Sources of price dispersion in U.S. residential solar installations

Gregory Nemet[1], Eric O'Shaughnessy[1,2], Ryan H. Wiser[3], Naïm R. Darghouth[3], Galen L. Barbose[3], Ken Gillingham[4], Varun Rai[5]

1 University of Wisconsin at Madison; 2 National Renewable Energy Laboratory; 3 Lawrence Berkeley National Laboratory; 4 Yale University; 5 University of Texas at Austin

SunShot U.S. Department of Energy

June 2017

ABSTRACT

Prices of solar PV have dropped dramatically, by half in just the past 6 years. But looking simply at prices paid today, there is considerable heterogeneity. For systems installed in 2014, the 10-90 percentile range for the observed $/W spans nearly a factor of 2. This apparent price dispersion raises policy-relevant questions, such as: why are consumers paying more than they need to? And would better-informed consumers increase the social benefits of solar PV? This paper analyzes price dispersion in U.S. residential PV installations between 2008 and 2014. Focusing on the most commonly used metric in previous studies of price dispersion, we use the quarterly coefficient of variation (CV) as our measure of price dispersion. We find higher levels of price dispersion in our data (0.22) than the average of 55 previous studies we reviewed (0.16). We also find that price dispersion has been persistent; it has remained above 0.15 since 2000 with no trend over that period. If anything, price dispersion has been increasing recently during the period for which we have complete data, 2008-14. Econometric analysis of the factors affecting price dispersion supports theories from the economic literature focusing on access to information and the costs and benefits of consumer search. Factors that increase the consumer payoffs of investing time in searching for information—system size and the value of solar—are associated with lower levels of price dispersion. Factors that reduce the costs of search—neighbors who have recently installed solar and having third-party quotes available—are also associated with less price dispersion. These results provide support for the importance of public efforts to enhance access to price information, e.g. by supporting private sector price quote providers. The results also point to the particular need for information in nascent markets for PV in which access to the experience of neighbors is not available.

Keywords: price dispersion; solar; PV

Sources of price dispersion in U.S. residential solar installations

Gregory Nemet[1], Eric O'Shaughnessy[1,2], Ryan H. Wiser[3], Naïm R. Darghouth[3], Galen L. Barbose[3], Kenneth Gillingham[4], Varun Rai[5]

[1]University of Wisconsin at Madison; [2]National Renewable Energy Laboratory; [3]Lawrence Berkeley National Laboratory; [4]Yale University; [5]Universityof Texas at Austin

Contents

1. INTRODUCTION

The halving of the price of residential solar photovoltaic (PV) systems since 2009 has been a major driver in the six-fold increase in grid-connected residential and non-residential PV installed capacity in the United States (U.S.) (Barbose and Darghouth 2016). In 2016, PV accounted for nearly 39% of *new* electricity generation in the U.S. Moreover, the cost of PV is rapidly approaching the U.S. Department of Energy's SunShot 2020 cost goals for utility, commercial, and residential applications (Woodhouse, Jones-Albertus et al. 2016). However, these impressive changes obscure the complication that not everyone receives low-priced solar. People pay dramatically different prices for what, on its face, appears to be a homogenous good, i.e., devices for producing kilowatt hours of solar electricity. The observed price per watt paid for installed residential PV systems in 2014 spans a factor of four. *Why is there such a range of prices for a seemingly indistinguishable good? Why are some people paying four times as much as others? Are these prices excluding some people from the market, and consequently limiting the benefits of solar PV? Is there a role for public policy to remedy this situation?* To address these questions, in this paper, we begin by examining the large set of studies in industrial organization that focus on the levels and sources of price dispersion in other sectors in section 2. In section 3, we outline our approach to measuring and analyzing price dispersion in U.S. residential PV, making use of a variable-rich PV system price data set. In section 4, we provide a broad set of descriptive statistics establishing the levels and trends in price dispersion. Section 5 includes results of our econometric analysis of the factors affecting price dispersion in PV. Section 6 concludes with a discussion of the implications of our findings.

2. REASONS FOR PRICE DISPERSION

Economic theory suggests that a homogenous good will sell for the same price in all locations, a concept commonly referred to as *the law of one price*. Price dispersion refers to a violation of the law of one price where consumers pay different prices for the same good or service. To understand price dispersion in PV, we start with the half a century of studies in industrial organization about why the law of one price does not always hold. From the first studies, costly access to information about attributes of a good has consistently provided an explanation of observed price dispersion (Stigler 1961, Salop and Stiglitz 1977, Reinganum 1979, Varian 1980) – a link that we further explore below. To be sure, studies have shown other explanations play a role as well, including: differences in production costs, distribution costs, price elasticity of demand, price discrimination, market structure, regulations, currency fluctuations, inflation, purchase-related amenities, and distinctions among apparently homogenous goods. Some of these factors can result in *product differentiation*, such that observed price heterogeneity for heterogeneous goods may be spuriously attributed to price dispersion. To determine if price dispersion actually exists, we need to control for product-differentiating factors, especially quality, market segments, and input costs. We discuss and control for product-differentiating factors that play a role in this case study. However, the center of this investigation is assessing the notion of heterogeneous access to information: that price dispersion can persist if information remains costly and some consumers choose not to become informed (Pratt, Wise et al. 1979, Burdett and Judd 1983). We focus on the consumers' incentives to search for information.

Search costs are particularly important in PV because consumers must choose a firm to install their system and may even have some choice about what type of system to install. Buying decisions are complicated for several reasons: the technology is new and dynamic; historically, almost no residential consumer is used to making capital investments to procure electricity; the value of PV to the consumer is associated with complicated rate designs and changing policies (Blackburn, Magee et al. 2014); and PV systems typically last 20 years or more, precluding repeat purchases. Furthermore, because the technology is new, still rare, and infrequently purchased, consumers likely have few acquaintances from which to obtain trustworthy experiential information (Rai, Reeves et al. 2016) (Rai et al. 2016). From a policy perspective, understanding the reasons why prices vary is important for improving the effectiveness and cost-efficiency of policies promoting the adoption of PV. We review the explanations of price dispersion to develop a model of PV prices, which we then empirically test.

2.1. Heterogeneous producers, goods, and consumers

Three categories of explanation are straightforward. Producers have uneven costs of production, the goods they produce are imperfect substitutes, and consumers value attributes differently.

2.1.1. Production costs and competition

Because positive consumer search costs can allow producers to act as local monopolists (Diamond 1971), pricing can diverge from a competitive equilibrium and thus differences in producer costs can create price dispersion (Carlson and McAfee 1983, Perloff and Salop 1985, Spulber 1995, Walsh and Whelan 1999). *Heterogeneous producer costs contribute to price dispersion* in studies of agriculture products (Kano, Kano et al. 2013), automobiles (Goldberg and Verboven 2005), electricity (Davis, Grim et al. 2013), and U.S. domestic imports more generally (Yilmazkuday 2014). Not only manufacturing costs but differences in other costs such as wages (Van Nieuwerburgh and Weill 2010), exchange rates (Goldberg and Verboven 2001), inflation (Van Hoomissen 1988), taxation (Chouinard and Perloff 2007), and transportation and distribution affect prices. Of particular relevance to this study, differences in advertising costs have also been found to affect prices (Baye, Morgan et al. 2006). A study of homogenous manufactured goods, including concrete, cardboard boxes, and steel cans, found that firm size contributed to much of the observed costs differences across firms within products (Roberts and Supina 1996).

These differences in costs often become especially relevant in a setting of imperfect competition. Somewhat paradoxically, several studies have found that price dispersion increases in the number of firms in the market (Carlson and McAfee 1983) due to: each firm having a lower likelihood of being the lowest price firm (Stahl 1989), enhanced incentives to set extreme prices (Chandra and Tappata 2011), and small effects of competition on the highest prices (Allen, Clark et al. 2014). Others have found the opposite relationship, with dispersion decreasing in competition (Spulber 1995, Barron, Taylor et al. 2004, Lin and Chen 2014). Furthermore, price dispersion may increase with concentration in competitive markets and decrease with concentration in less competitive markets (Dai, Liu et al. 2014). The relationship appears positive between market concentration and price dispersion when that concentration is high or low and negative when observed at medium values (Chakrabarty and Kutlu 2014). Price dispersion may increase incentives for entry (Gerardi and Shapiro 2009), stimulate cat-and-mouse pricing games as large sellers try to price small sellers out of the market (Menzio and Trachter 2015), and may also reflect predatory pricing among incumbents (Besanko, Doraszelski et al. 2014).

Many firm costs are decreasing in firm scale, and previous work has indicated large differences in firm scale in the case of PV. Similarly, firm experience also affects prices (Gillingham, Deng et al. 2016). Scale and experience are thus important factors influencing price dispersion. Note though that sometimes costs can increase in scale if firms shift to higher quality inputs (Atkin, Chaudhry et al. 2015). We see some indications of this effect in terms of the efficiency of PV panels that firms use. Quality is thus also important to control for.

2.1.2. Product quality

One reason why information, and the search for it, is so crucial to understanding price dispersion is that there may be quality differences among the products being considered. Apparent price dispersion can be, at least in part, due to real heterogeneity in the characteristics among products that are substitutes for one another (Shepard 1991, Brynjolfsson and Smith 2000, Goldberg and Verboven 2005, Imbs, Mumtaz et al. 2010). Firms may even intentionally differentiate their products, not necessarily to make them better, but to introduce price dispersion (Shepard 1991, Clay, Krishnan et al. 2002). Consumers may use higher prices as an indication of quality (Brucks, Zeithaml et al. 2000). This literature underscores the importance of accounting for quality differences that are observable to all consumers, but which study data may not include—including quality differences that are difficult for consumers to understand before, or even after, purchase (experience or post-experience goods).

Just as for other products, for solar PV it is also possible that an apparently homogenous good—installed capacity in kW— may actually be a heterogeneous one. Hence what seems to be price dispersion may in fact partially be a result of product heterogeneity. The consumer benefit from installed PV depends not just on the capacity (kW) of the system, but also how that system performs. Two systems with identical installed costs and capacity may produce different quantities of electricity based on installation factors such as equipment choice, system tilt, and site selection. Conceptually, a system with a higher kWh to kW ratio (more output per unit of installed capacity) has a greater kWh return on investment than a system with a lower kWh to kW ratio. Therefore, differences in installed system characteristics can help capture differences in the final consumer good involved, kWh of electricity. Installed systems are comprised of both a durable good (panels with 20-year warranty) and a service (installation); from our data, we can estimate that the latter constitute well over half of system costs. Certain installed system characteristics may be valued, e.g. reputation, capital cost over-runs, and help in accessing incentives. This raises the question of whether there is substantial variation in the quality of installed systems.

2.1.3. Consumer attributes

Consumer preferences also contribute to price dispersion. This might be due to differences in price elasticity of demand in certain markets (Goldberg and Verboven 2005) or consumer preferences for certain brands (Imbs, Mumtaz et al. 2010). With frequently purchased, homogeneous retail food goods, shopping preferences (‘supermarket lovers’ vs. ‘social shoppers’) play a large role in the existence of monopolistic retailers and small retailers both staying in business (Anania and Nistico 2014). Studies have found that dispersion increases in income and falls in educational attainment (Marvel 1976). Another finding suggests price dispersion negatively influences trust due to perceived risk (Wu, Vassileva et al. 2015).

In the case of PV, we have seen that, as predicted generally (Marvel 1976), prices do appear to reflect consumer incomes (Gillingham, Deng et al. 2016). More centrally to this topic, the one-time purchase, high-cost product aspects may make value sensitive to a consumer's appetite for risk; risk averse consumers may pay more for systems from installers they consider reliable.

2.2. Costs and benefits of search

From the perspective of the consumer, one can consider price dispersion resulting from heterogeneous consumers' assessments of the benefits and costs of investing in the search for information about a potential purchase. An important outcome is that some consumers will choose to stay uninformed.

2.2.1. Costly consumer search

Consumers face different costs in accessing information about product prices and attributes. Previous work has attributed price dispersion to differences in search costs in automobiles (Dahlby and West 1986), gasoline (Chandra and Tappata 2011), and prescription medications (Sorensen 2000). The past two decades have seen an increasing body of work looking at the effect of the internet in reducing the costs of accessing product information (Bakos 1997, Bailey 1998, Pan, Ratchford et al. 2002). And while these studies find much lower search costs, branding, awareness, and trust remain important to purchase decisions (Brynjolfsson and Smith 2000). Kutlu (2015) found that consumers – lacking perfect memory and price information – employ heuristics showing whether a product is expensive or not to help process price information. This raises another channel of information, trusted peers, who can provide information based on experience with trust and at low costs, typically as a co-benefit of social interactions (Bollinger and Gillingham 2012, Rai and Robinson 2013). Assessing this mechanism is central to our study.

2.2.2. Gains to consumer search

Consumers must also decide if investing in search is likely to generate a payoff. First, if price dispersion is perceived as large, the gains to information are also large (Pratt, Wise et al. 1979, Chandra and Tappata 2011, Jaeger and Storchmann 2011). There is a 'value of information' corresponding to a consumer's expected benefit of being informed (Pennerstorfer, Schmidt-Dengler et al. 2014). This value is the difference between the expected price and the lowest price in the market or the reduction in price due to search (Marvel 1976). This dynamic is one reason we use several years of data, under the assumption that opportunities for search from the existence of dispersion will have stabilized. Second, the gains to costly information search increase in the price of the good (Stigler 1961, Lach 2002, Eckard 2004, Jaeger and Storchmann 2011). The underlying assumption behind this claim is that search costs increase only modestly for more expensive goods; thus, there are scale economies in search. Third, the frequency of purchase affects search motivation, with consistent findings of lower price dispersion for frequently purchased goods (Nelson 1970, Sorensen 2000). This 'learning by buying' mechanism involves a combination of lower search costs (via information acquired through experience) and higher payoffs (multiple purchases). In the case of durable goods, such as PV, repeated purchases are rare and this motivation for search plays little role if any. Although increased search resulted in lower mortgage rates in one study (Lee 2014), a repeated finding is that despite the gains to search, many consumers still do not invest in accessing information (Ratchford 2009).

2.3. Measuring price dispersion

While previous studies have measured price dispersion in a variety of ways, the most predominant has been coefficient of variation (CV=variance/mean) (Eckard 2004, Allen, Clark et al. 2014). Others have used variance (Pratt, Wise et al. 1979, Dahlby and West 1986), standard deviation (Van Hoomissen 1988), range (Brynjolfsson and Smith 2000, Sorensen 2000), Gini coefficient (Borenstein and Rose 1994), 10–90 percentile range (Roberts and Supina 1996), and the difference between the two lowest prices in a market (Baye, Morgan et al. 2006). Other approaches control for explanatory factors and use the residuals to measure dispersion (Barron, Taylor et al. 2004, Jaeger and Storchmann 2011). Chandra and Tappata (2011) assess temporal price dispersion by measuring 'rank reversals', changes in a firm's price rank over time. We adopt CV as the primary measure in this study, in part due to its frequent use and suitability over long time periods (Baye, Morgan et al. 2006).

In sum, the existing literature on price dispersion indicates that if searching for product information is costly, and some consumers choose to remain uninformed, then we should expect dispersion in prices. We thus set out to examine whether these dynamics are at work in the U.S. PV market.

3. APPROACH, METHODS, AND DATA

We assemble a data set including a large sample of residential PV installations in the past 15 years to address the research question: *what causes the apparently large range in prices for installed small-scale PV systems in the residential sector?*

3.1. The data

As our primary data source we use installed PV system data from 59 PV incentive programs in 34 U.S. states, collected as part of the Lawrence Berkeley National Laboratory's (LBNL's) *Tracking the Sun* (TTS) report series (Barbose and Darghouth 2016). The full TTS data set accounts for about two thirds of U.S. PV installations since 2000 and is described in detail in the annual TTS report (Barbose and Darghouth 2016). We have data on systems installed between 2000 and 2014. We restrict our core analysis to 2008-14 to focus on more-recent data, and to avoid data quality problems in 2007. However, we include data back to 2000 in the descriptive analyses for which installer information is not used. In addition, we restrict system size to 1-15kW and price per watt to 1-25$/W. We drop the following types of installations: commercial, other types of non-residential customers (e.g. schools), appraised value third-party owned (TPO) systems, and those missing identification of the installer. We also drop the small portions of the data that represent building-integrated PV, systems with battery backup and/or tracking, self-installs, and systems placed in new construction. In the end, we have complete data on 234,666 installations from 2008-14.

To provide an overview of the dynamics behind our investigation of price dispersion, Figure 1 shows the distribution of installed price per watt for the entire 2000-14 data set. Focusing on the most recent year, 2014, Figure 2 shows the nature of dispersion, by illustrating prices within the largest markets (counties).

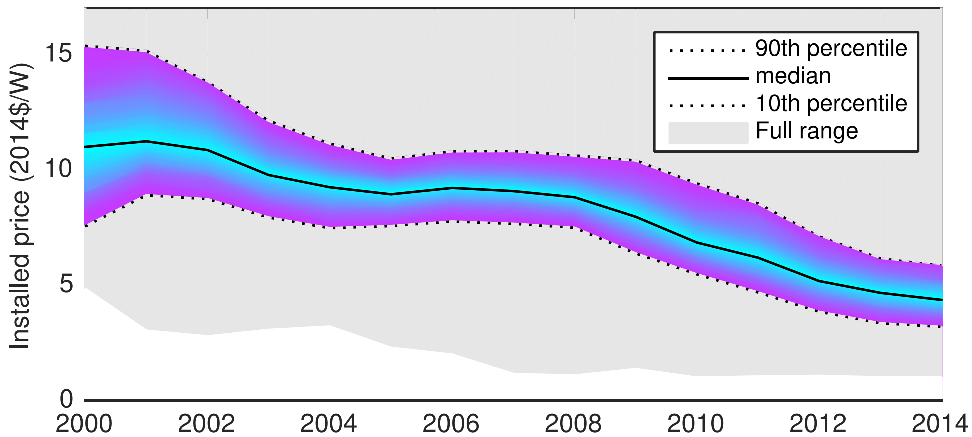

Figure 1. Distribution of installed prices (2014 $/W).

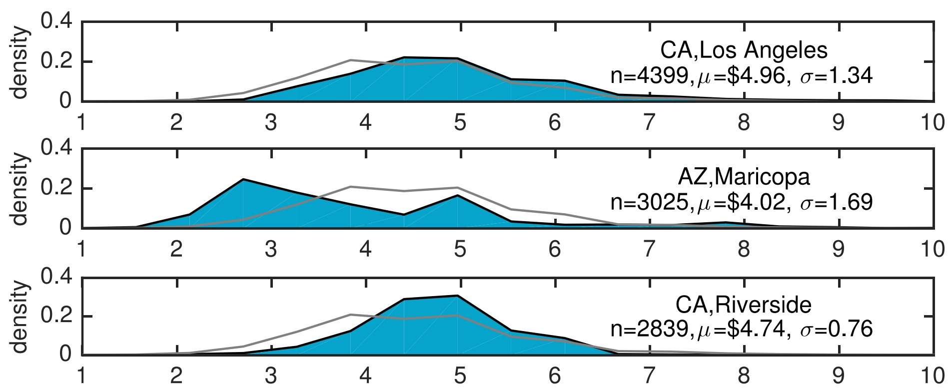

Figure 2. Distributions of prices (2014$/W) in largest markets (defined by counties), by installs in 2014. Gray curve is distribution for all U.S. installations.

3.2. Dependent variable: CV

We measure price dispersion in a given time and place using the coefficient of variation (CV). There are several ways to define CV with our data. For example, temporally, one can define CV over months or years. For our analysis we use quarters, an intermediate measure. Similarly, geographically, one could calculate CV over counties, states, or for the U.S. as a whole. We use counties as the relevant market from which consumers make choices. *In our analysis, we calculate CV by quarter within a county.* In Figure 3, we compare the CVs in our study—that is the average CV for each quarter and country—to 55 CVs in other studies of price dispersion for various products. While we would not necessarily expect the CVs to be the same, it is nonetheless instructive to see how the CVs for solar PV compare with others. Both the 2014 and 2008-14 measures of CV in our data set are above the mean of the other studies (0.16).

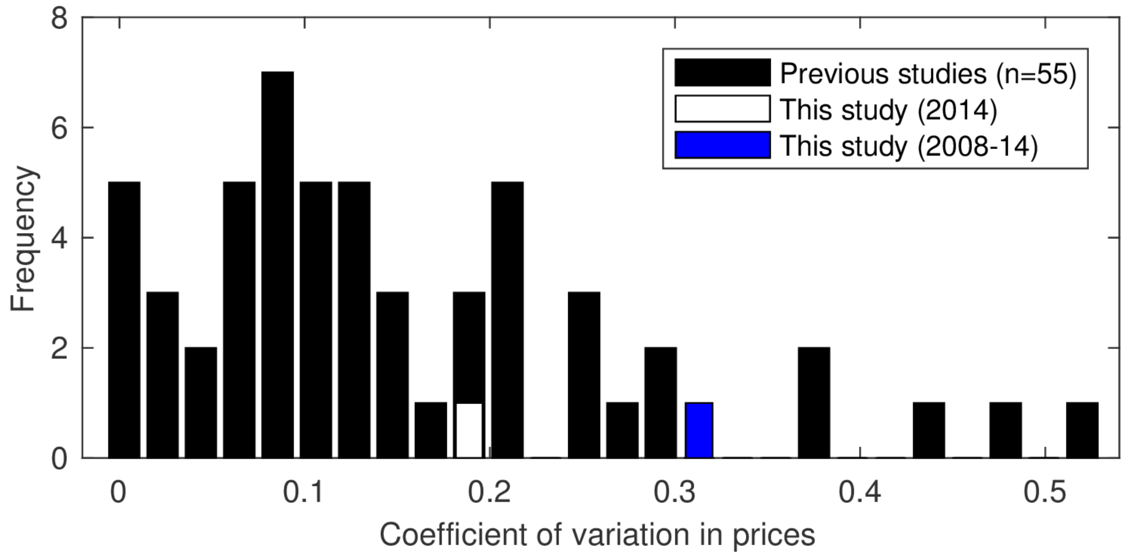

Figure 3. Comparison of coefficients of variation: values from the literature and values from PV data.

3.3. Independent variables

We use other variables in the TTS data set to estimate predictors of price dispersion. We build several variables using the name of the company that installed each system, which we refer to hence as installers. To account for the level of installer competition we calculate a variable for the concentration of market share by installer using a Hirschman-Herfindahl Index (HHI). We calculate measures of experience for each installer, i.e. the count of their previous installations. We use an index of labor costs in each county, as a proxy for firms' costs in that county and time. We use demographics for each county and time: educational attainment and household income. We use the density of households in a county (households/km^2) as a possible proxy for information distribution. We account for whether systems are third-party owned (TPO) or customer owned (CO). We include characteristics of the system (size in kW) and modules and inverter (efficiency, country of origin, thin-film, and micro-inverter). We use electricity prices, insolation, and all subsidies to estimate the consumer value of solar in each county and quarter. For details on how each of these variables was constructed see Nemet, O'Shaughnessy et al. (2016).

We also construct new variables to model access to information from neighbors with solar, access to quotes from third-party quote providers, and a variable to estimate system performance. First, to examine access to information from neighbors about solar, we use the addresses of installations to construct variables for the count of nearby neighbors who installed PV previously. As a base assumption, we use a 1 km radius to define a neighbor and apply 12 months to account for the installations being recent enough to provide relevant information for new purchases (Graziano and Gillingham 2015). We create similar variables for other radii values and time periods. An important rationale for using addresses is the clustering we see throughout the data (Figure 4). Second, we use datasets of quotes provided by third-party quote providers, companies that provide a platform for prospective customers to solicit quotes from installers, to assess whether installations had quotes available, and how many were provided (see Appendix for detail). Third, we attempt to account for the quality or system performance of the PV system. This is done in part simply by including the efficiency and country of origin of modules. To go further we use data on the performance of a small subset of systems, for which we have electricity output over multiple years. In the Appendix, we describe how we arrive at a performance ratio for each system in the third and fourth year that system was online. We create a dummy variable for installers who have a mean performance ratio above average and consider these installers high-quality installers.

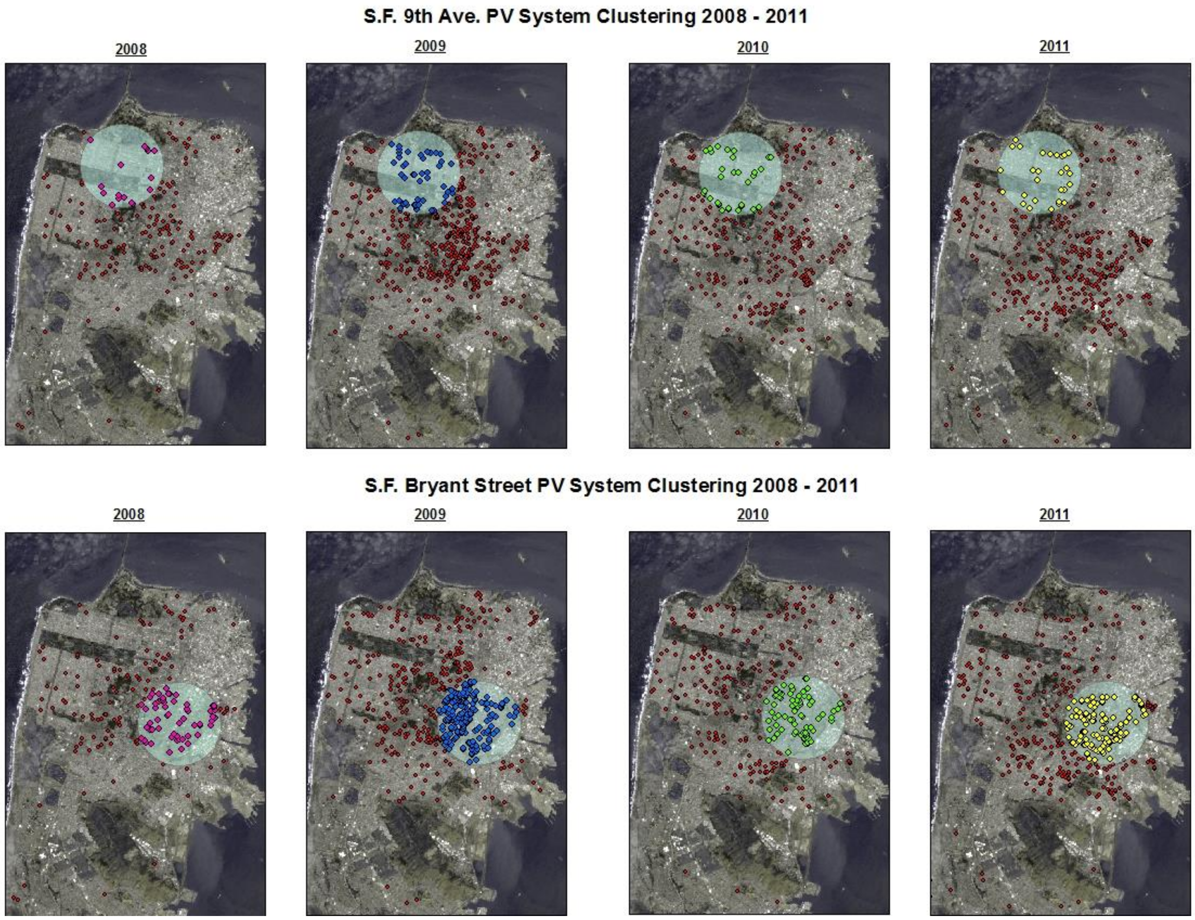

Figure 4. Annual new installations in two San Francisco Neighborhoods 2008 – 2011.

Finally, we use counties to define markets. This approach is common in previous studies (Gillingham, Deng et al. 2016), but we recognize that other approaches to defining markets may also be useful to explore. Counties are political jurisdictions that may represent somewhat arbitrary definitions for the space over which potential PV customers search for information about prices. Thus as a robustness check we draw on recent work using an alternative scheme for defining markets, based on the presence of installers rather than political jurisdictions (O'Shaughnessy, Nemet et al. 2016). The Appendix provides a summary of this approach. Figure 5 shows how the approaches differ using the example of the Phoenix metropolitan area. The variable most sensitive to the choice of market definition, installer HHI, has a correlation of 0.55 between the county definitions and the new installer-driven definitions.

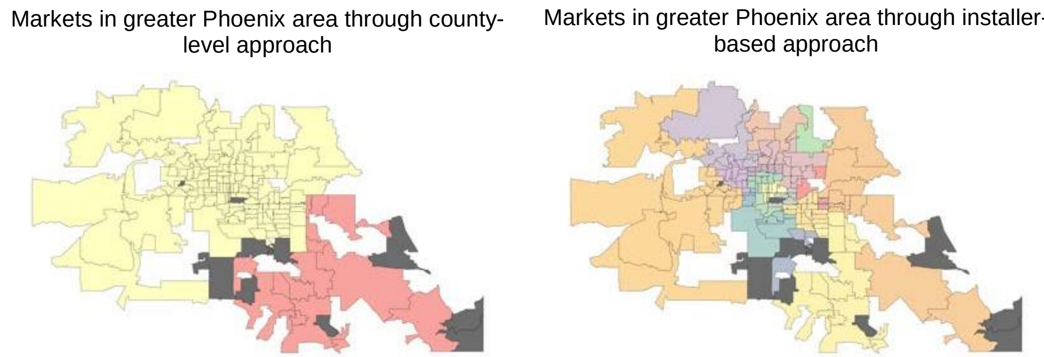

Figure 5. Illustration of differences in solar PV markets through a county-level approach (left map) and installer-based approach (right map) in the greater Phoenix area. Each color represents a separate market. Gray boundaries correspond to zip codes.

4. DESCRIPTIVE RESULTS

4.1. Differences in CV by groups

As a first look at the relationship between a subset of the variables that the literature establishes as predictors of price dispersion and CV in our data, we show mean prices and CV for several groups of the data set (Table 1).

Theories of consumer search predict that higher-value goods will show less price dispersion because the benefits for consumers of investing in search are higher for those than for goods with lower value. We use system size as a proxy for the size of the consumer's investment. The results for system size (watts) support this prediction; CV, which is calculated from unit prices ($/W), is lower for the largest category of systems, those >7kW, and the middle category CV is lower than that for the smallest systems. Similarly, in 2014 we see slightly lower variability in prices for markets where the value of solar is above the average price of an installed system in 2014 ($4.56/W), although we see the opposite relationship for the longer 2008-14 period. We also see much less price dispersion in TPO systems, where a customer hosts and buys power from a system owned by a third-party company, compared to customer owned (CO) systems. Third-party system owners generally purchase and own multiple systems. Lower price dispersion in TPO systems supports the notion that companies that pay for many systems will see higher gains from search than homeowners who purchase only one. Third-party system owners face similar information challenges around quality and costs, but face higher payoffs from search and also face lower costs from search in that they can use information from previous transactions to inform current ones. TPO providers also sometimes use standard pricing that they offer to installers; their search in this case thus consists of determining what price to offer installers.

The price dispersion literature also predicts that price dispersion will be lower when the costs of search are lower, such as due to easier information acquisition from neighbors. The neighbors variable shows some of this effect, with lower dispersion in areas with at least one install within 1km in the previous 12 months; we do not see as strong an effect at higher levels of neighbors. The availability of quotes from third party providers should also reduce price dispersion. We see this as well, with markets that had at least 1 quote per install showing lower CV than those with less than one quote per install. Of course, a multivariate approach is needed to identify these effects, as well as the impacts of other variables.

The literature is ambiguous about the effect of market concentration. For PV in 2014, we observe more price dispersion in less concentrated markets. However, this is not the case for the earlier years, perhaps underscoring the complex effects of market concentration. Because California accounts for more than half of our data set, we assess whether it is structurally different from the rest of the US. While prices are higher in California, CV is the same.

Table 1. Descriptive statistics of PV prices ($/W) grouped by predictors of price dispersion.

	2008–14			2014		
	mean	cv	n	mean	cv	n
Benefits of search						
Size 1-5 kW	6.93	0.30	92012	5.01	0.18	12576
5-7 kW	5.82	0.26	68760	4.56	0.16	14548
7-15 kW	5.46	0.26	73894	4.24	0.15	17280
VoS <4.56	4.63	0.22	53581	4.47	0.18	28503
≥ 4.56	6.59	0.29	181085	4.73	0.17	15901
CO	6.60	0.31	123880	4.54	0.21	19907
TPO	5.41	0.20	87229	4.62	0.14	21908
Costs of search						
Nbrs. 0	6.33	0.31	61506	4.56	0.18	12592
1-5	6.52	0.26	64237	4.73	0.15	9007
>5	6.27	0.24	28670	4.72	0.15	3049
Quotes <1	6.60	0.31	167220	4.53	0.19	23788
≥ 1	5.01	0.20	67446	4.60	0.18	20616
Other						
HHI <0.15	6.13	0.26	141869	4.56	0.22	24450
≥ 0.15	6.23	0.31	79881	4.55	0.19	15854
CA	6.55	0.31	118123	4.78	0.20	14929
non-CA	5.73	0.32	116543	4.45	0.20	29475
All	6.14	0.33	234666	4.56	0.20	44404

4.2. Trends in CV and HHI

Figure 6 shows the quarterly CV for the entire U.S. One can see no trend over the full 2000-14 period, although clearly an increase during the 2008-14 period, on which we focus. Note that the denominator for CV, price, has been declining steadily over this time, as seen in Figure 1. We also show the trend in quarterly CV for the largest counties (Figure 7). The 2 largest are clearly increasing while the 3rd and 4th are slightly decreasing. We show state level trends in Figure 8.

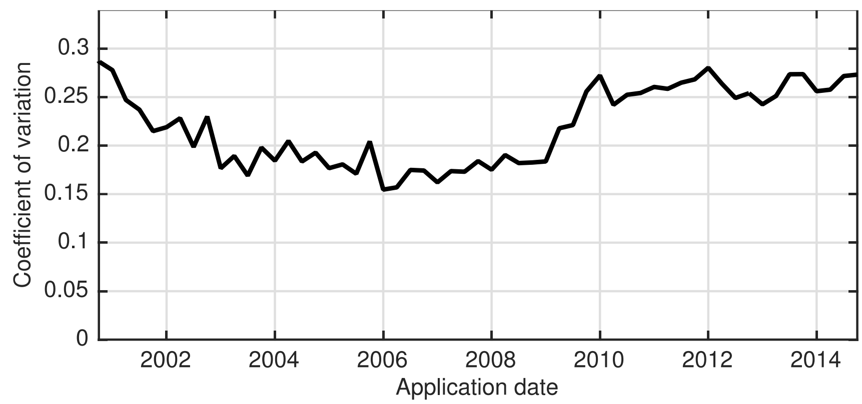

Figure 6. Quarterly coefficient of variation, all U.S.

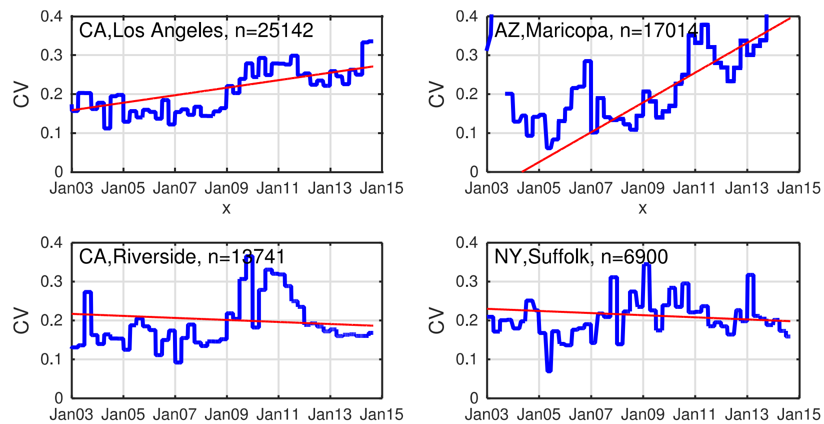

Figure 7. Trend in quarterly CV in largest counties (based on 2014 installs). N is for all years.

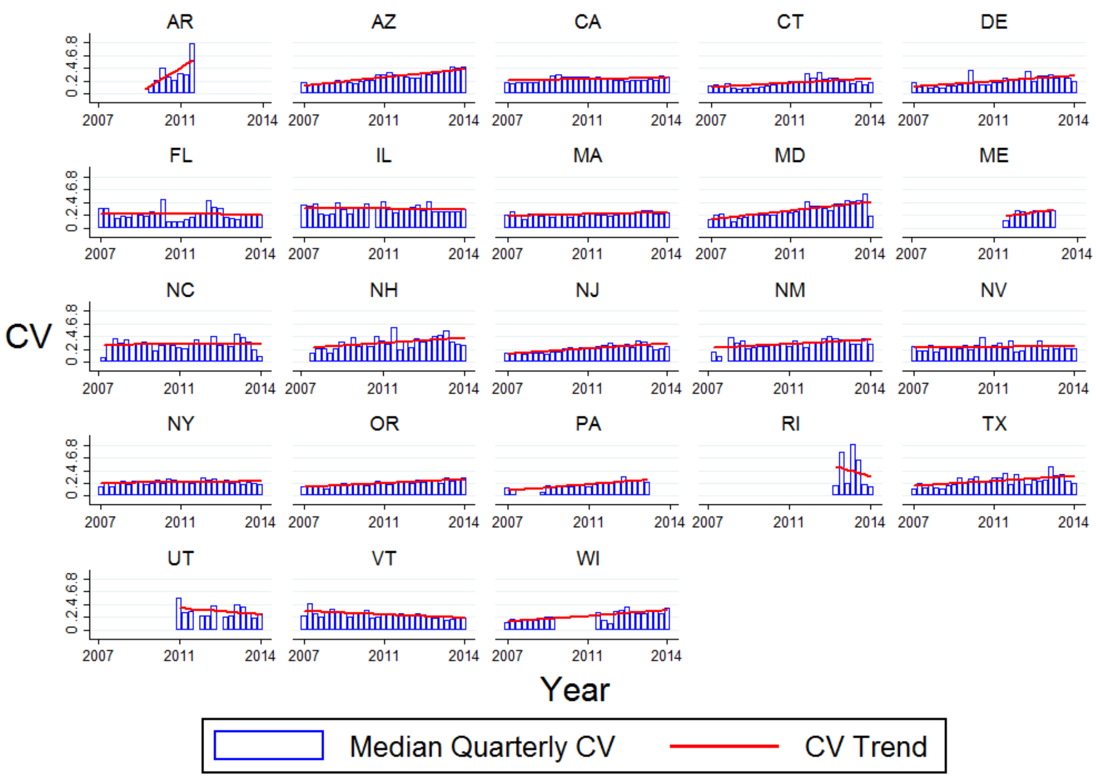

Figure 8. Trend in quarterly coefficient of variation (CV) by state.

In Figure 9 we compare the level of recent (2014) CV (horizontal axis) with the annual trend (% change per year) in CV from 2008-14 (vertical axis) for each state. The size of each circle represents the number of installations. Trends in CV that are significantly different from zero are marked in red. The total US means are marked by a small circle that is not scaled by count of installs. In some markets, CV has been rising and is now quite high (AZ and MD). CV is only decreasing in a few small markets. NJ's 2014 CV is below average but has been increasing. NY and MA are notable as large markets in which CV is below average and relatively stable. CA is about at the U.S. average (in part because it is influential in determining the average) and has been increasing at below the U.S. average. To sum up Figure 9, 1) in almost every state CV is above the average in the literature on price dispersion (0.16) and 2) in almost every state CV has been increasing.

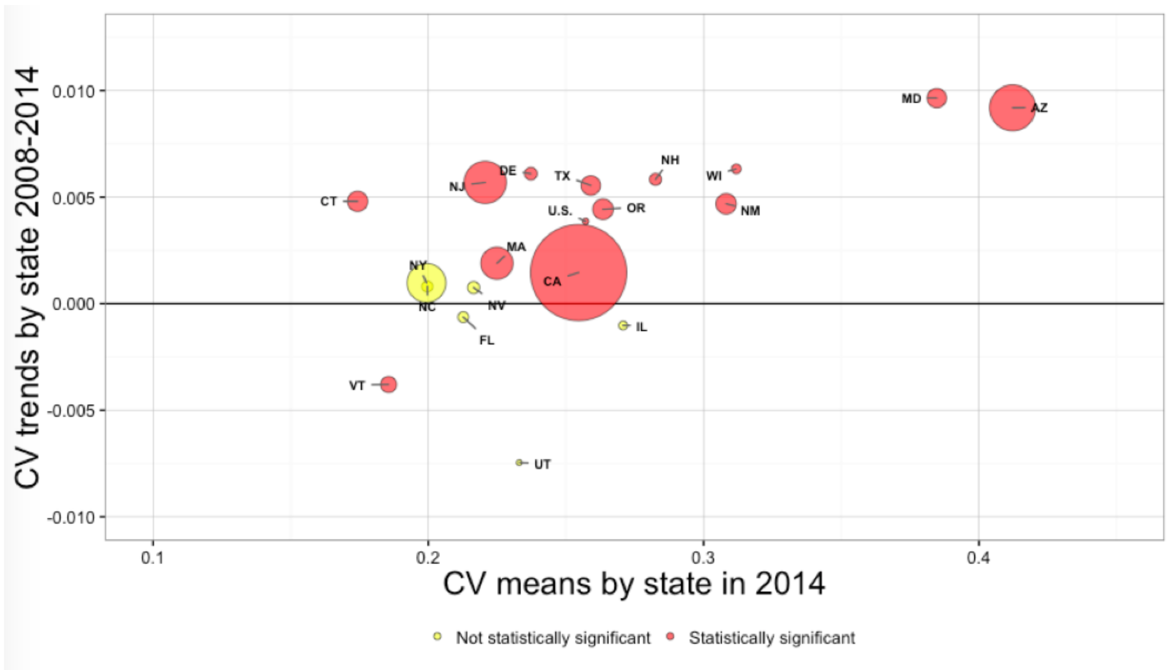

Figure 9. Trend in CV 2008-14 and level of CV in 2014 by state.

We also look at the trends in HHI by market by quarter. Figure 10 shows HHI for the 6 largest counties. Concentration has generally been increasing in recent years. Only in AZ (Maricopa) is it particularly high.

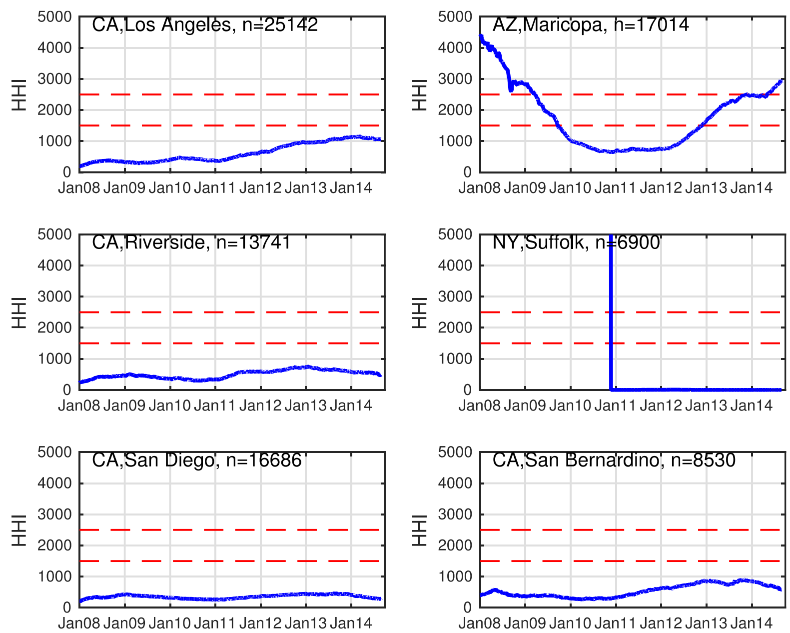

Figure 10. Trend in HHI in 6 largest markets using county definitions (based on 2014 installs). N is for all years. Dashed lines are DOJ thresholds for 'moderately and highly concentrated. Note that the third largest county, Suffolk County NY, is missing installer names.

5. ANALYTICAL RESULTS

5.1. Estimator

We regress our measure of price dispersion, the coefficient of variation (CV), on variables that the literature has pointed to as important for price dispersion. Except for the system characteristics, all variables are aggregated as means at the county level by quarter. All variables are in logs. In the base specification, we have 5153 observations.

$$CV_{it} = \beta_0 + \beta_1 COMP_{it} + \beta_2 FIRM_{it} + \beta_3 SRCHCOST_{it} + \beta_4 SRCHBENEFIT_{it} + \beta_5 MOD_{ist} + s + t + e_{it}$$

for each market *i*, and quarter *t*. COMP includes our variable for measuring competition in a market: market-level HHI. FIRM includes all firms' county-level experience, as well as labor costs in the market. SRCHCOST is a vector of variables that might affect the costs of search for pricing information: education, household income, TPO, market density (households per km^2), neighbors who have installed PV, and a proxy for the availability of third-party quotes. SRCHBENEFIT is a vector of variables that affect the payoffs of searching for information: system size and the value of solar. MOD is a vector of PV system characteristics including module price index (national average), module efficiency, whether the modules are manufactured in China or are thin film, and whether a microinverter is used. For the system characteristics, we measure the dispersion (CV) within a county quarter, rather than the means. To allow for state-quarter fixed effects, we also add binary variables for the state and the quarter. In Figure 11 we show a kernel density plot of the dependent variable, CV by quarter by county.

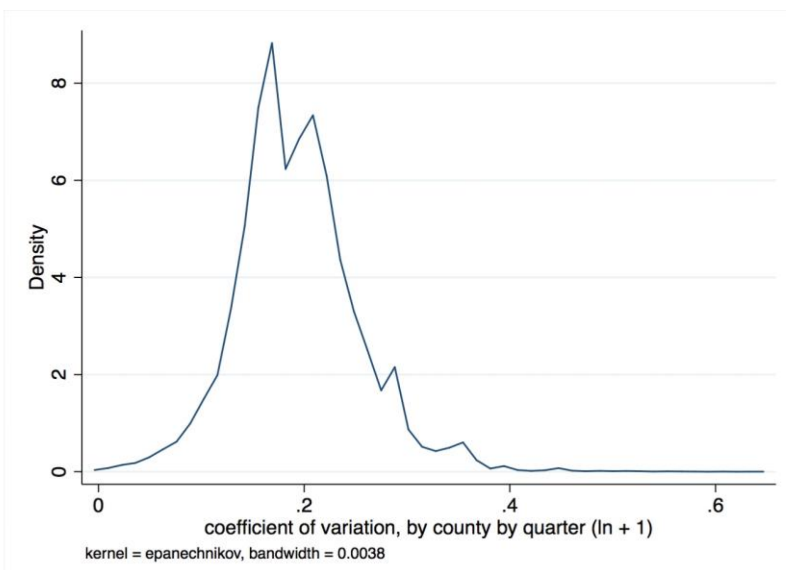

Figure 11. Dependent variable: logged (y=ln(cv) + 1) coefficient of variation by quarter by county 2008-14.

5.2. Main Results

Table 2 shows our main results regressing the coefficient of variation on the variables described above, with all variables in logs. The first nine variables in Table 2 (from "Concentration (HHI)" through "U.S. module prices") are used in every specification. Model 1, our base specification, uses market density (households/km) to represent the efficiency of information transmission. Model 2 substitutes the count of PV neighbors in a 1km radius for market density; as indicated in the 'N' row, note that we lose about 1/3 of observations that are missing geocoded addresses. Model 3 uses Model 1 and adds another form of information transmission, whether third-party quotes were available at the time of the application. Model 4 is the same as Model 1 but drops the state and quarter fixed effects to assess their impacts. Model 5 adds a proxy for installer quality: whether the installer's performance ratio is above average. Model 6 adds system characteristics, which might proxy for quality, to model 1. Both model 5 and 6 omit large portions of the database for which we are missing those added characteristics.

Table 2. Coefficient estimates from regressions by quarter by county on Xs for 2008-14 installs (dependent variable=log CV). State and quarter effects included. System characteristics measure dispersion within a county quarter. All other x's are the logged mean for each county and quarter.

	(1) Base	(2) Neighbors	(3) Quotes	(4) no FEs	(5) Instllr perf	(6) Module
Concentration (HHI)	-0.0273*	-0.0269	-0.0256*	-0.00190	-0.0624**	-0.0181
TPO (%)	-0.0440***	-0.0578***	-0.0412***	-0.0521***	-0.0685***	-0.0615***
Experience (installs)	0.00989***	0.00921***	0.0101***	0.00872***	0.00635***	0.00492**
Solar value	-0.0370***	-0.0332**	-0.0365***	-0.0302***	-0.0542***	-0.0411***
Size (kW)	-0.0270***	-0.0282**	-0.0279***	-0.0428***	-0.0270*	-0.00820
Education (bach.)	-0.0693**	0.0274	-0.0702**	-0.0522*	-0.0434	-0.0405
Income (100k+)	0.107***	0.0305	0.110***	0.141***	0.106**	0.0740*
Avg. wage	-0.00441	-0.00396	-0.00404	-0.0101	-0.00836	-0.00901
U.S. module prices	-0.526	-0.564	-0.540*	-0.0497***	-0.407	-0.0483
Market density	0.00127		0.00130	-0.00220	0.000302	0.00359*
Neighbors (1km,1yr)		-0.0108**				
Quotes (per install)			-0.00672**			
Installer perf. ratio					0.00570	
Module efficiency						0.128***
Chinese module						0.0435***
Thin film						0.0428*
Micro-inverter						0.0122
N	5153	3618	5153	5153	2349	3511
r2_a	0.132	0.113	0.133	0.089	0.253	0.183

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The base specification includes 5153 county-by-quarter observations, with substantially fewer observations in other specifications when we include variables that are missing for many counties: addresses, performance, and module information. Measures of fit make clear that much of the variation in CV is unexplained by the variables we include here—as is consistent with most studies of price dispersion. However, these fits are quite high compared to the set of studies on price dispersion that we review in section 2. Fit is notably higher in model 5, which includes installer quality. That variable however is not significant indicating that the high fit in model 5 may be due to restricting the data to primarily California installations.

To put the results shown in Table 2 in context, Figure 12 provides estimates of the magnitudes of the effects of each of the significant variables for the base model specification, as well as quotes (from model 2) and neighbors (from model 3). We use the 5^{th} to 95^{th} percentile range of each variable and estimate the effect on CV of moving from the mean of that variable to the end of this range. The variables are sorted from highest impact to lowest. For example, moving from the mean share of TPO systems in a county in a quarter (18%) to the 95th percentile (91% of systems are TPO) would reduce the CV by 18%. Conversely, moving TPO share from the mean to the 5th percentile (no TPO systems) would increase the CV by 4%. This effect produces the largest range of the variables considered, underscoring the importance of the ownership of the systems for price dispersion.

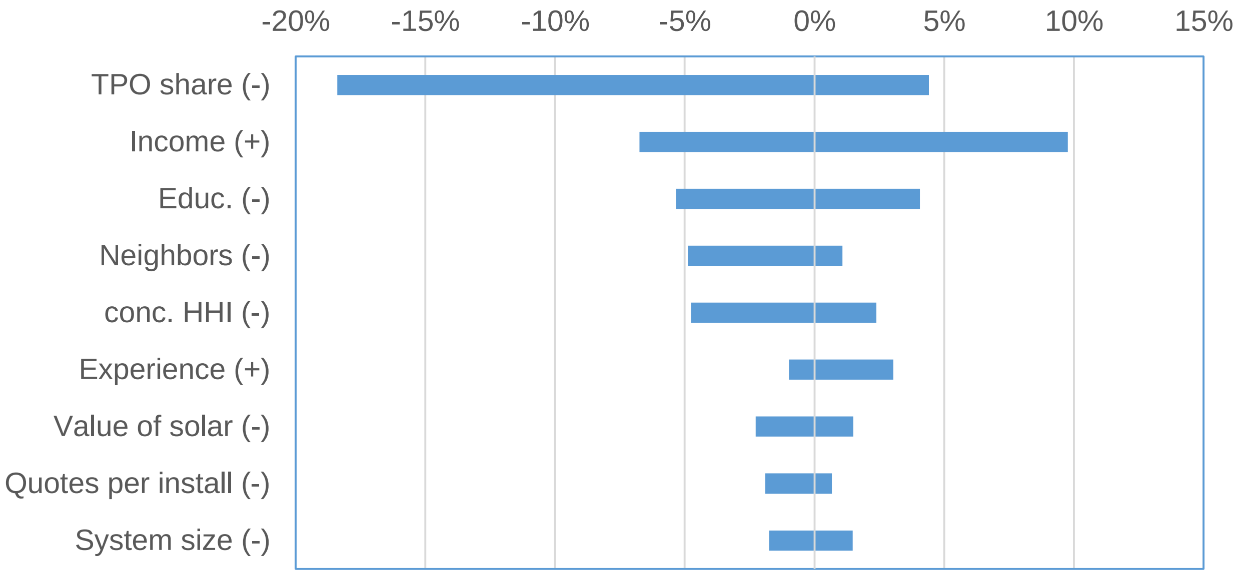

Figure 12. Sizes of effects of significant variables on coefficient of variation. Range of effects on CV shown for shifting each variable from its 5th to 95th percentile. The sign of each effect is included in parentheses.

Several variables stand out as important in these specifications:

1) *TPO*. As mentioned above, the effect of TPO share on CV is consistently negative and significant across all 6 specifications in Table 2. As mentioned in the example above, the variation in TPO has the largest effect on CV of all the variables. This result is expected given previous work showing that the price distribution of TPO systems is substantially narrower than that of customer owned systems (Nemet, O'Shaughnessy et al. 2016). We note that this variable is somewhat particular to the US PV industry, as third-party ownership is not as common elsewhere. Furthermore, the data on prices of TPO transactions are not quite as reliable as owned prices, due to the presence of standardized "block" pricing between third-party owners and solar installers as described earlier; this is true even after we have removed the clearly non-credible prices, those that use "appraised values."[1] However, these are real prices paid in the market. Moreover, there is an additional economic explanation that the TPO

[1] For "appraised value" systems, there are no customer/owner transactions for PV systems installed and owned by the same installer (integrated third-party owned installers). Hence reported prices are not transaction prices but appraised values; we have not included those systems in our sample. See Barbose and Darghouth (2016) for more details.

variable does reflect. The consumer search explanations of price dispersion discussed above emphasize the costs of consumer search, learning from repeated purchases, and payoffs from scale. For customer owned systems, there are no repeat purchases and scale is limited by roof area in most cases and is truncated at 15kW in our data. In contrast, third-party owners are involved in purchasing hundreds of systems and can spread the costs of search over megawatts rather than kilowatts. We consider alternative specifications in which we drop TPO systems in the robustness checks below, which generally show similar results as the main specifications.

2) *Neighbors*. The effect of neighbors on CV is also negative and significant (this variable is only present in Model 2). *Neighbors* here is defined as the count of systems installed in the previous 12 months within a 1 km radius of the installation. This result fits directly with the costs of consumer search explanation discussed in the economic literature. Having neighbors with recent purchase experience provides a channel for low cost information dissemination. It allows a one-time purchase to obtain some of the benefits of learning-by-shopping emphasized in the economic literature. Further, the local aspect of the information makes it likely to be both relevant and trusted, potentially providing the new customer with information about pricing strategies, negotiation experience, reliability, and installer quality. Note that we have included variables for education and income so this effect is not just a proxy for having neighbors of similar socio-economic status. We also note that the significance of income and education are lost when neighbors are added. From Figure 12 we see that the size of the effect is in the middle for these variables and that it is asymmetric, in part because many consumers have no recent purchases in their neighborhood. Moving from the mean number of neighbors (0.8) to the 95th percentile (4.1) reduces CV by 5%. We examine alternative constructions of the *neighbors* variable in the robustness checks below.

3) *Experience*. The average experience of installers in a county is associated with higher levels of price dispersion, an effect we see consistently significant across all specifications. This variable, the count of previous installations by installer, combines both knowledge gained from experience and scale. With more experience in the market there is a greater range in prices. Note that this effect is separate from a market power effect which is captured by the concentration variable; the two are negatively correlated and only weakly (-0.20). One interpretation is that part of the learning from experience by installers involves learning about the market, about adopters, and about the sales process. This could indicate that experience leads to more price dispersion because experienced installers are better able to price discriminate. We know from previous work that there appears to be evidence of value-based pricing (Gillingham, Deng et al. 2016).

4) *Value of solar*. The effect of value of solar, i.e. the sum of solar subsidies and electricity bill savings in $/W, on CV is consistently negative and significant. The higher the value of a solar system (on average in the area), the lower the price dispersion. From a consumer search perspective, this result fits generally with the literature. With high benefits to search, consumers will be more likely to invest the time required to find reputable installers and competitive pricing. Alternatively, those consumers who face a lower value of solar (due to lower retail rates, or a lack of local financial incentive) but that still install a PV system may be systematically less concerned about the underlying cost of PV and “finding the best deal”, and may instead be relatively more motivated by other considerations (the environment, impressing the neighbors, etc.).

5) *Quotes*. Our variable for *quotes* provides another measure of access to information. In model 3, the quotes variable is the number of third-party quotes provided in a county in a quarter divided by the number of systems installed. The coefficient is negative and significant. Third party provision of price quotes to consumers reduces the cost of accessing price information. Only a small number of parameters are needed for the consumer to begin to obtain quotes. Contrast this with negotiating pricing with individual installers, in which obtaining pricing is costlier and more time consuming, in part because providing pricing information is deeply embedded in the sales process. The result here is clearly in line with economic theory; having access to quotes reduces price dispersion. This result also provides some encouraging initial evidence on the effects of government programs to promote the availability of third-party quotes. The sizes of the effects are small. However, note that we are working with a subset of quote providers, so the range of 0-3 third-party quotes per installation understates the range of quotes provided.

6) *System size*: Results from economic studies of price dispersion generally indicate that as the magnitude of a purchase increases, consumers will find it beneficial to invest in search and thus price dispersion should be decreasing in purchase size. That is the result we find here. Coefficients for the size of the installation are negative in every specification. They are significant in each specification except for model 6, when the module characteristics are added and about a third of observations are dropped.

7) *Concentration*: Our measure of the effect of installer industry concentration (the HHI) is consistently negative, although not significant in every specification. The interpretation of this result is that the more concentrated the industry, i.e. less competition among installers, the lower the price dispersion. The economic literature is notably ambiguous about the expected direction of this effect. As discussed in Section 2, concentration would be expected to increase price dispersion due to market power and the ability to price discriminate. But other studies argue that the fewer installers operating in concentrated markets tend to have less dispersion in costs and weaker incentives to set extreme prices. Our results support the latter set of effects, but are not entirely clear, consistent with the ambiguity of the sign of the effect in the literature.

8) *Other variables*. We also see significant effects in other variables, but these are less robust across specifications. Education has a negative effect on price dispersion in all models except when neighbors are added in Model 2. It is significant in three of the models. Income is positive in all models but significant in only three. It also loses significance when the neighbors variable is added. The signs of these demographic variables are as expected from the literature: higher education would reduce the costs of search, and income would increase the opportunity cost of search. The result on neighbors is the dominant one, however, and has a clearer link to information transmission. Correlations between neighbors and these demographic variables are 0.27-0.30. None of the results for another variable, labor costs, are significant. One might test the result of within-county wage dispersion, but our experience in previous work is that prices are insensitive to wage variation (Gillingham, Deng et al. 2016).

That the effect of national average module prices is negative, although not generally significant, fits with the notion that consumers will find more value in engaging in search when the amounts at stake are higher. Note that we also include time fixed effects so module price really is the effect of price levels, not the underlying time trend in CV. It thus also makes sense that module price index becomes highly significant only in model 4 when the time trend fixed effects are dropped.

Market density, which might be a proxy for information transmission, is not significant. Our measure for installer quality, performance ratio, is not significant. We do see larger effects for value of solar and TPO in Model 5 when it is added. Unsurprisingly, the size of the experience effect is small in Model 5, implying that some of the experience effect is attributable to quality in the form of improved performance.

The effects of dispersion in system characteristics (efficiency, Chinese, and thin film), in a county quarter, are positive and significant. The fit is also higher. Expectedly, variation in system characteristics explains some of the dispersion in prices.

5.3. Robustness checks

We run several alternative specifications to assess whether the results above remain robust under alternative assumptions on market definitions, neighbors, quotes, third-party ownership, and time trend. First, in Table 3 we run the same models as in Table 2 but using the alternative definitions of markets based on O'Shaughnessy, Nemet et al. (2016), rather than counties. We find similar results. Fits are generally higher, now with more than twice as many observations in each model. All of the seven primary variables of interest retain their size and significance. Neighbors and quotes retain their signs and significance; the neighbors coefficient is smaller and the quotes is larger. Concentration is now consistently significant. We note that this is the variable for which we would expect the new market definitions to have the largest change. HHI is especially sensitive to whether markets are defined by counties or by installer activities. *It thus bolsters credence that price dispersion is decreasing in concentration rather than increasing.* As mentioned above, we again have no theoretical prediction about what direction the effect of HHI will have. The main reason we use the new market definitions as a robustness check, rather than as our preferred specification, is that there may be some endogeneity in how the markets are defined. That is, installers' decisions about entering and pricing in a market may play a role in how the borders of markets are defined. That is not true of counties, which are clearly exogenous to installers' decisions.

Table 3. Coefficient estimates from regressions of y=log CV by quarter by new market definitions on Xs for 2008-14 installs. State and quarter effects included. System characteristics measure dispersion within a county quarter. All other x's are the logged mean for each county and quarter.

	(1) Base	(2) Neighbors	(3) Quotes	(4) no FEs	(5) Instllr perf	(6) Module
Concentration (HHI)	-0.0944***	-0.0767***	-0.0945***	-0.0975***	-0.113***	-0.0918***
TPO (%)	-0.0432***	-0.0520***	-0.0401***	-0.0477***	-0.0602***	-0.0516***
Experience (installs)	0.00354***	0.00453***	0.00360***	0.00580***	0.00316**	0.00402***
Solar value	-0.0515***	-0.0491***	-0.0486***	-0.0466***	-0.0535***	-0.0502***
Size (kW)	-0.0375***	-0.0347***	-0.0373***	-0.0488***	-0.0345***	-0.0288***
Education (bach.)	-0.0169	-0.000853	-0.0166	-0.0335**	-0.0384**	-0.00597
Income (100k+)	0.0326*	0.0206	0.0345*	0.0556***	0.0387*	0.00687
Avg. wage	-0.00135	0.00828	-0.00141	-0.0310**	-0.00406	0.0143
U.S. module prices	-0.00423	-0.0524	-0.0178	-0.0408***	0.0519	-0.0456
Market density	0.00120		0.00120	-0.00165	0.00350**	0.00208
Neighbors (1km,1yr)		-0.00430*				
Quotes (per install)			-0.00826***			
Installer perf. ratio					-0.00122	
Module efficiency						0.0804**
Chinese module						0.0331***
Thin film						0.0607***
Micro-inverter						-0.0169
N	12384	9444	12384	12415	7101	10586
r2_a	0.154	0.124	0.156	0.117	0.217	0.165

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Second, we include further robustness checks on key variables related to consumer search: neighbors and quotes (Table 4). For comparison, we include Model 1 from Table 2, and in Model 2 the original neighbors variable from Table 2. Model 3 includes an alternative definition of neighbors, the count of PV installations within 1km in all previous periods, rather than restricting to the 12 months prior. Model 4 uses the quote variable from Table 2 (third-party quotes per installation) and Model 5 uses the alternative measure: whether any third-party quotes had been provided in that county in that quarter or any previous quarter. We find that the inclusion of the effect of neighbors who have installed PV more than a year previously is quite similar to the effect of recent neighbors. This is likely due to the imbalance in the data over time; there are many more installations in recent periods so older neighbors are relatively rare. In the Appendix (Table 9) we compare results for eight versions of the neighbors variable: radii of 0.5, 1, 2, and 5 kilometers over 1 and 2 years. Unsurprisingly, the effect diminishes with distance bolstering our confidence in the variable as an information transmission proxy. The effect also diminishes over time. The effects become insignificant beyond 2 km after one year, and beyond 0.5km after 2 years. *Nearby neighbors who have installed PV recently are most strongly associated with lower price dispersion.*

Quotes become insignificant with the alternative specification (Table 4). The quantity of quotes available in a market is significant but whether any quote had been made is not. While this does not enhance the robustness of the findings on the quote variable, it is also unsurprising. The binary quotes available variable is a blunter measure than the density of quotes and thus less powerful.

Table 4. Coefficient estimates from regressions of y=log CV by quarter by county on Xs for 2008-14 installs. State and quarter effects included. Each x is the logged mean for each county and quarter.

	(1) Base	(2) Neighbors	(3) Neighbors2	(4) Quotes	(5) Quotes2
Concentration (HHI)	-0.0273*	-0.0269	-0.0247	-0.0256*	-0.0263*
TPO (%)	-0.0440***	-0.0578***	-0.0567***	-0.0412***	-0.0423***
Experience (installs)	0.00989***	0.00921***	0.0116***	0.0101***	0.0101***
Solar value	-0.0370***	-0.0332**	-0.0312**	-0.0365***	-0.0377***
Size (kW)	-0.0270***	-0.0282**	-0.0286**	-0.0279***	-0.0275***
Education (bach.)	-0.0693**	0.0274	0.0294	-0.0702**	-0.0693**
Income (100k+)	0.107***	0.0305	0.0301	0.110***	0.108***
Avg. wage	-0.00441	-0.00396	-0.00544	-0.00404	-0.00454
U.S. module prices	-0.526	-0.564	-0.555	-0.540*	-0.525
Market density	0.00127			0.00130	0.00126
Neighbors (1km,1yr)		-0.0108**			
Neighbors (1km,all yrs)			-0.0104***		
Quotes (per install)				-0.00672**	
Quotes available					-0.00824
N	5153	3618	3618	5153	5153
r2_a	0.132	0.113	0.114	0.133	0.132

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Third, we consider alternative approaches to characterizing TPO systems and the time trend. In Table 10 in the Appendix we assess the changes to effects when dropping TPO systems altogether, rather than controlling for them with a dummy. The results are generally the same. Quotes and neighbors remain negative and significant, but in both cases are larger. Our interpretation is that the access to information that neighbors and quotes provide matters more for customer owned systems than for TPO systems, as would be expected given the different business processes underlying TPO system pricing. Finally, in Table 11 we include a time trend rather than quarter dummies. The time trend is negative and significant indicating that price dispersion has fallen over the 2008-14 time period after controlling for the variables included in models 2-6.

6. DISCUSSION AND CONCLUSION

This study assesses the existence of and factors affecting price dispersion in U.S. residential PV installations from 2008 to 2014. We find levels of price dispersion higher than those of the mean of previous studies of price dispersion in other areas. While price dispersion is empirically common, our results indicate that PV markets violate the law of one price even more than many other products. Some explanations are 1) that PV systems are not actually a homogenous good, and 2) that PV markets are not fully competitive. We aim to address point 1 by controlling for size, system characteristics and the value of solar. We address point 2 by carefully accounting for market concentration, using multiple market definitions. Still, we see wide dispersion in prices even after controlling for product differentiation and market structure. Crucially, our 7-year time series shows that price dispersion is persistent; it has not substantially diminished as PV markets have matured. The existence of price dispersion matters in part because some consumers are paying higher, and in some cases much higher, prices than others. The perhaps more important outcome of persistent price dispersion is the set of consumers that we do not see in our data: those which do not adopt solar because they perceive that prices are higher than what is actually available or because they are unwilling to invest the time in search (Rai and Beck 2015).

We thus conclude that finding ways to reduce price dispersion could have substantial public benefits. Unless there is a strong and persistent barrier to entry, reducing price dispersion will lower not raise consumer prices. The results we discuss in Section 5 are manifold. But perhaps the most robust insight is that the evidence of an information problem in the residential PV market fits well with the economic literature on consumer search. Where and when the benefits of search are high (large system size, value of solar), price dispersion is smaller. Where and when the costs of search are low (having neighbors with PV, having many quotes in the market) price dispersion is also smaller. These results are thus encouraging for policy programs that increase access to information, such as by making installer price quotes more broadly available. A successful example is the Dept. of Energy's incubator award for EnergySage. The results also make clear that pricing is more competitive, i.e. smaller CV, in clusters of installations where potential consumers can make use of a trusted and relatively accessible information source, their neighbors. On the one hand, this might suggest that targeting adoption in existing clusters has higher potential to result in low-priced solar. On the other hand, it suggests that establishing new clusters likely needs extra assistance in providing potential consumers with reliable information about installer quality and pricing.

A. APPENDIX: DATA SET DESCRIPTIVE STATISTICS, VARIABLE DEFINITIONS

A.1. Additional descriptives

Table 5. Variable definitions for variables used in main regression results.

Variable name	Variable label	Defiinition
hhi_county	hhi, cnty	HHI index (0-1) for county (last 12 months)
tpo_implied	TPO	dummy, 1 if TPO
agg_exp_county	agg exp cnty	depr. experience in county, all installers
solarvalue_w	value of solar, W	value of solar ($/W)
sys_size_kw	sys size kW	system size in kW
pop_edu4_more_bach_zip	edu college zip	percent completed Bachelor, in zip
num_hous_incom4_zip	inc 100k zip	pct HH income >100k, in zip
labor_cost_ind_100k	wage cnty 100k	labor cost index: 2.5 (admin) to 2 (roof) to 1 (electr), by county
mod_price_indx double	mod pr index	monthly module prices ($/W) at time of application
market_density_1km	HHs per sq km	local market density (owner-occupied HHs within county/sq.km.)
effic_1	mod eff	module efficiency
china_dum	china panel	dummy, 1 if panel made in China
thinfilm_dum	thin film	dummy, 1 if thin Film
microinv_dum	micro invrtr	dummy, 1 if micro-inverter

Table 6. Summary statistics for all variables for systems installed 2008-14. Note: This represents the full set of variables considered for the present work. Only a subset was ultimately used, focusing on those linked with previous economic literature on price dispersion.

Variable	Obs	Mean	Std. Dev.	Min	Max	P50
price per W	234666	6.143	2.135	1.006	24.978	5.789
price 000s	234666	35.221	17.262	1.312	298.208	31.809
CV, qtr new mkt	230152	.209	.082	0	1.102	.202
CV, qtr county	232453	.215	.074	0	.902	.208
CV, qtr state	234661	.238	.053	.067	.808	.228
installers per 1k HHs	233075	.13	.099	0	1.443	.112
mkt duration yrs	234666	11.458	3.227	0	16.668	11.904
hhi, cnty	234666	.15	.188	0	1	.093
hhi, new mkt	221750	.173	.177	.027	1	.111
active instllrs	234666	41.737	43.391	0	189	26
exp cnty	217022	87.327	174.324	0	5740.962	24.92
exp state	217022	394.957	693.717	0	11028.89	99.573
agg exp cnty	234666	1930.344	2628.321	0	12157.63	852.011
mkt share, cnty	234666	.233	.341	0	1	.063
mkt share, new mkt	234666	.368	.4	.001	1	.152
inst scale cnty	234666	29.576	62.452	0	977	6
inst scale st	234666	106.729	177.923	0	3112	28
HHs per sq km	233075	104.016	151.21	.078	2942.207	53.204
nbrs 1 km in 1 yr	154413	3.325	6.331	0	154	1
TPO	211109	.413	.492	0	1	0
edu college zip	232943	.365	.17	0	.923	.344
inc 100k zip	232943	.343	.153	0	.919	.336
pct demo cnty	231015	.554	.113	.098	.912	.526
value of solar, w	234666	6.379	2.349	1.392	17.054	5.948
value of solar, k	234666	37.441	20.955	3.752	214.636	32.809
pct srec	234666	.056	.12	0	.475	0
interconnect	204218	14.399	6.218	-3	27.5	11.5
wage cnty 100k	234137	.593	.165	.195	3.804	.563
mod pr index	234666	1.528	.922	.767	4.257	1.076
inv pr index	194132	.347	.08	.252	.52	.329
sys size kW	234666	6.038	2.718	1	15	5.61
mod eff	168899	.158	.02	.055	.212	.153
china panel	174386	.335	.472	0	1	0
thin film	234666	.013	.115	0	1	0
micro invrtr	195131	.278	.448	0	1	0
performance	308	.703	.113	.29	1.007	.714
Installer avg PR	66992	.73	.077	.36	.989	.73
Num quotes	234685	358.304	702.584	0	5362	16
install year	234666	2011.722	1.758	2008	2014	2012
app month	234666	167.494	21.275	117	201	170

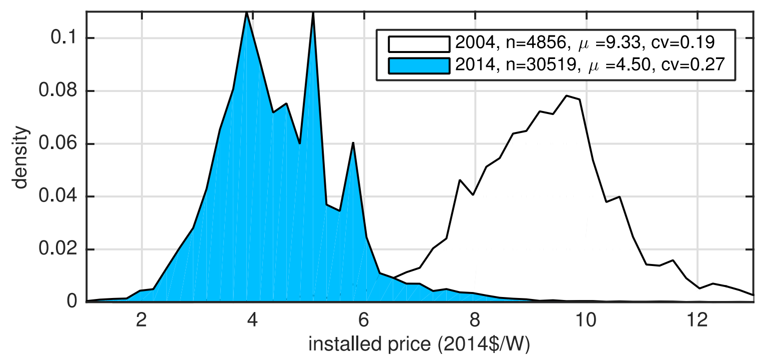

Figure 13. Distributions of installed prices (constant $/W).

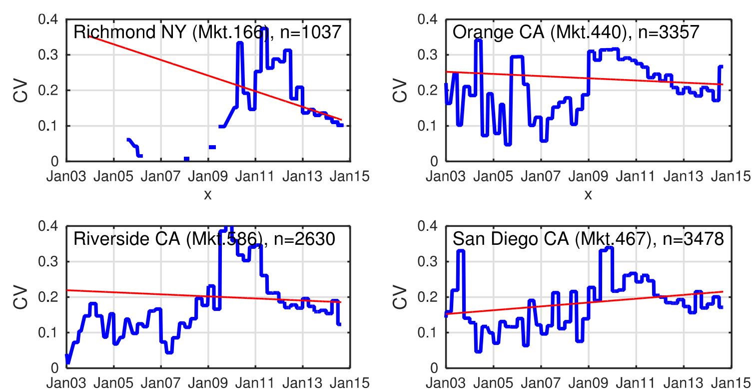

Figure 14. Trend in quarterly CV in the largest new market definitions (based on 2014 installs). N is for all years.

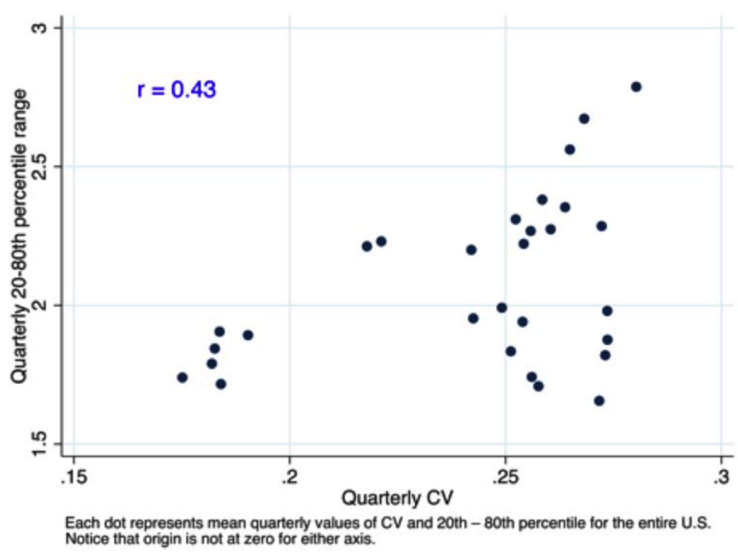

Figure 15. Comparison of CV and the measure of dispersion typically used in TTS, the 20-80 range.

Table 7. Correlations among regressors in specification #2.

	HHI	TPO	Exp.	VoS	Size	Educ	Income	Wage	Mod Price
HHI	1								
TPO	-0.241	1							
Exp.	-0.208	0.268	1						
VoS	-0.253	-0.102	0.023	1					
Size	-0.081	0.242	0.036	-0.131	1				
Educ	-0.197	0.026	0.145	0.030	-0.133	1			
Income	-0.244	0.217	0.253	0.138	0.044	0.772	1		
Wage	-0.077	0.075	0.095	0.038	-0.030	0.336	0.392	1	
Mod Price	0.103	-0.309	-0.118	0.534	-0.300	0.011	-0.021	-0.034	1
Neighbors	-0.236	0.226	0.314	0.257	-0.067	0.277	0.321	0.206	-0.076

Table 8. Summary statistics for neighbors variables for systems installed 2008- 14.

Variable	Obs	Mean	Std. Dev.	Min	Max
distance to nearest previous install (km)	154413	.788	4.111	0	505.605
count of all prev installs within nbrhood radius (2km)	154413	51.365	66.15	0	893
dist rcnt nbr	154413	19.964	90.452	0	1607.755
nbrs 1 km in 1 yr	154413	3.325	6.331	0	154
count of recent (past year) installs	154413	24103.55	7460.451	7935	34080
nbrs 1/2 km	154413	5.368	7.402	0	122
nbrs 1/2 km in 1 yr	154413	1.027	2.281	0	81
nbrs 1 km	154413	16.618	21.933	0	348
nbrs 1 km in 1 yr	154413	3.325	6.331	0	154
nbrs 5 km	154413	212.902	248.817	0	2479
nbrs 5 km in 1 yr	154413	45.101	67.262	0	713
nbrs 1/2 km in 2 yrs	154413	1.89	3.584	0	98
nbrs 1 km in 2 yrs	154413	6.094	10.686	0	262
nbrs 2 km in 2 yrs	154413	19.397	32.454	0	598
nbrs 5 km in 2 yrs	154413	82.417	120.664	0	1379

A.2. Additional robustness checks

We include the results of additional analyses using alternative estimators.

Table 9. Sensitivity to neighbors definitions. Coefficient estimates from regressions of y=log CV by quarter by county on Xs for 2008-14 installs. State and quarter effects included. Each x is the logged mean for each county and quarter.

	(1) 1km,1yr	(2) 0.5km,1yr	(3) 2km,1yr	(4) 5km,1yr	(5) 0.5km,2yr	(6) 1km,2yr	(7) 2km,2yr	(8) 5km,2yr
Concentration (HHI)	-0.0269	-0.0274	-0.0276	-0.0287	-0.0265	0.0189	0.0184	0.0177
TPO (%)	-0.0578***	-0.0581***	-0.0580***	-0.0595***	-0.0581***	-0.0658***	-0.0660***	-0.0662***
Experience (installs)	0.00921***	0.00883***	0.00952***	0.00944***	0.00925***	0.00857***	0.00860***	0.00864***
Solar value	-0.0332**	-0.0331**	-0.0333**	-0.0339**	-0.0327**	-0.00551	-0.00581	-0.00622
Size (kW)	-0.0282**	-0.0280**	-0.0279**	-0.0263**	-0.0282**	-0.0427***	-0.0425***	-0.0421***
Education (bach.)	0.0274	0.0246	0.0279	0.0199	0.0282	-0.0129	-0.0129	-0.0128
Income (100k+)	0.0305	0.0317	0.0290	0.0318	0.0310	0.0800*	0.0797*	0.0798*
Avg. wage	-0.00396	-0.00364	-0.00535	-0.00799	-0.00383	-0.0127	-0.0130	-0.0129
U.S. module prices	-0.564	-0.567	-0.568	-0.570	-0.568	-0.0488***	-0.0487***	-0.0485***
Neighbors (1km,1yr)	-0.0108**							
Neighbors (0.5km,1yr)		-0.0158**						
Neighbors (2km,1yr)			-0.00756**					
Neighbors (5km,1yr)				-0.00388				
Neighbors (0.5km,2yr)					-0.0144***			
Neighbors (1km,2yr)						-0.00283		
Neighbors (2km,2yr)							-0.00191	
Neighbors (5km,2yr)								-0.00141
N	3618	3618	3618	3618	3618	3618	3618	3618
r2_a	0.113	0.113	0.112	0.111	0.113	0.068	0.068	0.068

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 10. Effects of dropping TPO systems. Models 2-6 omit TPO systems. Coefficient estimates from regressions of y=log CV by quarter by county on Xs for 2008-14 installs. State and quarter effects included. System characteristics measure dispersion within a county quarter. All other x's are the logged mean for each county and quarter.

	(1) Base TPO	(2) Base	(3) Neighbors	(4) Quotes	(5) no FEs	(6) Module
Concentration (HHI)	-0.0273*	-0.0312**	-0.0330*	-0.0287**	-0.000338	-0.0227
TPO (%)	-0.0440***	0	0	0	0	0
Experience (installs)	0.00989***	0.00798***	0.00639***	0.00834***	0.00676***	0.00122
Solar value	-0.0370***	-0.0575***	-0.0457***	-0.0550***	-0.0381***	-0.0702***
Size (kW)	-0.0270***	-0.0222***	-0.0239**	-0.0230***	-0.0439***	-0.00000679
Education (bach.)	-0.0693**	-0.0581*	0.0562	-0.0600*	-0.0290	-0.00121
Income (100k+)	0.107***	0.0966***	0.00651	0.100***	0.118***	0.0441
Avg. wage	-0.00441	-0.00111	0.00218	-0.000409	-0.00934	-0.00845
U.S. module prices	-0.526	-0.485*	-0.469	-0.498*	-0.0385***	-0.0645
Market density	0.00127	0.00163		0.00163	-0.00288*	0.00422**
Neighbors (1km,1yr)			-0.0123***			
Quotes (per install)				-0.00743**		
Module efficiency						0.142***
Chinese module						0.0295**
Thin film						0.0395*
Micro-inverter						0.0122
N	5153	4964	3520	4964	4964	3380
r2_a	0.132	0.129	0.102	0.130	0.081	0.173

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 11. Include application year as an X to see trend in CV. Coefficient estimates from regressions of y=log CV by quarter by county on Xs for 2008-14 installs. State effects included (but not quarter effects). System characteristics measure dispersion within a county quarter. All other x's are the logged mean for each county and quarter.

	(1) Base (orig.)	(2) Base	(3) Neighbors	(4) Quotes	(5) no FEs	(6) Module
Concentration (HHI)	-0.0273*	-0.0349**	-0.0307*	-0.0313**	0.00135	-0.0250
TPO (%)	-0.0440***	-0.0374***	-0.0538***	-0.0352***	-0.0559***	-0.0553***
Experience (installs)	0.00989***	0.00975***	0.00890***	0.00998***	0.00995***	0.00442**
Solar value	-0.0370***	-0.0307***	-0.0290**	-0.0329***	-0.0411***	-0.0378***
Size (kW)	-0.0270***	-0.0278***	-0.0303***	-0.0290***	-0.0397***	-0.0101
Education (bach.)	-0.0693**	-0.0715**	0.0245	-0.0717**	-0.0582*	-0.0461
Income (100k+)	0.107***	0.112***	0.0328	0.115***	0.142***	0.0794*
Avg. wage	-0.00441	-0.00576	-0.00409	-0.00532	-0.00529	-0.00813
U.S. module prices	-0.526	-0.0833***	-0.0886***	-0.0904***	-0.0969***	-0.0733***
Market density	0.00127	0.000967		0.00110	-0.00247*	0.00325*
Year applied		-15.53***	-18.67***	-15.39***	-20.52***	-13.96**
Neighbors (1km,1yr)			-0.0105**			
Quotes (per install)				-0.00982***		
Module efficiency						0.126***
Chinese module						0.0501***
Thin film						0.0299
Micro-inverter						0.0218
N	5153	5153	3618	5153	5153	3511
r2_a	0.132	0.127	0.110	0.130	0.093	0.180

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

A.3. Calculating performance ratio

We use a performance ratio (PR) to proxy for system quality. Some installers are able to charge premiums, potentially based on reputation for high quality, which we proxy for as enhanced system performance over time. We define “quality” as the ability of a system to achieve production that closely approximates expected production over time. By constructing a ratio of observed average annual production to expected we seek to account for some of the variation in price resulting from the installer providing services resulting in quality not reflected in system size but in the installed price ($/Watt). Figure 16 shows the resulting performance ratio values.

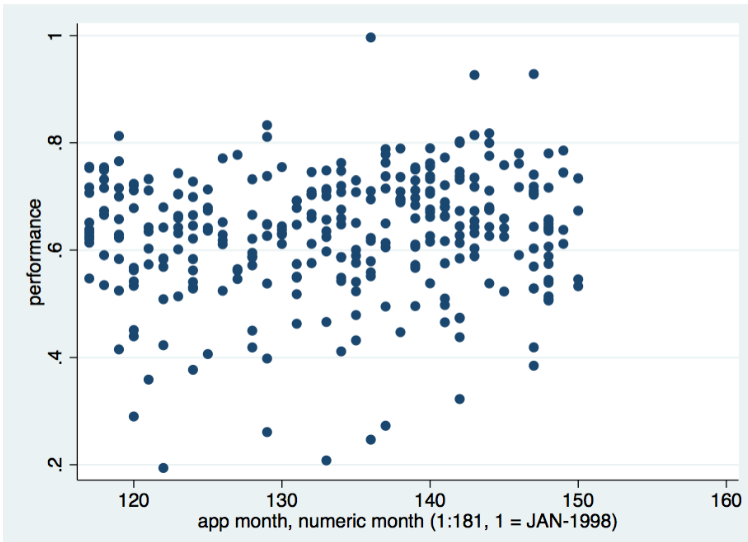

Figure 16. A proxy for quality: performance ratio 2010-14 by install date.

Data

Alternating current (AC) capacity comes from the California Solar Initiative (CSI) (https://www.californiasolarstatistics.ca.gov/data_downloads/).

Production data comes from CSI 15-Minute Interval Data (https://www.californiasolarstatistics.ca.gov/data_downloads/). Comprised of metered generation from 504 CSI systems starting in 2010 and ending in 2016 and time stamped in fifteen minute intervals, this data provides the measured production. Measured systems come from the three largest utility service areas in California: Pacific Gas & Electric, Southern California Edison, and San Diego Gas & Electric.

Irradiance data comes from the National Solar Radiation Database (NSRDB) Physical Solar Model (PSM). This database has a spatial resolution of 4x4 kilometers, at half-hour intervals, and includes years 1998 to 2014.

Tracking the Sun (TTS) system data are matched to production data to generate this data sample, which includes 161 different installers.

Census tract GIS layers came from U.S. Census cartographic Boundary shapefiles (https://www.census.gov/geo/maps-data/data/cbf/cbf_tracts.html).

Procedure

To produce system-specific PRs we matched systems in the TTS data set with production data from the California Solar Initiative 15-Minute Interval Data. Of the 504 systems matched in TTS, 424 comprise this analysis. We aggregated 15-minute interval CSI production data to the monthly level. We also filtered out negative production values. System outages due to equipment malfunction are included in the data sample used for analysis. These data are important to distinguish systems that are underperforming relative to their rated potential.

Generating system-specific PRs required matching locations of irradiance to system locations. The TTS dataset includes geocoded system locations, as do NSRDB irradiance measurements. Using ArcGIS system locations and irradiance measurements along with census tract data layers, we plotted and overlaid system locations and irradiance measurements on census tracts in California. We then identified census tract centroid points and used inverse distance weighting to generate a weighted average irradiance value for each census centroid associated with the closest system location. Similar to production data, we aggregated 30-minute average irradiance values to monthly insolation values for the PR calculation.

We control for outliers by finding values more than 2 standard deviations above and below the median for the systems in this sample. Defined as such, roughly 5% of the calculated PRs represent outliers. Using the 95th percentile as the value representing the threshold for 'healthy' measurements, all PRs are divided by the PR value representing the 95% percentile. This provides a dimensionless and normalized performance metric near unity when the system is operating near its design-specified full performance. This is an attempt to eliminate the influence of high outliers due to unknown module additions since completion of the original installation.

Median monthly PR values (Equation 1), averaged yearly, represent the average annual performance of each system. Importantly, even though production monitoring began for the majority (~95%) of systems in 2010 and 2011, the years of installation differ from initial monitoring dates. Therefore, we created an age index so that system PR comparisons come from their *i*th year of operation. The 3rd and 4th years of operation had the highest number of observations and are what we present to facilitate comparison as average annual PR values.

Performance Ratio

Equation 1 shows the PR adapted from a more general definition available from Sandia Laboratories (https://pvpmc.sandia.gov/modeling-steps/5-ac-system-output/pv-performance-metrics/performance-ratio/) with the measured divided by expected generation taken as a median annual value. Equation 1 converges in probability to the median in a normal distribution that matches the mean. The numerator represents monthly measured production in kWh. The AC rating differs from the more general definition provided above because instead of a rating based on standard test conditions (STC—1000w/m^2, 25 degrees C cell and ambient temperature, and 1.5 air mass) it uses the more conservative Photovoltaics for Utility Scale Applications (PVUSA Test Conditions or 'PTC'—1000w/m^2, 45 degrees C cell temperature, 20 degrees C ambient temperature, and 1 meter/second wind speed for module cooling). For this particular rating, the PTC–AC is multiplied by a design factor and inverter efficiency equaling the

CSI–AC rating. Importantly, it includes a design factor. The design factor is a ratio comparing a proposed system's expected generation with that of a baseline system that includes optimal tilt and azimuth, a proposed location versus the actual location, and applying an adjustment for the effect of mounting method on module temperature's impact on production. We use the CSI–AC system rating instead of the STC rating because it more closely characterizes the probable output of the entire system.

$$PR_j = \frac{1}{12}\sum_{i=1}^{12}\left[\frac{E_{ij}}{kW_{AC\_ij}\left(\frac{Ins_{ij}}{kW_{STC}}\right)}\right] \qquad (1)$$

Where:

t = Time period, monthly

$E_{i,j}$ = average metered energy output (kWh)

$kW_{AC\_i,j}$ = CSI Rating of each system (kW-AC)

$Ins_{i,j}$ = average insolation per square-meter at census centroid containing system i

kW_{STC} = irradiance at standard test conditions (STC) (1,000W/m^2)

An additional advantage of the CSI rating is that it does not penalize systems for shading and suboptimal tilt. We do not want to penalize systems with shading because installers may have little control over shaded site conditions. Although under the control of the installer, optimal tilt is unlikely for a number of reasons. First, residential installations infrequently build in tilt above the existing pitched roof. Second, on flat roofs, module density may be a more important design criterion that optimal tilt. In other words, systems are frequently designed for legitimate purposes other than maximizing annual energy output**.** This is especially true as module price continues to decline**.** In the event, we used a system rating that penalizes based on shade and tilt, systems designed for maximum energy harvest would likely have higher PR values than those designed based on other criteria. Our definition of quality seeks to capture how well a system performs over time given site-specific characteristics and design choices made to fit customers' needs.

Another choice that deviates from more conventional use of the PR that merits explanation in using the CSI rating is that it mixes AC with DC in the denominator. If the ratio between DC array and inverter capacity were much higher than 1.0, as can be frequently found in large commercial and utility scale installations, magnitudes could get distorted. This is not typically the case for residential PV systems. Additionally, when including DC rating instead of the CSI rating, the distribution is not significantly altered and we consider the inclusion of the attributes contained in the CSI rating important enough to warrant its use.

The gold standard for PV system performance is irradiance measured into the exact plane-of-array (POA) of each PV system, which is not feasible due to data limitations. Of the three kinds of irradiance data, global horizontal irradiance (GHI), diffuse horizontal irradiance (DHI), and direct normal irradiance (DNI), only GHI is used here. GHI is the sum of DHI and DNI and assumes arrays are horizontally tilted. In the absence of POA and where only GHI values exist, GHI is the preferred measure of irradiance because DNI gives values that are too high in the morning and evening and DHI misses diffuse light.

A.4. Defining non-political market definitions

Solar PV market research to date has used various boundaries based on political jurisdictions to define markets, primarily county-level boundaries. However solar PV markets have exhibited spatial clustering that defies jurisdictional boundaries (Bollinger and Gillingham 2012). The inadequacy of jurisdictional boundaries for market research in general has given rise to a body of literature on the study of market definition. The literature suggests that the use of jurisdictional boundaries (e.g., county) may misrepresent true market dynamics. In general, the task of market definition is to delineate market boundaries based on some conception of product substitutability (Massey 2000; Baker 2007; Davis and Garces 2010; Boshoff 2014).

Building on this literature, O'Shaughnessy et al. (2016) developed a market definition methodology for solar PV markets based on the spatial distribution of installers. The theoretical basis is that solar PV installers exert local price constraints on other installers within a limited geographic area. These price constraints decrease with distance because the vast majority of installers are local small-scale businesses. In order to proxy for these price constraints, an installer overlap coefficient (IOC) was developed to capture the extent to which an installer community in one region *x* resembled the installer community in another region *y*:

$$IOC_{x,y} = \rho_x^y * \rho_y^x \tag{2}$$

Where $IOC_{x,y}$ is the IOC between regions *x* and *y*, ρ_x^y is the percentage of installers in region *x* that also operate in region *y*, and ρ_y^x is the percentage of installers in region *y* that also operate in region *x.* For the purposes of the methodology, an installer is considered to operate in a given region if the installer has installed at least one system in that region over some given timeframe.

The following algorithm describes the market definition method:

A candidate market is identified as the region with the maximum summed value of IOCs

The six nearest neighboring regions to the candidate market are assigned to the candidate market

Additional regions are assigned to the candidate market if (a) the IOC with the candidate market exceeds a specified critical value (e.g., 0.25) and (b) the region is contiguous with the candidate market through other regions assigned to the candidate market.

Steps 1 through 3 are repeated until all regions are assigned to markets.

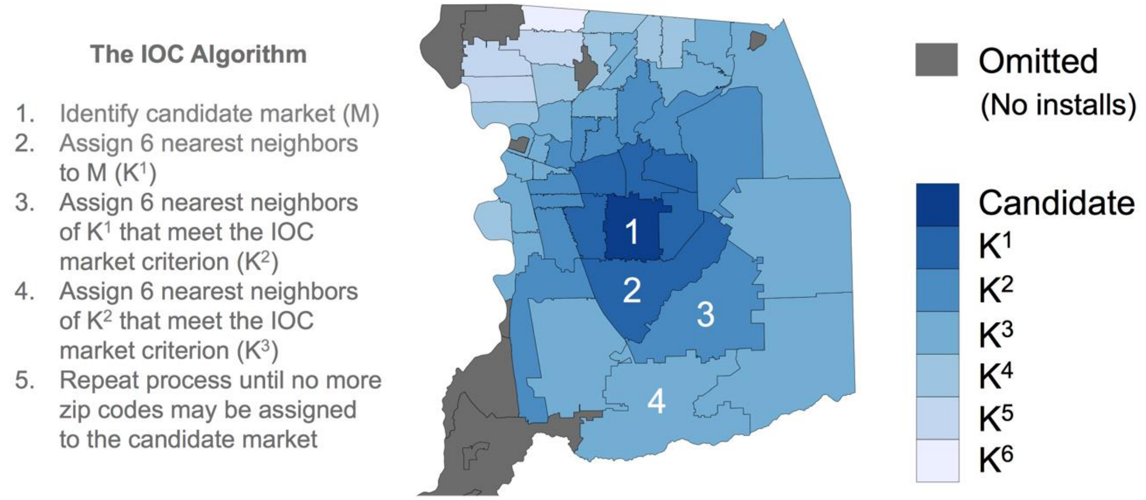

Figure 17. Visualization of the IOC algorithm in Sacramento County. Source: O'Shaughnessy, Nemet et al. (2016)

Several specifications of the IOC algorithm were explored and spatial statistical analyses were performed to study the appropriateness of various solar PV market definitions. In general, the IOC market definition approach results in more solar PV market granularity than a county-level approach. Further, the analytical results suggest that county-level market definitions may be overly broad in terms of studying solar PV market dynamics (O'Shaughnessy, Nemet et al. 2016).

We use the methodology developed in O'Shaughnessy, Nemet et al. (2016) to define markets for the current study. We use a critical IOC value of 0.25 applied to zip codes. In other words, a candidate market zip code is identified then nearby zip codes are assigned to the candidate market if their shared IOC is greater than 0.25 (indicating that at least 50% of the installers in one zip code are also present in the other zip code). We use a spatial smoothing process described in O'Shaughnessy et al. (2016) to group isolated zip codes into nearby markets.

The use of the installer-based market definition allows us to study local market dynamics with far more precision than would be possible under a county-level market definition. To illustrate, consider two installers "X" and "Y" that operated in Los Angeles County in 2013 and 2014. At the county level, both installers held similar market shares of below 4% of installs over this timeframe. However, the IOC market definition approach demonstrates that installers X and Y held significantly higher market shares in local pockets of Los Angeles County, with installer Y holding a 22.5% market share in one IOC-defined market in southeastern Los Angeles County (Figure 18). Applying the IOC market definition in the present study allows us to better understand the effects of local market shares below the county level.

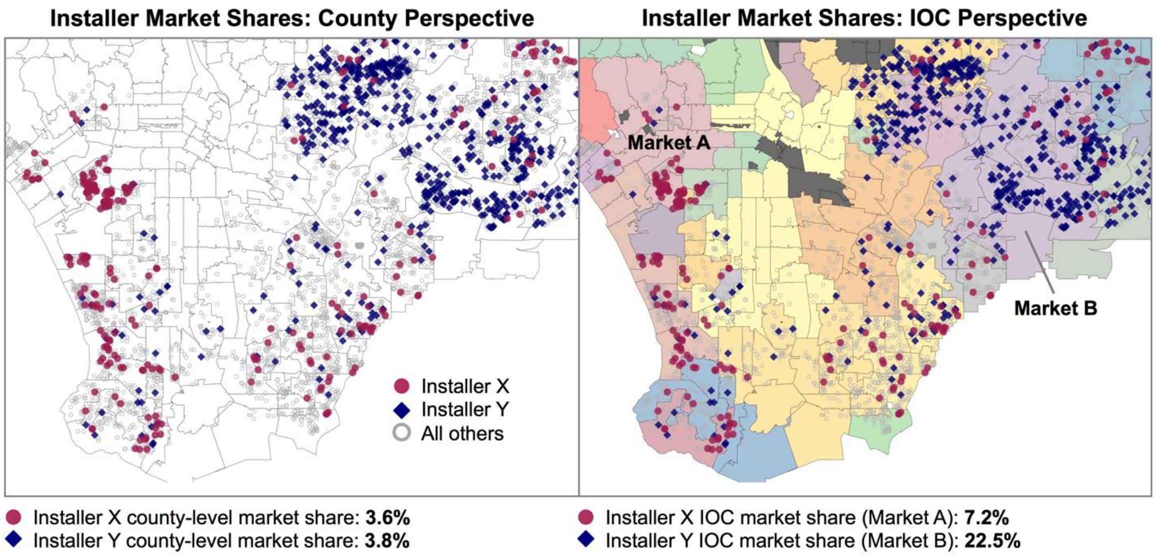

Figure 18. Depiction of installer market shares under a county-level market definition (left pane) and the IOC market definition (right pane). In the right pane, different colors correspond to separate markets. Source: O'Shaughnessy, Nemet et al. (2016)

A.5. Description of third-party quote data

A dataset of 647,728 quotes made from 2012 to 2016 for systems matching the characteristics of systems in our analysis (1-15 kW) was assembled from a third-party quote provider (Set A). The dataset was used as a proxy to determine where third-party quote provider services were generally available. Similarly to Tracking the Sun data, the quotes are heavily concentrated in California, but otherwise broadly distributed through every state included in our analysis. A second data set of 11,306 quotes from a second provider was also included (Set B). The TTS set includes PV installations in 8709 zip codes from 2008-14. The large set of quotes (A) included quotes in 5467 zip codes. The second smaller data set (B) included quotes from 331 additional zip codes, i.e., it added 6% to the set A quotes. 83% of the zips included in set B were also included in set A. The quotes variable was calculated as the number of quotes in a quarter in a county divided by the number of PV installation in that county quarter.

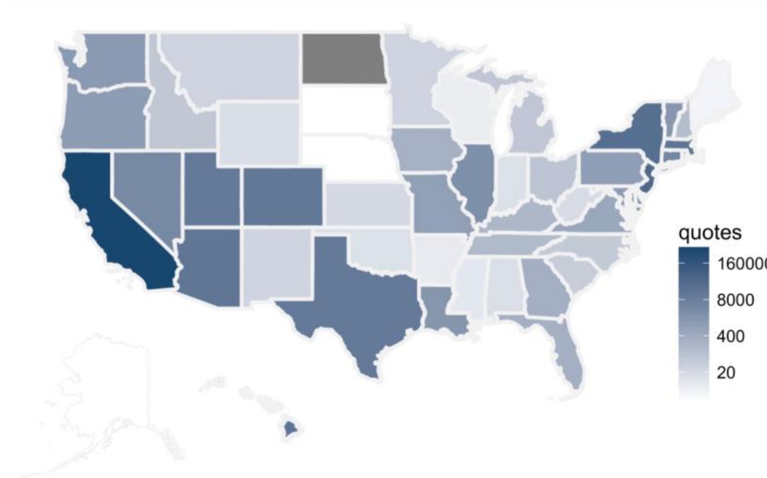

Figure 19. Quotes per state

References

Allen, J., R. Clark and J.-F. Houde (2014). "Price dispersion in mortgage markets." Journal of Industrial Economics **62**(3): 377-416.

Anania, G. and R. Nistico (2014). "Price dispersion and seller heterogeneity in retail food markets." Food Policy **44**: 190-201.

Atkin, D., A. Chaudhry, S. Chaudry, A. K. Khandelwal and E. Verhoogen (2015). "Mark-up and Cost Dispersion across Firms: Direct Evidence from Producer Surveys in Pakistan." National Bureau of Economic Research Working Paper Series **No. 20868**.

Bailey, J. P. (1998). Intermediation and electronic markets: Aggregation and pricing in Internet commerce. PhD, MIT.

Bakos, J. Y. (1997). "Reducing buyer search costs: Implications for electronic marketplaces." Management science **43**(12): 1676-1692.

Barbose, G. L. and N. R. Darghouth (2016). Tracking the Sun IX: The Installed Price of Residential and Non-Residential Photovoltaic Systems in the United States, Lawrence Berkeley National Laboratory.

Barron, J. M., B. A. Taylor and J. R. Umbeck (2004). "Number of sellers, average prices, and price dispersion." International Journal of Industrial Organization **22**(8,Äì9): 1041-1066.

Baye, M. R., J. Morgan and P. Scholten (2006). Information, search, and price dispersion. Handbook on economics and information systems. M. R. Baye. Amsterdam, Elsevier Press. **1**.

Besanko, D., U. Doraszelski and Y. Kryukov (2014). "The Economics of Predation: What Drives Pricing When There Is Learning-by-Doing?" American Economic Review **104**(3): 868-897.

Blackburn, G., C. Magee and V. Rai (2014). "Solar valuation and the modern utility's expansion into distributed generation." The Electricity Journal **27**(1): 18-32.

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

Borenstein, S. and N. L. Rose (1994). "Competition and Price Dispersion in the US Airline Industry." Journal of Political Economy **102**(4).

Brucks, M., V. A. Zeithaml and G. Naylor (2000). "Price and brand name as indicators of quality dimensions for consumer durables." Journal of the Academy of Marketing Science **28**(3): 359-374.

Brynjolfsson, E. and M. D. Smith (2000). "Frictionless commerce? A comparison of Internet and conventional retailers." Management Science **46**(4): 563-585.

Burdett, K. and K. L. Judd (1983). "Equilibrium price dispersion." Econometrica: Journal of the Econometric Society: 955-969.

Carlson, J. A. and R. P. McAfee (1983). "Discrete Equilibrium Price Dispersion." Journal of Political Economy **91**(3): 480-493.

Chakrabarty, D. and L. Kutlu (2014). "Competition and price dispersion in the airline markets." Applied Economics **46**(28): 3421-3436.

Chandra, A. and M. Tappata (2011). "Consumer search and dynamic price dispersion: an application to gasoline markets." The Rand Journal of Economics **42**(4): 681-704.

Chouinard, H. H. and J. M. Perloff (2007). "Gasoline price differences: Taxes, pollution regulations, mergers, market power, and market conditions." The BE Journal of Economic Analysis \& Policy **7**(1).

Clay, K., R. Krishnan, E. Wolff and D. Fernandes (2002). "Retail strategies on the web: Price and non–price competition in the online book industry." The Journal of Industrial Economics **50**(3): 351-367.

Dahlby, B. and D. S. West (1986). "Price Dispersion in an Automobile Insurance Market." Journal of Political Economy **94**(2): 418-438.

Dai, M., Q. Liu and K. Serfes (2014). "Is the effect of competition on price dispersion nonmonotonic? evidence from the us airline industry." Review of Economics and Statistics **96**(1): 161-170.

Davis, S. J., C. Grim, J. Haltiwanger and M. Streitwieser (2013). "Electricity Unit Value Prices and Purchase Quantities: US Manufacturing Plants, 1963–2000." Review of Economics and Statistics **95**(4): 1150-1165.

Diamond, P. A. (1971). "A model of price adjustment." Journal of economic theory **3**(2): 156-168.

Eckard, E. W. (2004). "The “law of one price” in 1901." Economic Inquiry **42**(1): 101-110.

Gerardi, K. S. and A. H. Shapiro (2009). "Does competition reduce price dispersion? New evidence from the airline industry." Journal of Political Economy **117**(1): 1-37.

Gillingham, K., H. Deng, R. H. Wiser, N. Darghouth, G. Nemet, G. L. Barbose, V. Rai and C. Dong (2016). "Deconstructing Solar Photovoltaic Pricing: The Role of Market Structure, Technology, and Policy." The Energy Journal **37**(3): 231-250.

Goldberg, P. K. and F. Verboven (2001). "The evolution of price dispersion in the European car market." The Review of Economic Studies **68**(4): 811-848.

Goldberg, P. K. and F. Verboven (2005). "Market integration and convergence to the Law of One Price: evidence from the European car market." Journal of International Economics **65**(1): 49-73.

Graziano, M. and K. Gillingham (2015). "Spatial Patterns of Solar Photovoltaic System Adoption: The Influence of Neighbors and the Built Environment." Journal of Economic Geography **15**: 815-839.

Imbs, J. M., H. Mumtaz, M. O. Ravn and H. Rey (2010). "One TV, One Price?" The Scandinavian Journal of Economics **112**(4): 753-781.

Jaeger, D. A. and K. Storchmann (2011). "Wine Retail Price Dispersion in the United States: Searching for Expensive Wines?" The American Economic Review **101**(3): 136-141.

Kano, K., T. Kano and K. Takechi (2013). "Exaggerated death of distance: Revisiting distance effects on regional price dispersions." Journal of International Economics **90**(2): 403-413.

Kutlu, L. (2015). "Limited memory consumers and price dispersion." Review of Industrial Organization **46**(4): 349-357.

Lach, S. (2002). "Existence and persistence of price dispersion: an empirical analysis." Review of Economics and Statistics **84**(3): 433-444.

Lee, J. H. (2014). "Gain From Consumers Information Searching and Price Dispersion." International Journal of Human Ecology **15**(2): 59-67.

Lin, Y.-C. and C.-M. Chen (2014). "Price And Price Dispersion On The Web: Evidence From The Online Book Industry Of Taiwan." Hitotsubashi Journal of Economics: 51-70.

Marvel, H. P. (1976). "The economics of information and retail gasoline price behavior: an empirical analysis." Journal of Political Economy **84**(5): 1033-1060.

Menzio, G. and N. Trachter (2015). "Equilibrium price dispersion with sequential search." Journal of Economic Theory **160**: 188-215.

Nelson, P. (1970). "Information and Consumer Behavior." Journal of Political Economy **78**(2): 311-329.

Nemet, G., E. O'Shaughnessy, R. H. Wiser, N. R. Darghouth, G. L. Barbose, G. Kenneth and V. Rai (2016). Characteristics of Low-Priced Solar Photovoltaic Systems in the United States. Berkeley, CA, Lawrence Berkeley National Laboratory.

O'Shaughnessy, E., G. F. Nemet and N. Darghouth (2016). Using the Spatial Distribution of Installers to Define Solar Photovoltaic Markets, NREL (National Renewable Energy Laboratory (NREL), Golden, CO (United States)).

Pan, X., B. T. Ratchford and V. Shankar (2002). "Can price dispersion in online markets be explained by differences in e-tailer service quality?" Journal of the Academy of Marketing Science **30**(4): 433-445.

Pennerstorfer, D., P. Schmidt-Dengler, N. Schutz, C. Weiss and B. Yontcheva (2014). Information and Price Dispersion: Evidence from Retail Gasoline, Working Paper.

Perloff, J. M. and S. C. Salop (1985). "Equilibrium with product differentiation." The Review of Economic Studies **52**(1): 107-120.

Pratt, J. W., D. A. Wise and R. Zeckhauser (1979). "Price differences in almost competitive markets." The Quarterly Journal of Economics **93**(2): 189-211.

Rai, V. and A. L. Beck (2015). "Public perceptions and information gaps in solar energy in Texas." Environmental Research Letters **10**(7): 074011.

Rai, V., D. C. Reeves and R. Margolis (2016). "Overcoming barriers and uncertainties in the adoption of residential solar PV." Renewable Energy **89**: 498-505.

Rai, V. and S. A. Robinson (2013). "Effective information channels for reducing costs of environmentally-friendly technologies: evidence from residential PV markets." Environmental Research Letters **8**(1): 014044.

Ratchford, B. T. (2009). "5 Consumer search and pricing." handbook of Pricing Research in Marketing: 91.

Reinganum, J. F. (1979). "A simple model of equilibrium price dispersion." The Journal of Political Economy **87**(4): 851.

Roberts, M. J. and D. Supina (1996). "Output price, markups, and producer size." European Economic Review **40**(3-5): 909-921.

Salop, S. and J. Stiglitz (1977). "Bargains and ripoffs: A model of monopolistically competitive price dispersion." The Review of Economic Studies: 493-510.

Shepard, A. (1991). "Price discrimination and retail configuration." Journal of Political Economy: 30-53.

Sorensen, A. T. (2000). "Equilibrium price dispersion in retail markets for prescription drugs." Journal of Political Economy **108**(4): 833-850.

Spulber, D. F. (1995). "Bertrand competition when rivals' costs are unknown." The Journal of Industrial Economics: 1-11.

Stahl, D. O. (1989). "Oligopolistic pricing with sequential consumer search." The American Economic Review: 700-712.

Stigler, G. J. (1961). "The economics of information." The journal of political economy: 213-225.

Van Hoomissen, T. (1988). "Price dispersion and inflation: Evidence from Israel." Journal of Political Economy **96**(6): 1303-1314.

Van Nieuwerburgh, S. and P.-O. Weill (2010). "Why has house price dispersion gone up?" The Review of Economic Studies **77**(4): 1567-1606.

Varian, H. R. (1980). "A model of sales." The American Economic Review: 651-659.

Walsh, P. P. and C. Whelan (1999). "Modelling price dispersion as an outcome of competition in the Irish grocery market." Journal of Industrial Economics: 325-343.

Woodhouse, M., R. Jones-Albertus, D. Feldman, R. Fu, K. Horowitz, D. Chung, D. Jordan and S. Kurtz (2016). On the Path to SunShot: The Role of Advancements in Solar Photovoltaic Efficiency, Reliability, and Costs. Golden, CO (United States), National Renewable Energy Laboratory (NREL), .

Wu, K., J. Vassileva, Z. Noorian and Y. Zhao (2015). "How do you feel when you see a list of prices? The interplay among price dispersion, perceived risk and initial trust in Chinese C2C market." Journal of Retailing and Consumer Services **25**: 36-46.

Yilmazkuday, H. (2014). "Price dispersion across US districts of entry." Economics Letters **123**(3): 361-365.

Key Report Contacts

Gregory Nemet
University of Wisconsin
gnemet@lafollette.wisc.edu

Ryan H. Wiser
Lawrence Berkeley National Laboratory (LBNL)
rhwiser@lbl.gov

Download the Report

http://emp.lbl.gov/reports

Acknowledgments

This work was supported by the Office of Energy Efficiency and Renewable Energy (Solar Energy Technologies Office) of the U.S. Department of Energy under Contract numbers DE-AC02-05CH11231 (LBNL) and DE-AC36-08GO28308 (NREL). For supporting this work, we thank Elaine Ulrich, Odette Mucha, Ammar Qusaibaty, Joshua Huneycutt, Dan Boff and the entire DOE Solar Energy Technologies Office team. For assistance and reviewing earlier versions of this report, we also thank Charlie Gay, Andrew Lick, and Ammar Qusaibaty.

June 2017