growing, empirical analysis of cost variation in the solar photovoltaic market are scarce. Several articles have analyzed the dramatic reduction in photovoltaic costs over time [Nemet, 2006, Candelise et al., 2013], but do not attempt to decompose the considerable cross-sectional variation. [Wiser et al., 2006] presents an analysis of local variation in costs in California, but the analysis goes only up to 2005. Since that time, the market has expanded by more than a factor of 20 and the structure of the market has changed substantially. This article aims to provide an up-to-date analysis of costs in the California solar photovoltaic market, using a methodology that is effectively able to decompose global from local factors.

3 Data and The California Solar Initiative

I use publicly available data[3] on approximately 100,000 solar photovoltaic systems installed in the state of California between 2007 and 2014. A cleaned data set is also available on my

[3] http://www.californiasolarstatistics.ca.gov/

	5%	25%	50%	75%	95%
Date of Installation, Years Since 2007	1.53	3.74	5.24	6.34	7.26
Cost, $ per kW	3850.00	4900.00	5730.00	7246.38	9677.08
Cost, $ per kW, inc. Subsid.	3570.45	4619.53	5149.24	6182.47	8658.76
Nameplate Capacity, kW	2.20	3.70	5.16	7.00	11.98
Total # Observations	106551				
% Leased	52.5				
% Using Chinese Manufactured Panels	28.5				

Table 1: Summary statistics for California rooftop solar panel installations from 2007 through 2014

website.[4]

The data includes all installations covered by the California Solar Initiative, which provided rebates for solar panel installations on existing single and multi-family homes, commercial and governmental buildings. Large utility-owned projects are not included in this program. The dataset includes variables on the size of the system, installation date, the amount of subsidy provided by the state, the location of the installation, the contractor who installed the system and the manufacturer of the component panels and inverters. Table 1 1 shows summary statistics for key variables.

The California Solar Initiative was launched in January of 2007 and scheduled to last until the end of 2016 or until the allocated funds of approximately 2.1 billion dollars were exhausted [California Public Utilities Commission, 2014]. As of the end of 2014, approximately 1700 megawatts (MW) out of a goal of 1940 mW was installed. The rebates covered customers of the largest three investor-owned utilities – Pacific Gas and Electric Company, Southern California Edison, and San Diego Gas and Electric – combined representing approximately 70 percent of California's load. The size of the subsidy depends on the size of installation, as well as how much capacity had already been installed state-wide; the subsidies were designed to decline over time as more capacity was installed.

In addition to the California solar incentives, incentives at the federal level also exist to

[4] `http://jmaurit.github.io\#calsolar2`

encourage solar power investment. The Investment Tax Credit (ITC) for solar, established in 2006, provides a 30 percent tax credit for solar systems on residential and commercial properties. In addition, solar power system owners benefit from an additional tax benefit since they qualify for the Modified Accelerated Cost Recovery System (MACRS), which allows for an accelerated cost depreciation over five years.

Both of these programs were constant in the period studied, so I do not directly address them in the analyses. However, I do discuss the potential effects of the ITC in distorting the reported cost data in the results section.

4 A Semi-parametric Model of Solar Panel System Costs

A problem with estimating a model of solar panel system costs is the existence of unobserved time-varying variables. Such unobserved variables are primarily composed of technological change and economies of scale in the production of component solar panels and inverters. These unobserved variables are likely correlated with the local variables of interest over time and are likely to bias the results. More so, the shape of the unobserved function over time is likely to be highly non-linear, reflecting bursts of technological progress or increased economies of scale in manufacturing.

To control for the effects of the unobserved variables over time, I use a semi-parametric model within the General Additive Model frameworks of Hastie and Tibshirani [1990]. I use a smoothed cubic regression spline to control for the unobserved effects of technological change and economies of scale over time, while the observed variables of interest enter linearly. In the economics literature, such models are known as partial linear models [Yatchew, 1998]. Because of the inclusion of this non-parametric function, the linear variables of interest can be interpreted conditionally on the general cost level over time and will not be biased if the variable is correlated with the direction of the general cost trend.

The model can be written as in equation 1.

$$
\begin{aligned}
Log(cost\_per\_kw_i) = \alpha &+ f(time\_since\_2007_i) + \delta_{\mathbf{sector}} \\
&+ \beta_1 nameplate\_kw_i + \beta_2 county\_year\_total_i + \beta_3 zip\_year\_total_i \\
&+ \beta_4 incentive\_per\_kw_i + \zeta_1 lease_i + \zeta_2 chinese_i + \epsilon_i
\end{aligned}
\tag{1}
$$

Here the left-hand-side variable is the log costs per kW of installed nameplate capacity, $Log(cost\_per\_kw_i)$. On the right hand side is the non-parametric function of time, $f(time\_since\_2007_i)$, that is meant to capture the non-linear effect of technological change and manufacturing economies of scale through time. The variable is measured in years since January 1st, 2007 – the date from which the data is available.

I include fixed effects for the host sector of the solar photovoltaic system, which is represented by $\delta_{\mathbf{sector}}$. The categories included are residential, commercial, non-profit and government. Because of the inclusion of these fixed effects, the variables of interest should be interpreted as "within" estimates from the four sectors.

The nameplate capacity, $nameplate\_kw_i$ of each solar panel system is included in the model in order to capture the effects of cost economies of scale in the size of the solar system. Within a reasonable range of system size, some cost components can be assumed to be fixed or at least vary less than one-to-one with size, and therefore lead to economies of scale. Inverters[5], permitting, and marketing are examples of cost components that likely vary less than one-to-one with size.

The variables $county\_year\_total_i$ and $zip\_year\_total_i$ represent the amount of capacity installed for a given year in, respectively, the same county and zip code of a solar photovoltaic system. These variables are meant to capture the local economies of scale in the market. For example, fixed contractor costs such as advertising and marketing, could be spread over more installations. Plausibly, in an area with many installations, costs could also be pressed

[5]An inverter converts the DC current generated by the solar photovoltaic system into AC current compatible with the grid.

down by competition between contractors.

The variable $incentive\_per\_kw_i$ represents the subsidy provided by the state of California per kW of nameplate capacity. Since the left-hand-side variable, $cost\_per\_kw_i$, represents costs before incentives, the inclusion of the subsidy on the right hand side does not, then, simply reflect an identity. Instead, the inclusion of the variable is meant to capture indirect effects that subsidies have on costs. Lower subsidies may lead contractors to focus on their costs and to lower their prices in order to stay competitive. Conversely, the presence of generous subsidies may lead to inflated costs.

Two dummy variables are included in the model: $lease_i$ and $china_i$. $lease_i$ represents whether a solar panel system is owned by the host or whether it is third-party owned, most often by the contractor or a subsidiary. In the industry, a lease most often refers to a payment of a fixed monthly fee, that will increase at a contractually agreed upon rate. The indicator variable $lease_i$ also includes power purchase agreements, where payments by the host are made based on the actual power produced at a contractually agreed upon price.[6] What both these arrangements have in common is that the solar photovoltaic system host makes little to no down-payment, nor do they own the solar system. Instead, they sign a contract for the long-term leasing of the system or the purchasing of the power from the system – most commonly for 20 years.

Ideally the variable $lease_i$ would capture the real cost variation between leased solar photovoltaic systems and those owned outright. In reality, the variable will also capture differences in reporting costs. For a system that is sold outright, the reported cost is simply the price paid to the solar contractor for the system. For a leased system, the reported cost is rather an estimate of the total of component costs plus a mark-up for general sales and administrative costs and a profit margin. This cost estimate is also supplied to the federal government in order to receive the federal investment tax credit (ITC) for solar photovoltaic investments, which covers 30 percent of the investment cost. However, this also gives the

[6] The contractual price in a PPA need not be fixed, but can and usually is designed to increase over time

contractors an incentive to inflate their reported costs in order to claim a larger tax credit.

The dummy variable $china_i$ indicates whether solar panels used in the installation came from a Chinese manufacturer. Anecdotal evidence suggests that the emergence of the Chinese solar panel manufacturing industry and the eventual concentration of manufacturing scale in that country was a driver of lower-cost panels and in turn installations.[7] The effect that the emergence of China as a solar photovoltaic manufacturing hub had on average prices over time would be captured by the smooth term. The coefficient on the *china* variable should instead capture the direct effect on costs of choosing panels from a Chinese manufacturer at a given point in time.

The smoothed function is estimated using a cubic regression spline. A cubic polynomial function is used to fit the shape in sections, separated at points known as knots, but continuous up to the second derivative. The regression spline can be represented in linear form, $\boldsymbol{X\beta}$, and thus standard, efficient matrix algebra techniques can be used to fit the model.

The smoothed component of the model is fit by minimizing equation 2.

$$\|\mathbf{y} - \mathbf{X}\beta\|^2 + \lambda \int_0^1 [f''(x)]^2 dx \tag{2}$$

The latter term, $\lambda \int_0^1 [f''(x)]^2$, is an estimate of the second derivative of the function. This serves as a penalty for the "wiggliness" of the function. The λ can be adjusted to control the level of smoothing. Instead of setting λ arbitrarily, however, cross-validation is used. Intuitively, each data point is, one-by-one, left out and the smooth term that provides the best average predicted fit over all the data is chosen. For further details, I refer to Wood [2006].

With a single-variable cubic regression spline, the smoothed terms can be interpreted directly, and thus we can do a visual check of the appropriateness of the smoothed function. Figure 2 shows the estimated smoothed function of costs over time. All other co-variates

[7] See for example the article from The Economist `http://www.economist.com/news/business/21696941-solar-power-reshaping-energy-production-developing-world-follow-sun?zid=313&ah=fe2aac0b11adef572d67aed9273b6e55`

are held at their average value, except the fixed-effects and dummy variables. The $sector_i$ variable is held at the value indicating a residential host, $lease_i$ is held at zero, and $china_i$ is also held at zero. The shaded value represents a point-wise 95% confidence interval. As can be seen, with over 100,000 observations, the curve can be estimated fairly precisely.[8].

The non-parametric function shows an initial period without a clear directional trend. However, following 2010 a steep downward cost trend is evident. This coincides with the emergence and expansion of Chinese photovoltaic panel manufacturing and export. In addition, two prominent bumps are apparent in the smoothed function corresponding to approximately 2009 and 2011. A likely explanation can be seen in figure 3. In the figure, each point represents the log-transformed cost per-kW of a solar power system installation. The visible dark horizontal bands coincide with large lease contractors who appear to change their reported costs in discontinuous jumps. This could plausibly lead to the observed bumps in the smooth function. Later in the article, I discuss in more detail the effect of reported-costs of leased systems and how that may affect the results of the analysis.

I show results from a model specification that includes county fixed effects in order to control for regional variation in costs that may be present. Plausibly, the different host sectors could experience substantially different cost trends over time. In addition, as discussed earlier, the $lease_i$ indicator variable likely captures both real differences in costs associated with 3rd-party ownership of solar panel systems, as well as variation in reported costs. The fixed effects controlling for the differences between sectors and ownership-models may not sufficiently take into account these idiosyncratic variance components and their correlation with the smoothed term.

To control for these sources of variation, I also run a regression with *sector* and *lease* random effects. The smoothed function can then be written as $f(time\_since\_i, t * Z)$. Where $Z \sim N(0, \sigma)$ and σ is a vector of variance terms to be estimated that represents the host

[8]This confidence interval is based on a Bayesian simulation of the posterior distributions. This is analogous to confidence intervals based on bootstrapped standard errors, but where the setting of Bayesian priors based on the penalized form of estimation allows for better computational efficiency. For details I refer to Wood [2006]

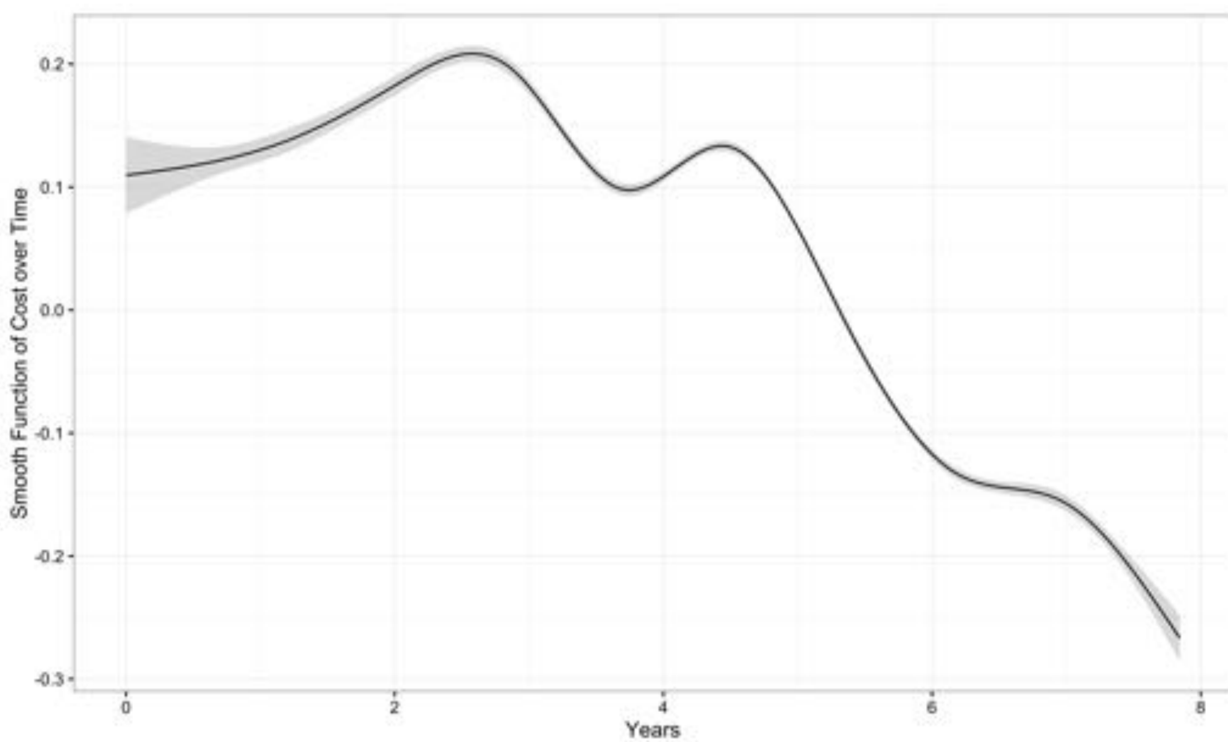

Figure 2: The smoothed function of solar panel system costs over time. The x-axis is measured in years since 2007. The y-axis can be read as the log difference from the mean value. -.1 could then, for example, be read as average costs being 10% below the mean value for the period studied. All other continuous variables are set at their mean value. The shaded area represents a point-wise 95% confidence interval. After approximately 2009, a steep and sustained downward trend in solar panel costs is evident.

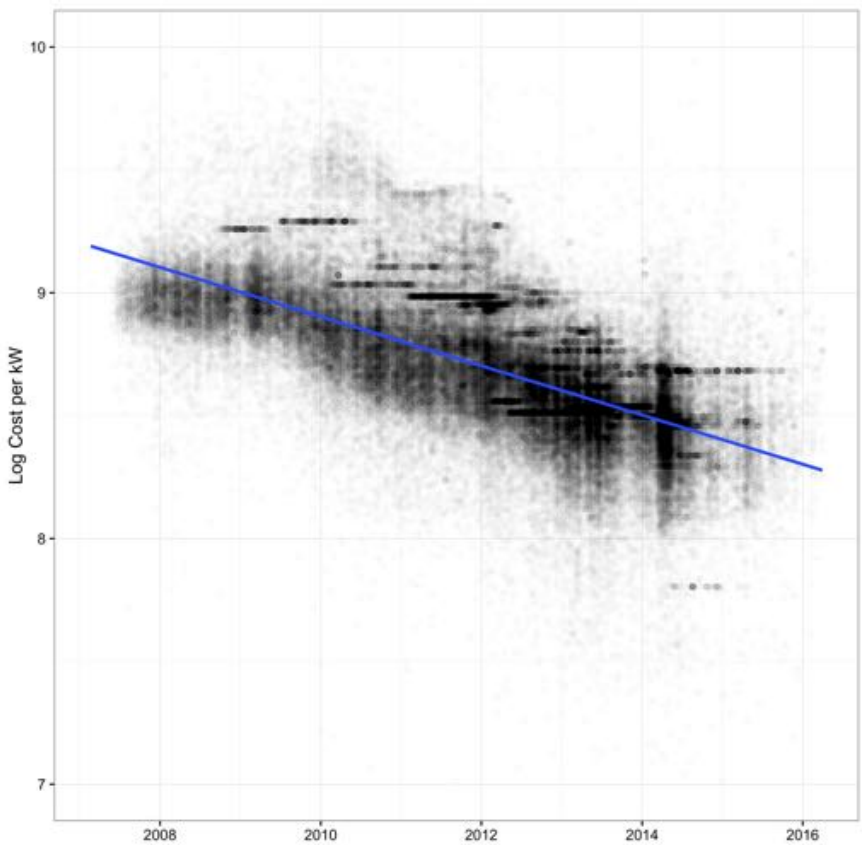

Figure 3: Each point represents the log-transformed cost per-kW of a solar power system installation. The visible dark horizontal bands coincide with large lease contractors who appear to change their reported costs in discontinuous jumps. This is a likely explanation for the bumps visible in the smoothed function.

sectors, the *lease* indicator and their interactions.

5 Results

The results of the regression are shown in table 2 below. The first column shows results from the simple model with sector fixed effects. The second column shows results when county fixed effects are added. The third column shows results from the model with random effects for host sector and lease. My preferred model is with sector and county fixed effects. Both the deviance explained metric as well as the R^2 indicates that adding county-level factors substantially improves the fit of the model. On the other hand, the model with sector and lease random effects appears not to substantially add to the goodness-of-fit.

In the regression, the commercial sector is left out as the comparison factor. Residential installations are shown to have average costs nearly identical to commercial. However, the non-profit sector is shown to have costs nearly 9% lower while the government sector has costs that are approximately 13% higher than the base case. Higher costs for the government sector could potentially reflect procurement regulations. For example, some local governments only award contracts to companies with a unionized workforce. The higher costs could also reflect an agency problem. Government employees may not have the right incentives in place in order to find the most cost-effective solar panel contractor.

The reasons for the lower costs for non-profit hosts is not altogether clear. However, some contractors, like the non-profit Grid Alternatives, do offer reduced cost installations to low-income homeowners and non-profit organizations.

The estimated parameter on the $nameplate\_kw_i$ variable indicates that larger solar photovoltaic systems tend to have slightly lower costs per kW of capacity. However, the estimated effect is small in magnitude, with on average a 10 kW increase in system size leading to only about a .3% decrease in cost.

The coefficient on the variable $zip\_year\_total_i$ gives evidence for the existence of local

	Sector FE	County & Sector FE	County FE, Lease & Sector RE	Diff-in-Diff
Intercept	8.6171^{***}	8.6608^{***}	9.0724^{***}	10.98^{***}
	(0.0067)	(0.0077)	(0.0074)	(0.0083)
Government Sector	0.1169^{***}	0.1307^{***}		0.1754^{***}
	(0.0125)	(0.0124)		(0.0122)
Non-Profit Sector	-0.0920^{***}	-0.0960^{***}		-0.02
	(0.0145)	(0.0143)		(0.0143)
Residential Sector	0.0159^{*}	0.0155^{*}		0.0329^{***}
	(0.0065)	(0.0064)		(0.0065)
nameplate	-0.0005^{***}	-0.0003^{***}	-0.0005^{***}	-0.0004^{***}
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
county_year_totals (MW/year)	0.0010^{***}	-0.0000	-0.0042^{***}	0.0002
	(0.0001)	(0.0002)	(0.0002) (0.0001)	
zip_year_total (MW/year)	-0.0347^{***}	-0.0130^{***}	-0.0150^{***}	-0.0390^{***}
	(0.0031)	(0.0033)	(0.0033)	(0.0031)
incentive_per_kw ($10/kW)	0.0012^{***}	0.0010^{***}	0.0006^{***}	
	(0.0000)	(0.0000)	(0.0000)	
contractor_market_share %	0.0016^{***}	0.0015^{***}	0.0024^{***}	0.0010^{***}
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
lease	0.0205^{***}	0.0183^{***}		0.0194^{***}
	(0.0021)	(0.0021)		(0.0021)
china	-0.0575^{***}	-0.0564^{***}	-0.0700^{***}	-0.0500^{***}
	(0.0024)	(0.0023)	(0.0024)	(0.0024)
steps				13.7186^{***}
				(0.9644)
year				0.0591^{***}
				(0.0041)
steps:year				-0.0068^{***}
				(0.0005)
EDF: s(time_years)	8.9142^{***}	8.9072^{***}		8.9201^{***}
	(8.9979)	(8.9975)		
EDF: s(time_years,lease,sector)			7.9536^{***}	
			(8.0000)	
AIC	1876602.0511	1873838.6614	1878554.7616	1870358.3493
BIC	1876802.3335	1874536.8480	1879205.5109	1870577.7749
Log Likelihood	-938280.1113	-936846.4235	-939209.4272	-935156.2546
Deviance	277851294007.7000	270473798188.6884	282740536291.1357	274823699286.9071
Deviance explained	0.3758	0.3924	0.3648	0.3817
Dispersion	2608171.0511	2540158.7390	2655238.4685	2587231.5087
R^2	0.3757	0.3920	0.3644	0.3816
GCV score	2608658.6051	2541874.1518	2656907.9934	2587765.4073
Num. obs.	106551	106551	106551	106245
Num. smooth terms	1	1	1	1

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 2: Model results. Estimated coefficients are shown for the linear terms. Standard errors are in parenthesis. For the smoothed terms, estimated degrees of freedom are displayed, where the p-values are from F-tests of whether the smooth terms significantly improve the fit of the model.

economies of scale in solar photovoltaic systems. An additional 10 MW of installed capacity in a given year and given zip code will lead to solar photovoltaic systems that are between 13% and 35% cheaper in the same year and zip code. I use the results from the model without county fixed effects as an upper bound as the fixed effects likely absorb substantial amount of the relevant variation.

The economies of scale appear to be mainly at the geographical level of zip codes and not the larger county-level geography. When leaving out county level fixed effects, the $county\_year\_total_i$ variable is close to zero. This can be interpreted to mean that economies of scale are at best slight when the effects of scale at the zip-code geography are taken into account.

The coefficient on the variable $incentive\_per\_kw_i$ is meant to capture potential inflationary effects that subsidies may have on solar panel costs. The estimated coefficient appears to indicate a slight but positive and significant inflationary effect. An additional $100 of incentives per kW of capacity is estimated to increase the cost of an installation by approximately 1%.

However, some care is warranted in interpreting this result. Larger subsidies will tend to encourage larger installations, and thus the effects of subsidies and size on cost per kW could be conflated. This issue is at least partially dealt with by the inclusion of the nameplate capacity variable as a controlling variable for installation size. In addition, the design of the subsidy scheme suggests another identification approach. Subsidies were designed to decrease in steps, from 1 to 10 – 1 indicating the highest per-kW subsidy – according to the total installed capacity in each utility area. Within a limited range of total installed capacity, the exact point where the subsidy is lowered can be considered to be arbitrary, and a difference-in-difference approach could be used for identification of the subsidy inflationary effect.

I run a regression similar to the preferred model, but instead of including the variable $incentive\_per\_kw$, I include the variables $steps$, $years$ and their interaction in order to make a difference-in-difference estimation. The variable *steps* is an integer between 1 and 10 that represents the subsidy level. The variable *year* represents the year of installation. The coefficient on their interaction is the difference-in-difference estimator. The identifying assumption is that the exact cut-off point of when a new subsidy is put in place is arbitrary within a limited time span. The coefficient on the interaction effect represents the contrast between the per-kW cost directly before and after a change in the subsidy step within a given year and with all other observed factors held constant.

I show the results of the difference-in-difference estimation in the fourth column of table 2. The coefficient on the interaction effect, $steps : year$, is shown to be slightly negative but significant. Since an increasing step corresponds to a lower subsidy, the sign of the coefficient is also consistent with a slight inflationary effect of solar subsidies.

While geographic economies of scale are associated with lower costs, larger contractors cannot be shown to have a cost advantage. For a one percentage point increase in the state-wide market share of a contractor in a given year, average costs increase by approximately .15%. While the magnitude of the effect is slight, it appears that some contractors are gaining market share without necessarily having a cost advantage over their competitors. In the next section, I discuss potential non-cost market drivers.

Finally, the $lease_i$ and $china_i$ indicators have the expected sign. Leased solar systems are reported to have higher costs than those sold outright. While good reasons exist for believing that a leased system may bare higher real costs, because of the federal investment tax credit, contractors and third-party-owners have a perverse incentive to inflate their reported costs in order to benefit from a larger tax credit. The magnitude of the effect, whether it reflects real or artificial factors, is nonetheless modest. Leased systems have on average 2% higher costs than those sold outright, all else equal.

The model results also provide evidence that the use of Chinese panels confers a substantial cost advantage. Systems that used Chinese panels tended to be on average 6% less costly than other installations. However, as noted earlier, this magnitude only reflects the direct effect of using Chinese manufactured panels verses other panels at a given time.

In addition to the comparison of the four model specifications above, I have run several robustness checks. Cross validation, used to obtain an optimal smoothed curve, may lead to over-fitting if the residuals of the model are serially correlated [Dimitropoulos and Yatchew, 2016]. An autocorrelation function (ACF) plot of the residuals did not, however, detect any significant autocorrelation.

As an additional model robustness check, I ran a differencing model [Yatchew, 1998] of the ordered dataset along the time direction. This is a much simpler semi-parametric model that nonetheless takes into account the non-linearities of the unobserved components that vary over time. The results from this model, presented in the appendix, are generally in line with the results presented above.

6 Non-cost drivers of the California Solar Photovoltaic Market: A Case Study of SolarCity

A trend that comes across clearly in the data is an increasing market concentration among contractors after 2010. In particular, as the top panel in figure 4 shows, the solar photovoltaic installer and contractor firm, SolarCity was able to capture as much as 20 percent of the market by 2014. This is all the more striking since there are several hundred active contractors and installers in the market, and there appears to be few barriers to entry.

At the same time, the analysis above does not provide any evidence that would suggest that firms gained greater market share by being low-cost providers of solar photovoltaic systems. In fact, the regression indicates that increased market share was associated with somewhat higher costs.

In this section I explore non-cost explanations for why some contractors were able to both expand their market share, and presumably also expand the market as a whole. Because I am studying only a few firms, a rigorous econometric analysis is not feasible, as the actions

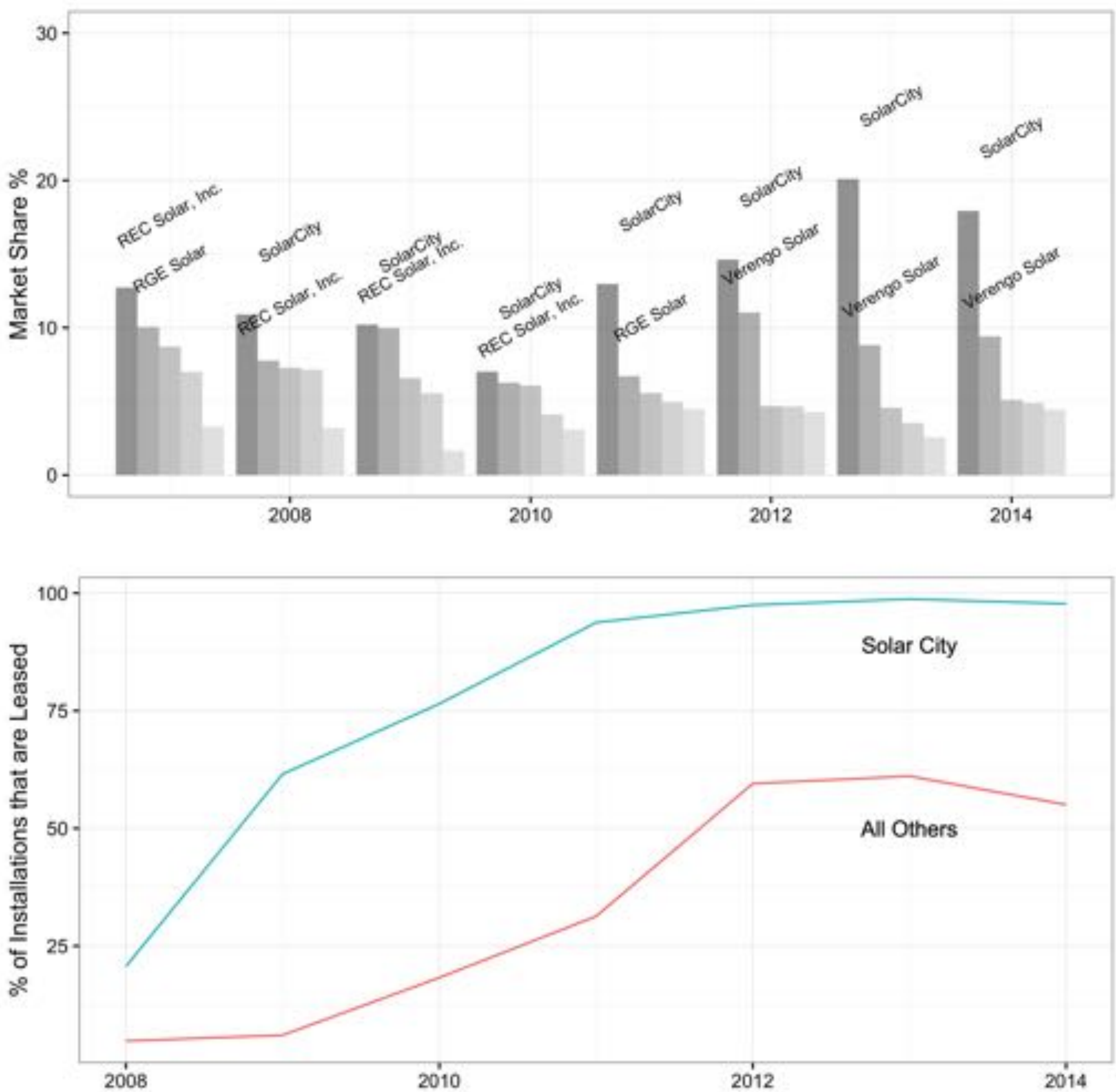

Figure 4: The top panel shows the market share of the largest five contractors in a given year. After 2010, a pattern of increasing market concentration is evident, with the contractor SolarCity gaining as much as 20 percent of the market. The bottom panel shows the percentage of installations from SolarCity and all other firms that are leased rather than sold outright, suggesting a mechanism for SolarCity's gain in market share.

of a single firm would tend to dominate the analysis.

Instead, I focus on the strategies of the dominant contractor, SolarCity, and take an exploratory approach to the data in order to try to glean insights into the market. This section should then be considered as an extended discussion, where the findings are necessarily suggestive rather than conclusive.

The increasing use of third-party ownership of solar photovoltaic systems – I will simply call it leasing from now – is one potential driver of both increased installations and increased market concentration in the contractor market. As the bottom panel of figure 4 shows, leasing became generally more popular in the market over time. However, SolarCity quickly increased the share of their installations that were leased after 2008, and nearly all their installations were leased by 2012. Further analysis reveals that most of the leased installations were handled by large contractors: Leasing and increased market concentration are strongly correlated.

Leasing is attractive to consumers and businesses for several reasons. They provide cash-constrained homeowners and small businesses the ability to place solar photovoltaic systems on their roofs without a large initial expenditure that may require financing. In addition, a leased solar photovoltaic system can allay uncertainty about the complexities of owning a solar panel system. When a system is leased, the host customer generally is not responsible for repairs and maintenance, which homeowners or businesses may consider a nuisance. Leasing can then attract new demographics to host photovoltaic systems [Drury et al., 2012].

Importantly, a lease also helps allay uncertainty related to a homeowner generating their own electricity. Solar photovoltaic panels along with supporting equipment like the inverter are complicated and sophisticated technology where quality can vary greatly between suppliers. At the same time, quality can be difficult and expensive to verify by a homeowner or small business with limited technical and financial resources. I discuss this further below in relation to the introduction of Chinese manufactured solar panels.

A leasing model also confers several advantages to a contractor who is able to offer the service. The market of potential customers is expanded to those that are cash constrained, but could otherwise benefit from solar panels. At the same time, the electricity generated from the solar panels provides a relatively certain guarantee of future income from the investment, irrespective of the credit-worthiness of the host consumer.

While the wider solar contractor market has few barriers to entry, the narrower market for leasing solar photovoltaic systems does impose significant hurdles for potential entrants. Solar panel installations are capital intensive, and the calculated levelized cost of electricity are heavily dependent on financing costs. A contractor wishing to offer a solar leasing service needs access to substantial amounts of capital at relatively inexpensive cost.

Between 2012 and 2015, SolarCity was able to raise $80 million through a 2012 IPO as well as the issuance of 3.4 million shares in common stock in 2013 at a price of $46 per share. They secured financing of over $1.5 billion from banks including Goldman Sachs and JP Morgan. In addition, SolarCity was the first company to issue Asset Backed Securities (ABS) based on the income stream of their solar photovoltaic assets.[9] This level of large and sophisticated financing is available to only a few firms with appropriate scale and creditworthiness. Access to finance amounts to a substantial barrier to entry.

In the model results, the use of Chinese manufactured panels was shown to be associated with lower-cost photovoltaic systems. The data also indicates that SolarCity was quicker and more decisive than the rest of the market to switch to cheaper Chinese panels. Figure 5 shows the four largest suppliers of solar panels to SolarCity from 2007 through 2014. Up to 2011, SolarCity had used panels from established manufacturers based in Europe (BP Solar, SolarWorld), the US (Evergreen, First Solar) or Japan (Kyocera, Sharp, Sanyo). However, in 2011 SolarCity began sourcing a substantial number of panels from a Chinese manufacturer, Yingli. In the years that followed, SolarCity sourced the vast majority of their panels from Chinese manufacturers (Yingli, Trina, AU Optronics).

[9]Information on SolarCity's financing was gathered from the company's press releases: `http://www.solarcity.com/newsroom/press`.

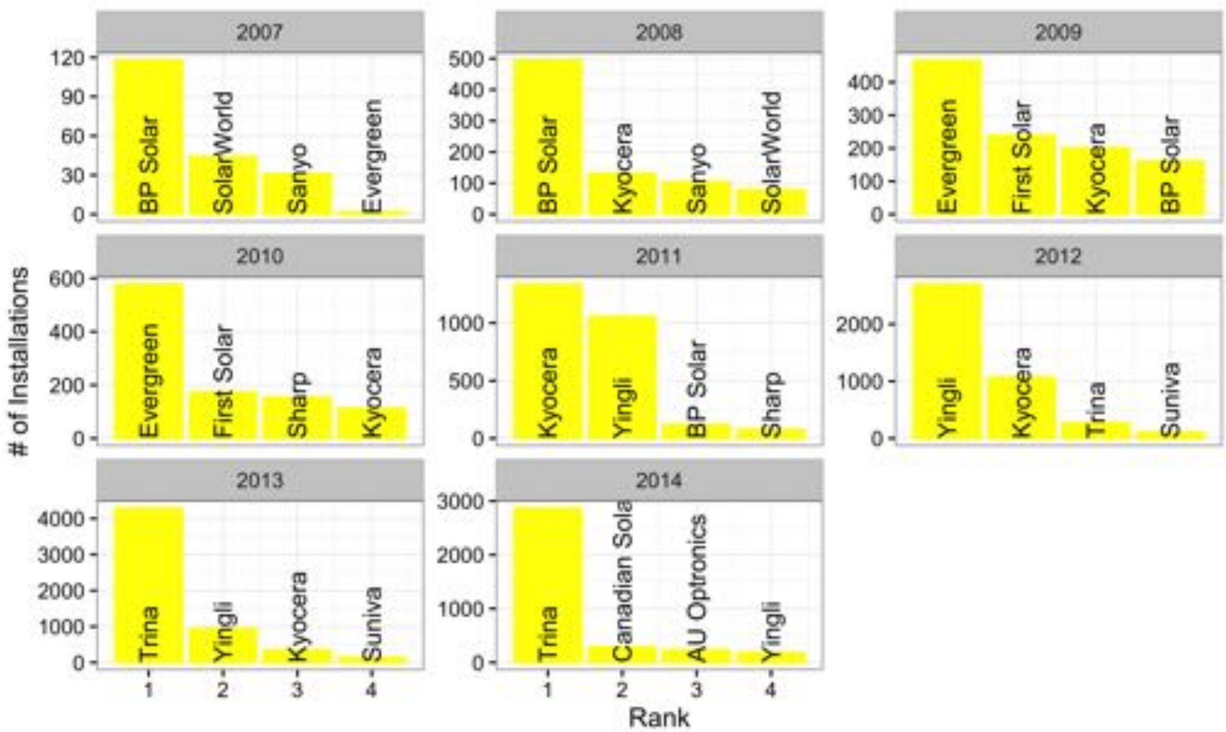

Figure 5: The figure shows the four largest suppliers of solar panels to SolarCity for each year from 2007 through 2014. After 2011, SolarCity began to increasingly rely on panels from Chinese manufacturers.

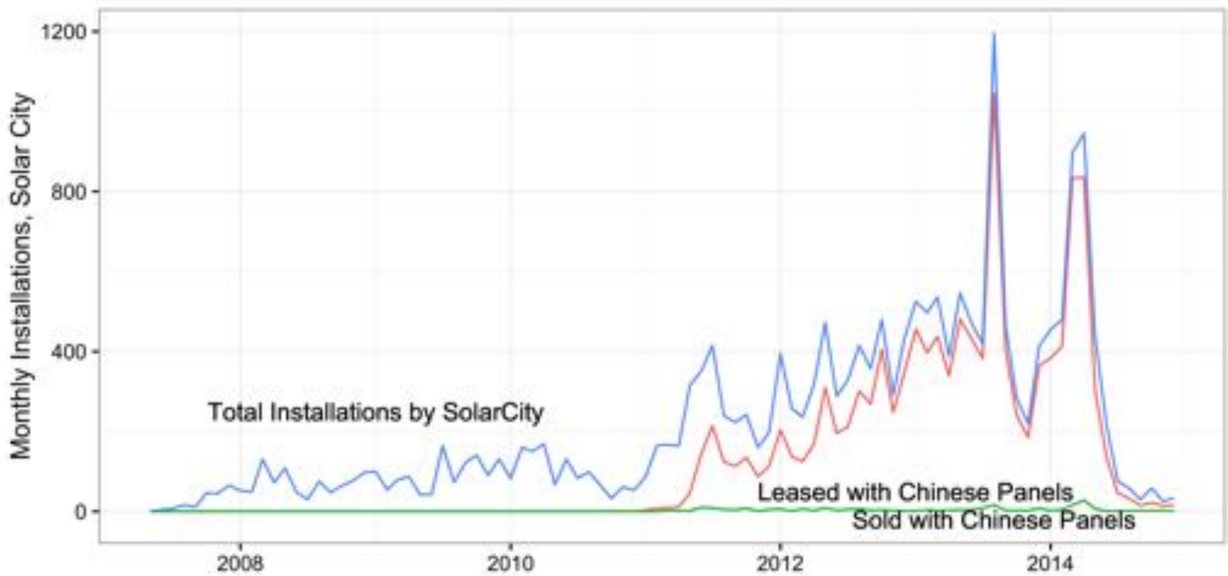

Figure 6: The figure shows that the sourcing of Chinese manufactured solar panels coincided with SolarCity's use of a leasing business model.

Figure 6 shows that the introduction of Chinese panels by SolarCity coincided to a great extent with both an increased use of leasing and with a strong growth in monthly installations. By 2013, leased solar photovoltaic systems that used Chinese panels made up the large majority of solar panel installations completed by SolarCity.

Cheaper Chinese solar panels likely helped make a leasing model financially feasible for large contractors. Presumably, the increased prevalence of the leasing model also helped expand the market substantially, as discussed earlier. However, a subtler rationale also exists for the observed relationship between the introduction of Chinese solar panels and a leasing business model that comes from economic theory of asymmetric information of quality.

Solar panels are long-lived assets that currently must last at least a decade in order to be financially profitable for the owner. More so, judging the quality of solar panels is beyond the technical abilities of the vast majority of consumers and contractors and thus most will rely on reputation and ratings of existing manufacturers.

However, this presents a problem for Chinese manufacturers that have not had an earlier presence on the market. A type of "lemons" problem of asymmetric information arises [Akerlof, 1970]. Consumers, with poor information on the quality of panels from new Chinese manufacturers, will be weary of purchasing them. At a minimum they will demand a lower price than a comparable system with panels from an established manufacturer. The finding in the regression that solar photovoltaic systems that used Chinese manufactured panels were substantially cheaper may, in part, reflect this.

Consumers likely had reason to worry about the varying quality of Chinese manufactured panels. A New York Times article [10] from 2013 details how a spate of solar photovoltaic systems – primarily using Chinese panels – saw significant rates of failure and deterioration after only a couple years of use.

A contractor that offers a leasing model, however, can overcome these information asymmetry problems. In effect, they can aggregate the information asymmetry and attending risk and deploy resources to ensure the quality of a supplier.

While verifying the quality of panels from a previously unknown manufacturer is expensive, a large contractor can take steps like having experts test the quality of modules and visiting manufacturing facilities that ordinary homeowners and businesses would find prohibitive. In fact, specialized companies, like Solar Buyer (`http://solarbuyer.com`), exist to inspect and verify the quality of solar panels – for a substantial fee.

More so, a leasing model is likely superior to issuing a guarantee in overcoming the information asymmetry. A guarantee issued by a contractor to a homeowner is good only as long as the contractor remains solvent. Since the solar contractors are themselves often new firms, such a guarantee may not be seen as sufficient.

[10] `http://www.nytimes.com/2013/05/29/business/energy-environment/solar-powers-dark-side.html?pagewanted=all&_r=0`

7 Conclusion

Solar photovoltaic systems have become attractive for individual homeowners, businesses and government organizations to install and operate in many parts of the world – thanks in part to rapidly falling cost and government incentives. The distributed nature of the investment decision distinguishes photovoltaics from most other forms of electricity generation. The decision of whether or not to invest is not made by an informed electricity utility executive, but rather by regular home- and business-owners with limited industry knowledge and financial and engineering resources.

Investment costs in photovoltaics vary greatly. Global factors such manufacturing economies of scale and technological change affect costs over time. The distributed nature of photovoltaics also means that local factors affect costs across geography, demographics, and the structure and strategies of contractors. In this article I use a semi-parametric regression to decompose cost variation into a smoothed non-linear cost curve over time and linear unbiased estimates of the coefficients on local factors.

I find that the use of panels from Chinese manufacturers are associated with costs that are on average 6% lower. Additionally, I find evidence of local economies of scale. On average an additional 10 MW of capacity installed in a certain year in a given zip code is associated with solar photovoltaic systems that are between 13% and 35% less costly. Evidence is also found for subsidies having an inflationary effect on costs. A $100 increase in subsidies per kW of capacity is associated with an increase in costs of approximately 1%.

The time period studied was characterized by increasing concentration of the photovoltaic contractor market. However, this trend does not appear to be driven by costs. Contractors with higher market shares are actually associated with having slightly higher costs. In an extended discussion and case study, I explore non-cost explanations for both increasing concentration in the contractor market as well as for the expansion of the solar photovoltaic market as a whole.

I suggest that the introduction of leasing played a pivotal role. Leasing loosened consumer capital constraints by switching ownership to large contractors with access to multiple sources of financing. Transferring ownership to large contractors may also have allayed concerns about both the complexity of maintaining a solar photovoltaic installation as well concerns about the quality of components.

Exploring solar public policy is not the primary objective of this article, but several implications from the research do emerge. A direct policy implication concerns the design of flexible subsidies. In several US states only homeowners who themselves own the solar system on their roof can collect government production subsidies. For example, legislation was recently introduced in North Carolina to allow 3rd-party owners of solar panel systems to benefit from net-metering rules.[11] The flexibility of California's rules allowed for the introduction of leasing models, which likely helped to expand the market.

Trade policy is also indirectly related to the subject of this article. In 2014, after the period studied, tariffs of at least 30 percent were imposed by the US Department of Commerce on Chinese and Taiwanese solar panels. A full analysis of the merits and fairness of these sanctions is beyond the scope of this article. However, this article clearly shows how competition from Chinese manufacturers helped drive down overall system costs and spurred increased installations. Subsidizing solar systems while at the same time imposing tariffs on imported panels seem like contradictory actions if the aim is to increase renewable energy production.

This article only looks at a few of many interesting research questions related to the economics of the rapid expansion of distributed energy generation in the United States. The externality cost or benefit of distributed energy on the electricity grid and environment is an important practical concern for regulators seeking to design appropriate long-term incentive structures for investment. This article discusses and analyses the impact of subsidies on investment, but makes no attempt to say whether and at what level the subsidies are appropriate. This is a research question with clear relevance. The discussion of possible issues of

[11] http://www.greentechmedia.com/articles/read/north-carolina-bill-would-launch-opportunity-for-third-

asymmetric information in affecting both pricing and industry structure also deserves more formal treatment than has been provided here. Time series data of production from a selection of solar panel installations is available, and could potentially be used to test whether quality differences can be observed between solar panels that are leased and sold outright.

8 Software

For data cleaning and manipulation, I used the python package Pandas [McKinney, 2012]. I use the R statistical programming package for the analysis in this article [R Core Team, 2013]. I use the R package ggplot2 for plotting [Wickham, 2009], texreg for table formatting [Leifeld, 2013] and mgcv for implementation of the Generalized Additive Models [Wood, 2011].

References

George A. Akerlof. The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics*, 84(3):488–500, August 1970. doi: 10.2307/1879431. URL `http://www.jstor.org/stable/1879431`.

Erin Baker, Meredith Fowlie, Derek Lemoine, and Stanley S. Reynolds. The Economics of Solar Electricity. *Annual Review of Resource Economics*, 5(1):387–426, June 2013. ISSN 1941-1340, 1941-1359. doi: 10.1146/annurev-resource-091912-151843. URL `http://www.annualreviews.org/eprint/xRx96NggYH7YgQTkHBEU/full/10.1146/annurev-resource-091912-151843`.

Bryan Bollinger and Kenneth Gillingham. Peer Effects in the Diffusion of Solar Photovoltaic Panels. *Marketing Science*, 31(6):900–912, September 2012. ISSN 0732-2399. doi: 10.1287/mksc.1120.0727. URL `http://pubsonline.informs.org/doi/abs/10.1287/mksc.1120.0727`.

Severin Borenstein. The Market Value and Cost of Solar Photovoltaic Electricity Production. *Center for the Study of Energy Markets*, January 2008. URL `http://escholarship.org/uc/item/3ws6r3j4`.

Chrystie Burr. Subsidies and Investments in the Solar Power Market. *Working Paper*, March 2014. URL `http://burr.faculty.arizona.edu/sites/burr.faculty.arizona.edu/files/burr_solar_2014r.pdf`.

California Public Utilities Commission. *California Solar Initiative Program Handbook May 2014*. May 2014. URL http://www.gosolarcalifornia.ca.gov/documents/CSI_HANDBOOK.PDF.

Chiara Candelise, Mark Winskel, and Robert J. K. Gross. The dynamics of solar PV costs and prices as a challenge for technology forecasting. *Renewable and Sustainable Energy Reviews*, 26:96–107, October 2013. ISSN 1364-0321. doi: 10.1016/j.rser.2013.05.012. URL http://www.sciencedirect.com/science/article/pii/S1364032113003158.

Samuel R. Dastrup, Joshua Graff Zivin, Dora L. Costa, and Matthew E. Kahn. Understanding the Solar Home price premium: Electricity generation and Green social status. *European Economic Review*, 56(5):961–973, July 2012. ISSN 0014-2921. doi: 10.1016/j.euroecorev.2012.02.006. URL http://www.sciencedirect.com/science/article/pii/S0014292112000244.

Dimitiri Dimitropoulos and Adonis Yatchew. Discerning Trends in Commodity Markets. *Macroeconomic Dynamics*, Forthcoming, 2016.

Easan Drury, Mackay Miller, Charles M. Macal, Diane J. Graziano, Donna Heimiller, Jonathan Ozik, and Thomas D. Perry IV. The transformation of southern California's residential photovoltaics market through third-party ownership. *Energy Policy*, 42:681–690, March 2012. ISSN 0301-4215. doi: 10.1016/j.enpol.2011.12.047. URL http://www.sciencedirect.com/science/article/pii/S0301421511010536.

Martin Feldstein and Bernard Friedman. Tax subsidies, the rational demand for insurance and the health care crisis. *Journal of Public Economics*, 7(2):155–178, April 1977. ISSN 0047-2727. doi: 10.1016/0047-2727(77)90044-5. URL http://www.sciencedirect.com/science/article/pii/0047272777900445.

T. J. Hastie and R. J. Tibshirani. *Generalized Additive Models (Chapman & Hall/CRC Monographs on Statistics & Applied Probability)*. Chapman and Hall/CRC, 1 edition edition, June 1990. ISBN 0412343908.

Jonathan E. Hughes and Molly Podolefsky. Getting Green with Solar Subsidies: Evidence from the California Solar Initiative. *Journal of the Association of Environmental and Resource Economists*, 2(2):235–275, June 2015. ISSN 2333-5955. doi: 10.1086/681131. URL http://www.journals.uchicago.edu/doi/abs/10.1086/681131.

Philip Leifeld. texreg: Conversion of Statistical Model Output in R to LATEX and HTML Tables. *Journal of Statistical Software*, 55(8), November 2013.

Wes McKinney. *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. O'Reilly Media, 1 edition edition, October 2012.

Gregory F. Nemet. Beyond the learning curve: factors influencing cost reductions in photovoltaics. *Energy Policy*, 34(17):3218–3232, November 2006. ISSN 0301-4215. doi: 10.1016/j.enpol.2005.06.020. URL http://www.sciencedirect.com/science/article/pii/S0301421505001795.

John Pucher, Anders Markstedt, and Ira Hirschman. Impacts of Subsidies on the Costs of Urban Public Transport. *Journal of Transport Economics and Policy*, 17(2):155–176, 1983. ISSN 0022-5258. URL `http://www.jstor.org/stable/20052679`.

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2013. URL `http://www.R-project.org/`.

Hadley Wickham. *Ggplot2: Elegant Graphics for Data Analysis*. Springer Publishing Company, Incorporated, 2nd edition, 2009. ISBN 0387981403, 9780387981406.

Ryan Wiser, Mark Bolinger, Peter Cappers, and Robert Margolis. Letting The Sun Shine On Solar Costs: An Empirical Investigation Of Photovoltaic Cost Trends In California. *Lawrence Berkeley National Laboratory*, April 2006. URL `http://escholarship.org/uc/item/22p2m6vr`.

Simon Wood. *Generalized Additive Models: An Introduction with R (Chapman & Hall/CRC Texts in Statistical Science)*. Chapman and Hall/CRC, 1 edition edition, February 2006. ISBN 1584884746.

Simon N. Wood. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(1):3–36, 2011. ISSN 1467-9868. doi: 10.1111/j.1467-9868.2010.00749.x. URL `http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9868.2010.00749.x/abstract`.

Adonis Yatchew. Nonparametric Regression Techniques in Economics. *Journal of Economic Literature*, 36(2):669–721, 1998. ISSN 0022-0515. URL `http://www.jstor.org/stable/2565120`.

A Coefficient estimates on county fixed effects

	County & Sector FE	Lease & Sector RE
Amador	-0.079^{***}	-0.097^{***}
Butte	-0.118^{***}	-0.134^{***}
Calaveras	-0.097^{***}	-0.100^{***}
Colusa	-0.093^{**}	-0.108^{**}
Contra Costa	0.005	0.005
El Dorado	-0.045^{***}	-0.054^{***}
Fresno	-0.076^{***}	-0.066^{***}
Glenn	-0.083^{***}	-0.103^{***}
Humboldt	0.000	-0.000
Imperial	-0.285^{*}	-0.249
Inyo	-0.108^{*}	-0.131^{**}
Kern	-0.051^{***}	-0.040^{***}
Kings	-0.101^{***}	-0.114^{***}
Lake	-0.026	-0.048^{**}
Lassen	-0.210^{*}	-0.209^{*}
Los Angeles	-0.011^{*}	0.067^{***}
Madera	-0.082^{***}	-0.087^{***}
Marin	-0.072^{***}	-0.083^{***}
Mariposa	-0.111^{**}	-0.122^{***}
Mendocino	-0.015	-0.016
Merced	-0.109^{***}	-0.125^{***}
Mono	-0.125^{***}	-0.106^{***}
Monterey	-0.049^{***}	-0.059^{***}
Napa	-0.018	-0.026^{**}
Nevada	-0.112^{***}	-0.120^{***}
Orange	0.016^{***}	0.061^{***}
Placer	-0.059^{***}	-0.058^{***}
Plumas	-0.087^{***}	-0.090^{***}
Riverside	-0.019^{***}	0.055^{***}
Sacramento	-0.064	-0.073
San Benito	-0.075^{**}	-0.099^{***}
San Bernardino	0.006	0.052^{***}
San Diego	-0.021^{***}	0.034^{***}
San Francisco	0.149^{***}	0.155^{***}
San Joaquin	-0.077^{***}	-0.087^{***}
San Luis Obispo	-0.054^{***}	-0.067^{***}
San Mateo	0.003	-0.015^{*}
Santa Barbara	-0.090^{***}	-0.082^{***}
Santa Clara	-0.062^{***}	-0.045^{***}
Santa Cruz	-0.090^{***}	-0.098^{***}
Shasta	-0.087^{***}	-0.097^{***}
Solano	0.007	-0.007
Sonoma	-0.067^{***}	-0.051^{***}
Stanislaus	-0.119^{***}	-0.123^{***}
Sutter	-0.082^{***}	-0.095^{***}
Tehama	-0.047^{*}	-0.070^{**}
Trinity	0.159	0.186
Tulare	-0.067^{***}	-0.068^{***}
Tuolumne	-0.070^{***}	-0.077^{***}
Ventura	-0.038^{***}	-0.031^{***}
Yolo	-0.014	-0.024^{**}
Yuba	0.013	-0.001

Table 3: County Coefficients

B Differencing model results

As a robustness check, I ran a simple differencing model, which is also able to take into account the unobserved variables that vary non-linearly over time. Instead of estimating an explicit smoothed curve over time, this model takes the first-difference of the left-hand-side variable and continuous right-hand-side variables ordered by date of installation. At the limit, this will eliminate the unobserved time-varying components.

Following the notation of Yatchew [1998], consider the partial linear model written as in equation 3.

$$y_i = \mathbf{Z_i}\beta + f(x_t) + e_i \tag{3}$$

Here y_i represents the dependent variable of log cost per-kW of solar panel systems. $\mathbf{Z_i}$ represents a vector of continuous explanatory variables. For simplicity I have excluded categorical and binary variables. β represents the vector of coefficients on the variables. $f(x_i)$ represents the function of unobserved time-varying variables, x_i.

When variables are ordered by time then a differencing can be written as in equation 4

$$y_{i,t} - y_{i-1} = (\mathbf{Z_i} - \mathbf{Z_{i-1}})\beta + f(x_i) - f(x_{i-1}) + e_i - e_{i-1} \tag{4}$$

As the time difference approaches zero between observations, $f(x_i - x_{i-1})$ also approaches zero and is removed as a confounding factor from the simple OLS estimation. For a more in-depth discussion of differencing, I again refer to Yatchew [1998].

The results of the differencing model are shown in table 4. The sign and approximate magnitudes of the coefficient estimates are in line with the estimates from the cubic regression spline estimates above.

	Differencing Model
(Intercept)	0.0000
	(0.0008)
nameplate	-0.0002^{***}
	(0.0000)
county_year_total_mw	0.0009^{***}
	(0.0001)
zip_year_total_mw	-0.0151^{***}
	(0.0020)
incentive_per_kw	0.0001^{***}
	(0.0000)
contractor_market_share_perc	0.0065^{***}
	(0.0002)
AIC	27323.6139
BIC	27390.6479
Log Likelihood	-13654.8070
Deviance	8060.5552
Num. obs.	106541

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 4: Regression results from a differencing model. The results are roughly in line with results from the earlier estimation. Binary and categorical variables have been excluded from this regression, so some deviations are to be expected.