Junior Management Science 7(3) (2022) 643-667

Junior Management Science

journal homepage: www.jums.academy

Demand Estimation for Solar Photovoltaics in the United States – An Instrumental Variable Approach

Veronique Clara Bukow

Technische Universität München

Abstract

Worldwide the demand for solar photovoltaics (PV) has increased significantly over the past decades. This was driven by a price reduction for solar PV systems. A two-stage least squares linear regression yields insights into the price sensitivity for residential customers in the U.S., and California in particular. The specification includes instrumental variables as well as fixed effects to account for the common issues of endogeneity and data heterogeneity in demand estimation problems, respectively. The variation in the sales tax rate on solar PV and the movements of polysilicon spot prices are used to instrumentalise PV price changes. The regression results imply an inelastic demand with a long-term price elasticity of -0.443, accounting for differences over state and time. Investigating price elasticities for various income groups shows that lower-income customers react more strongly to price changes compared to those with relatively high income (-0.521 vs. -0.195). Likewise, regions with lower population density are more sensitive to price changes (-0.473 vs. -0.338). Besides price, installation costs and technological efficiency majorly impact the system size installed. Results of this study can provide data-driven guidance to efficient policy design and pricing strategies.

Keywords: Price elasticity; solar photovoltaic; instrumental variables; demand estimation.

1. Introduction

1.1. Central Issue

> *"The path towards sustainable energy sources will be long and sometimes difficult. But America cannot resist this transition; we must lead it."*
>
> *Barack Obama, 2013*

Mounting greenhouse gas emissions (GHG) and global climate change put especially industrialized countries under pressure to act. In an attempt to reduce emissions and limit global warming, the development and deployment of renewable energy sources is increasing worldwide as well as in the United States (U.S.) (EIA, 2019c; IEA, 2019). Hence, the importance of and focus on solar photovoltaics (PV) as one source of renewable energy has risen continuously in the past years. According to the International Energy Agency (IEA), the technology is expected to be the main accelerator in renewable capacity growth from 2019 to 2024 (IEA, 2019). It is little surprising, therefore, that investors, governments, and researchers are taking interest in understanding which factors predominantly drive the solar PV demand. Especially in the U.S., one of the world's largest economies with many different regional characteristics and incentive policies, this issue is of vital importance for both firms and policymakers (Gillingham & Tsvetanov, 2019a). Naturally, the installation price and subsequent maintenance costs play an important role when considering product demand. Installed price[1] reductions have led to a significant rise in installed systems throughout the past decades (Barbose & Darghouth, 2019). However, how sensitive are residential investors to a change in prices? Which other factors are important when it comes to deciding how much capacity to invest in? And do the answers to these questions vary for different customer groups?

Relevant insights regarding this topic can be of tremendous importance, as they enable federal and state governments to design better tailored, more efficient policy incentives and regulations while producing firms can make

[1] Hereinafter, installed price refers to the total costs for solar PV installation, including hardware costs as well as soft costs (customer acquisition, system design, installation labour, permitting, inspection, etc.) and Balance-of-System costs (racking, wiring, etc.) (Barbose & Darghouth, 2019).

DOI: https://doi.org/10.5282/jums/v7i3pp643-667

informed decisions on product design, pricing, and market forecasts.

1.2. Research Aims

This study's goal is threefold, placing particular emphasis on the first of the three research questions:

(1) How high is the price elasticity of demand[2] for solar photovoltaic systems in the U.S.?
(2) Which other factors impact the installed system size, and in what way?
(3) Can differences be observed for distinct subgroups of the population?

The price for solar PV is probably easier to influence externally than other demand driving factors. Therefore, knowledge of the marginal impact of price changes can be a crucial lever to design better tailored policies and pricing in order to globally increase the demand for and thereby the share of solar energy generation.

To provide sound and data-driven answers to the key questions outlined above, this study focusses on an econometric approach to estimate demand elasticities. The model specification includes instrumental variables as well as regional and time fixed effects. This way, it addresses some of the main challenges in demand estimation, accounting for endogeneity in regressors and heterogeneity on a regional level and over time, respectively (Cui, 2018; Gillingham & Tsvetanov, 2019a). The analysis is mainly based on a subset of U.S. pooled cross-sectional data collected and pre-processed by the Lawrence Berkeley National Laboratory (NBNL) on distributed, grid-connected PV systems for residential as well as non-residential customer segments from 1998 to 2018. The data set contains key attributes of installed systems, including system size, installed price, received financial incentives, location, module technology, and efficiency. A subset of observations is used for the estimation. I supplement these data with information on production factors like polycrystalline silicon prices and installer labour wages, as well as on incentive programs, electricity prices, and income and tax levels[3].

The analysis aims to derive insights that help to provide tangible and actionable policy implications to promote economic investment in solar PV and maximise the benefits of political and commercial interventions. The estimates can further be used to calculate program effectiveness and assess social desirability by comparing the derived costs of carbon emission abatement associated with solar PV rebates and tax incentives to the estimated (social) cost of carbon emissions assumed by the U.S. government (Gillingham & Tsvetanov, 2019a). Also, looking beyond the scope of this study, identifying a valid method of demand estimation might be beneficial to a much wider range of estimation problems, especially for early-stage technologies (Gillingham & Tsvetanov, 2019a).

The remainder of the paper at hand is structured as follows. Section 2 provides a brief introduction to today's U.S. energy infrastructure in general and the prevalence of solar photovoltaics in particular, underlining the importance of the topic in the light of current global environmental challenges. It outlines current deployment and development of solar PV, and gives insights into policies as well as the political environment regarding the technology in the U.S. The following section 3 turns to the estimation of the solar PV demand curve. It sheds light onto insights gained in relatively scarce prevalent literature on the topic and describes the Instrumental Variable (IV) estimation methodology, its application to the problem at hand, relevant data, and the estimation results. Section 4 discusses the political and economic relevance and implications of the findings. Lastly, section 5 critically assesses the study, touches upon limitations, and draws a comprehensive conclusion, including an outlook on potential future research.

2. Solar Photovoltaics in the U.S.

2.1. U.S. Energy Infrastructure and Solar Capacity today

The United States, as the world's largest economy, have a substantial influence on the global energy consumption as well as its consequences. The energy mix in the U.S. is dominated by fossil fuels, with petroleum, coal, and natural gas making up more than 80% of the country's energy production in 2018, and fossil fuel consumption even having increased by 4% relative to previous year levels (EIA, 2019a, 2019b). As a result, the country generates about 15% of the global energy-related CO_2 emissions (Center for Sustainable Systems, 2019). According to the U.S. Environmental Protection Agency (EPA), U.S. GHG emissions, 80% of which are CO_2, have increased by 1.3% since 1990. Most of them stem from burning of fossil fuels in transportation and electricity generation (EPA, 2020).

Notwithstanding, clean energy sources have gained in importance (EIA, 2020a). From the 1990s onward, renewable sources other than hydropower and biomass started to take a share in the U.S. energy mix (EIA, 2020a). Both U.S. production and consumption from non-fossil energy sources reached record levels in 2019, constituting 20% of the states' total energy consumption (EIA, 2020b). A total of 19% of the U.S. electricity was generated from renewable resources that year, thereof 15% by solar power[4], crowding out less efficient or less ecological alternatives such as coal and oil (EIA, 2020a).

[2]The price elasticity of demand is the percentage change in quantity demanded caused by a one percent change in price, moving along the demand curve. The elasticity can be expressed as the slope of the relationship between the natural logarithm of quantity and price.

[3]An overview of selected variables can be found in Appendix A7, Table 18.

[4]One typically distinguishes between two types of solar power: solar thermal and solar photovoltaic (Khan & Arsalan, 2016). Solar thermal converts sunlight into heat which can subsequently be used for multiple purposes. Solar photovoltaic applications, on the other hand, directly generate electricity from sunlight, using a semiconductor technology (Burr, 2014; Singh,

Solar energy is abundant, inexhaustible, and amongst the cleanest sustainable energy source to date (Denholm & Margolis, 2007; Parida, Iniyan, & Goic, 2011), although negative externalities are evidently not absent, arising during fabrication, construction, and operation (Khan & Arsalan, 2016; Nugent & Sovacool, 2014; Raman, 2013).

Solar power is one of the fastest-growing sources of energy, both globally and in the United States. In 2018, solar energy accounted for about 2% of the total U.S. energy consumption, but exhibited a growth of 22% compared to 2017 levels, highlighting the strong focus and large potential of the technology (EIA, 2019c). The country's total installed solar PV capacity has reached over 81 GW[5] in Q1 2020, following extensive investment in the past years (Perea et al., 2020b). For comparison, global PV installations reached 627 GW by the end of 2019 (Feldman & Margolis, 2020). The U.S. installations in 2019 constitutes a 23% year-over-year increase and represent nearly 40% of the total new U.S. electricity generating capacity installed that year (Perea et al., 2020a). Today's capacity was expected to more than double by 2025 before forecasts declined moderately due to the impacts from the coronavirus pandemic, which will most probably cause less utility PV to be built in the coming years (Perea et al., 2020b). However, especially residential solar saw a record-high in capacity additions in 2019, while, on the other hand, non-residential PV growth declined slightly due to unfavourable policy changes in several states (Barbose & Darghouth, 2019; EIA, 2019a).

2.2. Characteristics of Installations

Solar PV installations in the U.S. vary across customer segments[6] in numerous aspects, including system size, efficiency, and module and inverter technology, constitute Barbose and Darghouth (2019) from a representative U.S. data set on PV installations[7]. Overall, systems grew in size, with a median capacity of 6.4 kW for residential and 47 kW for non-residential installations in 2018 (s. Appendix A1, Figure 1). Module efficiency is highest for residential applications, likely due to greater space constraints compared to non-residential sites (s. Appendix A1, Figure 2). The residential share of monocrystalline silicon modules compared to lower-quality polycrystalline silicon is largest (Barbose & Darghouth, 2019), while production nowadays focusses on even more efficient products such as monocrystalline p-type PERC and n-type PERT, also for large-utility-scale systems (Blakers, 2019; Burr, 2014; Platzer, 2012) (s. Appendix A1, Figure 3). Furthermore, efficiency-enhancing module-level power electronics[8] (MLPE) like microinverters or DC power optimizers are used particularly in residential installations as small roof-top systems are constrained most with regard to orientation and flexibility. On the contrary, ground-mounting and tracking[9] are more common for large non-residential and utility-scale installations. According to Barbose and Darghouth (2019), in 2018, only 3% of all residential systems are ground-mounted. Less than 1% can track the sun, even though many residential rooftop installations do not offer the flexibility to freely choose the panel orientation, and only slightly more than half of the systems were oriented southward in 2018. This might be explained by lower rooftop installation costs, higher market penetration, and the fact that systems became economically viable also without an optimal panel orientation or tracking functionality (Barbose & Darghouth, 2019).

2.3. Development of Installed Prices

Overall, renewable energy sources have become the lowest-cost sources of power in many countries (IRENA, 2019). As one of those, solar photovoltaic has made huge steps towards becoming a mature technology throughout the last decade. In the U.S., average PV prices fell by 50% between 2013 and 2018 (IRENA, 2019). This development can primarily be attributed to higher efficiency, lower module prices, and decreasing system costs and is expected to continue in the coming decades (IRENA, 2019). Capital costs of solar PV include hardware costs as well as Balance-of-System (BoS) and soft costs (Barbose & Darghouth, 2019; Elshurafa, Albardi, Bigerna, & Bollino, 2018). While hardware mainly refers to the PV module and inverter, BoS costs include racking and wiring as well as soft costs for customer acquisition, system design, permitting, and labour costs for installation and inspection (Barbose & Darghouth, 2019; Elshurafa et al., 2018). Unlike hardware costs, these BoS costs differ regionally due to their strong dependence on local wage rates, taxes, and competition (Elshurafa et al., 2018; Gillingham et al., 2016).

Barbose and Darghouth (2019) identify several key trends in prices prior to incentives, using U.S. data from 30 states in the past 20 years. In 2019, national median installed prices in the U.S. ranged from \$3.7/W over \$3.0/W to \$2.4/W for residential to small and large non-residential systems, respectively (s. Appendix A1, Figure 4). These persisting differences can mainly be attributed to higher economies

2013). These range between distributed small-scale residential to utility-scale power generation facilities. Solar PV is the more mature and commercially established technology (Khan & Arsalan, 2016).

[5]All energy is expressed in direct current (DC) units. Direct current describes the flow of energy into one direction only. All solar PV nowadays produce DC power (Zainudin & Mekhilef, 2010).

[6]Distributed PV comprises residential as well as non-residential rooftop PV installations of any size and ground-mounted systems of less than 7,000 kW. Non-residential systems are divided into small and large non-residential, with a threshold of 100 kW, in accordance with Barbose and Darghouth (2019).

[7]For a more detailed description, see section 3.4.1.

[8]Direct current power optimizers and microinverters are collectively referred to as module-level power electronics, or MLPE. They have replaced standard string inverters in the past years, 55% of all residential PV systems using some form of MLPE by 2014 (NREL, 2015). Both microinverters and power optimizers can monitor the performance of individual solar panels, rather than the solar panel system as a whole. They improve performance for solar panels by reducing shading losses and the impact of multiple roof planes (Deline, Meydbray, Donovan, & Forrest, 2012).

[9]Tracking is the technical ability of a system to flexibly change its orientation towards the sun compared to fixed-tilt systems (Barbose & Darghouth, 2019).

of scale for larger installations (Barbose & Darghouth, 2019). In the long term, installed prices fell due to the reduction in both hardware, BoS, and soft costs. Modules and inverters made up about 55% of these total cos reductions and fell most between 2008 and 2012 (Barbose & Darghouth, 2019). The remaining 45% can be attributed to reduced BoS costs. While for residential PV, the decrease was mainly driven by hardware cost, non-residential installers benefited to a major extent from reduced BoS and soft costs (Barbose & Darghouth, 2019). The decline in non-hardware costs cannot be linked to a single factor, but rather a changing market and policy environment as well as mechanical aspects. Regarding technical reasons for price reduction, drivers are twofold: On the one hand, installing larger systems on average reduced costs per watt as fixed costs for permitting and customer acquisition occur irrespective of installation size. On the other hand, hardware technologies have improved and – due to the extreme price deflation of PV in the past years – were able to push less efficient polycrystalline modules out of the market. Increased module efficiencies in turn lead to a disproportionately low increase in area-related costs like racking and wage costs for installation of a given capacity (Barbose & Darghouth, 2019). Overall, in 2019, prices continued their declining trend, though at a slower pace. This reduced marginal change is mainly due to lowered financial incentives and higher customer acquisition costs as most early adopters already installed solar PV. Also, cost reductions and efficiency gains become increasingly difficult to capture as the market matures (Barbose & Darghouth, 2019). With rising grid penetration, more significant cost reduction becomes necessary in order to make solar PV profitable for the remaining potential customers, especially in states exhibiting already high PV shares.

Looking at an absolute price level, Barbose and Darghouth (2019) and Gillingham et al. (2016) find that costs vary substantially across states (s. Appendix A2, Figure 9). Smaller markets are generally associated with higher prices but more significant cost reductions. Likewise, urban areas tend to show higher prices. According to Barbose and Darghouth (2019), even on an individual installer-level, median installed prices deviate substantially. They are significantly higher for systems with premium efficiency modules and MLPE as well as ground-mounted systems. Interestingly, tax-exempt customers, mostly non-residential, also exhibit higher prices on average.

Understanding the key drivers and characteristics of PV system price changes as well as distinct regional differences is essential for the following estimation and interpretation of the results, especially because changes in the installed price need to be approximated through variation in correlated (instrumental) variables.

2.4. Incentive Policies and Political Environment

The United States implemented several utility, state, and federal incentive mechanisms to foster growth of solar PV (Barbose & Darghouth, 2019; Consumer Energy Alliance, 2018; Platzer, 2012; Shrimali & Jenner, 2013). Mostly, these comprised of cash incentives through the state or utility PV programs in the form of rebates or grants, performance-based incentives (PBIs), and federal and state investment tax credits (ITC) for both distributed and utility-scale systems. Furthermore, tax exemptions, rights for accelerated depreciation, retail rate net metering, a market for solar renewable energy certificates (SRECs), and non-rebate marketing programs had been established, some of which still exist (Barbose & Darghouth, 2019; Gillingham & Bollinger, 2019; Shrimali & Jenner, 2013). Tax exemptions for schools, governments, and non-profits result in a disproportionately large share of reduced tax costs for non-residential customers, find Barbose and Darghouth (2019). ITC supports investment since 2005 by providing a dollar-for-dollar reduction in tax liabilities (Platzer, 2012; SEIA, 2020). Distributed as well as large-scale utility installations are eligible to tax credits of up to 30% of purchase and installation costs. In 2015, the ITC was extended to 2021 and 2022 for residential and commercial applications, respectively, but the credit value will start to decline in 2020 (SEIA, 2020). However, not only ITC but also cash incentives have decreased throughout the past decade (Barbose & Darghouth, 2019). Many regions plan to phase out local government incentive programs in the coming years or have already done so (Gillingham & Tsvetanov, 2019a). At peak times providing cash incentives of $4-6/W, those expired in most larger markets or diminished to less than $0.5/W on average. However, other forms of financial support like SRECs have become more profitable and thus more prevalent (Barbose & Darghouth, 2019). On a state level, Renewable Portfolio Standards (RPS) encourage investment in green technologies by requiring certain contributions of renewable sources to the state's energy generation (Yin & Powers, 2010). SRECs markets facilitate compliance with these obligations. PV system owners have the possibility to sell SRECs from their installations, offering indirect cash incentives. Several states prefer these generation-based incentives over standard offer-based ones (Barbose & Darghouth, 2019). If one can assume that demand side subsidies are directly considered in the purchase decision and supply side subsidies are passed on (at least partly) to consumers (Dong, Wiser, & Rai, 2018; Gillingham & Tsvetanov, 2019a), then the changes in incentives directly impact installed prices and consequently solar PV demand[10].

In spite of numerous programs put in place to promote solar power usage, American energy policy has changed drastically under the current administration. It reversed several former agreements and targets and has, for instance, replaced the Clean Power Plan with a weaker Affordable Clean Energy Rule (Keyes et al., 2019; Krupnick et al., 2018). As the U.S. are one of the largest exporters of crude oil and natural gas (EIA, 2019a), economic interests seem to oppose the

[10]Gillingham and Tsvetanov (2019a) find a pass-through rate of cost reductions from the installer to the consumer of 84%, Dong et al. (2018) even find nearly 100% incentives pass-through for residential customers in California, implying a competitive market and well-operating subsidy programs from a pass-through perspective.

goal of deep decarbonisation and economy-wide emission reductions of 80% by 2050, stated in the U.S. Nationally Determined Contribution (NDC) of the Paris Agreement (Dennis, 2019; United Nations, 2016). On November 4^{th}, 2019, the U.S. government officially announced to withdraw from the Paris Agreement, although their pledge remains legally valid until November 2020 (Dennis, 2019; Zhang, Dai, Lai, & Wang, 2017). As a response, some states formed subnational climate initiatives and continue to strive for the previously set goals (Center for Climate and Energy Solutions, 2019; Friedman, 2019).

Furthermore, the current administration made some changes regarding solar energy in particular. In January 2018, the U.S. government placed a 'Section 201 Solar Tariff' on imported solar cells and modules, rendering investment more expensive – especially for utility-scale applications, as hardware costs increase (SEIA, 2019). These frequent policy and price changes make it an even more pressing matter to understand their effective impact on the demand for more renewable energy sources like solar photovoltaics.

3. Estimation of the Demand Curve

3.1. Evidence on Price Elasticity of Demand

So far, existing research on the demand for solar photovoltaic systems is very limited. Most research rather focusses on price elasticity of electricity demand than on the demand for the generation technology itself (Bernstein & Griffin, 2006; Mewton & Cacho, 2011; Miller & Alberini, 2016). Also, as solar PV can still be considered a maturing technology (Khan & Arsalan, 2016; van der Hulst et al., 2020), demand and supply conditions are constantly changing, making it harder to capture influences that remain valid over time. Recently, Gillingham and Tsvetanov (2019a) were the first to simultaneously address three main empirical challenges in estimating the demand for residential solar PV: price endogeneity, unobserved geographic heterogeneity, and excess zeros in the outcome variable with count data. Using panel data[11] on Census block level from Connecticut on the count of annual solar PV systems installed, Gillingham and Tsvetanov (2019a) account for heterogeneity in block group-specific characteristics by including geographic fixed effects and year dummies (Wooldridge, 2005). Furthermore, the address the issue of excess zeros in count data by applying a two-stage Poisson hurdle model[12] consisting of a logit regression with a control function and a truncated Poisson estimated by a General Method of Moments estimator. They include local roofing contractor wage rates and state incentives for PV systems as instrumental variables to eliminate the endogeneity in the price regressor. Their results suggest that residential consumers are relatively price insensitive (−0.65), meaning that the demand decreases less than proportional to the price increase.

Cui (2018) takes a slightly different approach to estimate both demand and supply functions of rooftop solar panels in California using data from the California Solar Initiatives rebate program. Like Gillingham and Tsvetanov (2019a), she estimates a hurdle model with count data to account for zeros in installation numbers aggregated by zip code and month. Likewise, Cui (2018) assumes a two-part non-linear model and uses a control function instead of a two-stage least squares estimator to account for endogeneity. Employing changes in rebates as exogenous variable to estimate supply and demand function simultaneously, she finds very different results compared to Gillingham and Tsvetanov (2019a), with a demand elasticity of -3.824 and supply elasticity of 5.572. She also specifies one model estimating the system size instead of installation count but does not obtain significant estimates. According to Cui (2018), customers are highly price sensitive, wherefore rebates are a very effective way to promote PV adoption. Cui (2018) further states that elasticity is not constant, but that consumers and sellers get more price inelastic as prices decrease. Besides the different model specifications, regional characteristics might to some degree provide explanations for the deviating results of these two studies. Gillingham and Tsvetanov (2019b) state that, unlike in Connecticut, the phase-out of subsidies could be anticipated beforehand in California, probably impacting the timing of investment decisions. Another aspect leading to differing estimates could be the limited sample of rebate installations as well as the slightly less granular assessment on zip code rather than Census block level.

Exploiting the changes in rebate rates for residential systems in California, Hughes and Podolefsky (2015) use a reduced form equation to estimate the number of installations. They find relatively high rebate elasticities of about -1.2, accounting for mean and utility specific unobservable characteristics that affect PV adoption and vary over time. Rogers and Sexton (2014) conclude that rebate elasticities are slightly lower, estimating a reduced form rebate elasticity of -0.4 for California.

Overall, the need for further research becomes apparent, as prevalent insights are both divergent and scarce. Numerous research designs are employed, ranging from different predictors over various forms of model specifications and estimation methodologies. Therefore, there is no consensus on typical demand curve characteristics and elasticities for solar photovoltaics so far. To my knowledge, there is no study yet providing substantial insight on the price elasticity of solar photovoltaic installations' system size installed, as will be the focus of this study.

3.2. Methodology and Research Design

3.2.1. Issues in demand estimation

Estimating demand and supply curves and their factor elasticities inherently poses the issue of simultaneous causal-

[11] Panel data are multidimensional data that include measurements pooled over space and time. They are a combination of cross-section and time series data (Baltagi, 2008).

[12] Hurdle models are motivated by sequential decision-making. The can represent the process of first deciding whether to buy or not and secondly deciding on the (positive) quantity to buy (Gillingham & Tsvetanov, 2019a).

ity[13] because the observed data of prices and quantities represent a set of market equilibria where supply equals demand (Angrist & Krueger, 2001). The price of a good influences its quantity bought and vice versa. An ordinary least squares (OLS) regression is incapable of isolating the effect of a price increase on one of the two curves, making alternative methods of estimation indispensable (Stock & Watson, 2020). A general linear model of solar PV demand would be specified as follows (Eq. (1) and (2)):

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} + u_i \tag{1}$$

Or, in matrix algebra

$$\boldsymbol{Y} = \boldsymbol{\beta X} + \boldsymbol{u} \tag{2}$$

For $i = 1, \ldots, n$ observations, let Y denote the dependent variable to be estimated: the size measured in watts of a solar PV system installed. X represents the k demand shifters or regressors incorporated in the model to estimate Y. The βs measure the effect size of the respective variables and are the coefficients of interest. The demand function is assumed to be linear in its parameters here.

A standard OLS approach to estimate the demand curve (Eq. (1)) makes several assumptions. One of them is that regressors $X_{1i}, \ldots, X_{ki}$ and error term u_i are uncorrelated, i.e., the conditional expectation of the error given the regressors is zero ($E[u_i | X_i] = 0$). Those regressors are called exogenous. In other words, there is no unobserved variable that is correlated with X and simultaneously changes Y. If this does not hold true, resulting OLS estimates are inconsistent even for large samples and other methods of estimation are necessary (Stock & Watson, 2020). This is where one challenge arises in this study. As for every supply-demand problem, the above-stated assumption of exogenous regressors is violated. An OLS estimation is biased because the simultaneous causality induces a correlation between the price regressor and the unobservable error term. Thus, how can this demand function be estimated consistently with the present endogeneity in prices? First, I define the model more precisely, separating endogenous and exogenous regressors:

$$\boldsymbol{Y} = \boldsymbol{\beta X} + \boldsymbol{\gamma W} + \boldsymbol{u} \tag{3}$$

Y still denotes the system size. The right-hand side combines its determining factors as well as the error term u. X now represents the endogenous regressors correlated with u while W denotes the truly exogenous regressors which are not correlated with u. In this estimation, only price per watt is assumed to be endogenous. With an endogenous regressor X the estimate of β will be incorrect (Stock & Watson, 2020). It captures both the effect of independent changes in X as well as changes in the error u due to simultaneous changes in Y associated with X.

There are several possible solutions to obtain consistent estimates in the case of simultaneous equations (Wooldridge, 2015). One, and probably the most frequently used one, is an Instrumental Variable approach, first employed by Philip Wright in 1928 (Angrist & Pischke, 2008; Wright, 1928).

3.2.2. Theory of instrumental variable estimation

The basic idea behind an IV estimation is to eliminate any correlation of the endogenous regressors X with the error term u by finding other variables that can be used in the regression in their stead. These variables are called instruments. Demand and supply estimation problems were the first applications of instrumental variables, initially called 'curve shifters' (Angrist & Krueger, 2001). For demand estimation, these curve shifters are used to trace out the slope of the curve by an exogenous variation in the supply, modifying cost conditions without affecting demand conditions (Angrist & Pischke, 2008; Stock & Watson, 2020).

An instrument, let it be Z, needs to satisfy two conditions of validity in order to produce meaningful results (Stock & Watson, 2020): Firstly, it must be *relevant*, i.e., highly correlated with the endogenous regressor to be replaced in order to serve as a good proxy variable ($Cor(Z, X) \neq 0$). The more variation in X can be explained by variation in the instrument, the more information will be sustained in the IV regression. If an instrument explains only a minor part, it is called weak (Stock & Watson, 2020). In this case, a weak instrument could be the price for an input factor that accounts only for a very small share of the final price or – with heterogeneous outcomes – influences only few observations, like the price for a rare module technology. Secondly and equally important, an instrument must be *exogenous*. This means that it must not affect the left-hand side of the equation – the variable to be estimated ($Cor(Z, u) = 0$). The variation in X that is related to Z is not related to u, neither through a direct effect of Z on Y nor through a variable that is not included in the model but is causally linked to both Y and Z (Stock & Watson, 2020). Given these assumptions hold true, the exogenous part of the variation in the endogenous regressor X can be isolated via changes in Z and can subsequently be used for a consistent estimation of Y (Angrist & Pischke, 2008; Stock & Watson, 2020). For this to work, one endogenous regressor must have at least one but can have several instruments. An IV model is called overidentified if the number of instruments exceeds the number of endogenous variables (Stock & Watson, 2020). Otherwise, it is exactly identified. Both relevance and exogeneity of instruments can be tested statistically, at least for overidentified models, e.g., by using the first stage F-statistic and the test on overidentifying restrictions (J-Statistic), respectively (Stock & Watson, 2020).

According to Angrist and Krueger (2001), the most efficient way to obtain estimates in an IV regression, especially when using numerous instruments, is the two-stage least squares (TSLS) approach. It builds upon two OLS regressions run in a row to compute the TSLS estimators.

(1) In the first stage, the part in the variation in X that is uncorrelated with the error u is isolated by regressing the

[13] Simultaneous causality means that "causality runs 'backward' from Y to X as well as 'forward' from X to Y" (Stock & Watson, 2020, p. 428). It is one cause for endogeneity in regression problems.

endogenous variable X on the instruments Z and all further exogenous variables W. X is split into two components: one that is correlated with the error and one that is not. For each endogenous regressor – in this study only the price –the reduced form equation (4) needs to be estimated by OLS (Stock & Watson, 2020):

$$\boldsymbol{X} = \pi \boldsymbol{Z} + \boldsymbol{\delta} \boldsymbol{W} + \boldsymbol{v} \tag{4}$$

(2) In the second stage of TSLS, the idea is to estimate fitted values $\hat{X}$ for all observations building on the first stage results and subsequently use these fitted values $\hat{X}$ instead of the original values X for estimation of the actual model specification. The original equation (3) is estimated again by OLS, only that X is replaced by $\hat{X}$:

$$\boldsymbol{Y} = \boldsymbol{\beta} \widehat{\boldsymbol{X}} + \boldsymbol{\gamma} \boldsymbol{W} + \widetilde{\boldsymbol{u}} \tag{5}$$

This regression provides consistent estimates $\widehat{\beta}^{TSLS}$, converging asymptotically towards the true parameter as the sample size increases (Stock & Watson, 2020). Consequently, researchers should work with significant sample sizes when applying an IV approach. However, if the explanatory variable is in fact not endogenous, both TSLS and OLS estimator are consistent, but the latter is more efficient. Therefore, it is important to ensure the presence of endogeneity, for instance by using the Durbin-Wu-Hausman test (Baum, Schaffer, & Stillman, 2007).

3.2.3. Potential issues in instrumental variable estimation

Like many other statistical models, an instrumental variable approach makes some model assumptions that need to be satisfied in order to obtain valid estimates (Stock & Watson, 2020). For IV models, these assumptions are modifications of the OLS assumptions for causal inference outlined in Stock and Watson (2020). In practice it is often very hard to meet all the requirements. Therefore, their validity in this application is discussed in Appendix B1.

More generally, Angrist and Krueger (2001) point out that IVs can solve the first-order problem to remove omitted variable bias[14] (OVB) only for a well-defined population. With heterogeneous responses, not every single observation can be explained by variation in the instruments as they only capture part of the true variation in prices. For example, where module costs do not differ significantly but mainly other hardware, BoS, or soft costs are drivers of price changes, an instrument shifting module prices will not accurately depict the price variation. Including several instruments can possibly counteract this to some extent. However, due to the bias-variance trade-off (James, Witten, Hastie, & Tibshirani, 2013), using more instruments might increase the variance of the estimators (Angrist & Krueger, 2001).

Lastly, another pitfall in IV estimation mentioned by Angrist and Krueger (2001) are functional form issues for both stage estimations. They emphasise that in a TSLS estimation procedure, the consistency of the final estimates $\widehat{\beta}^{TSLS}$ does not depend on the correct functional form of the first stage regression (Kelejian, 1971). Therefore, I estimate a linear regression for the first stage as a more complex non-linear model does not generate consistent estimates unless the fit is exactly right (Angrist & Krueger, 2001).

3.3. Application to Solar PV Demand Estimation

3.3.1. Specifying a multiple log-log linear regression model

The difficulties of demand estimation problems as well as one possible solution to solve them have been introduced in the previous sections. In the following, I apply this to the estimation of solar PV price elasticity. The preferred model[15] specification in this study is a multiple linear regression with log-transformed continuous outcome and predictor variables, including time and regional fixed effects on state and year level, controlling for several potentially confounding variables, and instrumenting for price. It is specified as follows:

$$\log(\boldsymbol{Y}) = \boldsymbol{\beta} \log(\boldsymbol{X}) + \boldsymbol{\gamma} \log(\boldsymbol{W}) + \boldsymbol{\alpha} + \boldsymbol{\mu} + \boldsymbol{u} \tag{6}$$

Y is the system size installed in watts, X represents all explanatory variables (exogenous and endogenous), W are control variables, and α and μ are fixed effects. The primary objective is to assess the constant elasticity of the system size installed with respect to the explanatory variables. Therefore, I estimate a log-log additive linear model (Eq. (6)). Its coefficients can directly be interpreted as an expected percentage change in system size given a regressor increases by one percent (Benoit, 2011). Rather than focussing on absolute differences, I estimate the relative change, the elasticity. If the price elasticity is constant, a percentage increase in price can be expected to cause a proportionate change in PV size demand over a wide range of prices. Thus, assuming a constant elasticity implies that absolute changes in system size may differ depending on the former level. This seems reasonable because a price increase by 1% when prices are low might well induce a shift in system size installed that is different from the shift caused when initial prices are relatively high. This means that a log transformation natively handles non-linear relationships between system size and independent variables (Benoit, 2011; James et al., 2013). For the data used here, regression plots show that the relationship between system size and the explanatory variables after log-transformation is at least slightly more linear (s. Appendix A4, Figure 18 to Figure 21). Apart from simplifying the functional form, log-transformed variables often follow an approximately normal distribution for otherwise skewed variables. In this case, several variables are right skewed

[14]"If the regressor [. . .] is correlated with a variable that has been omitted from the analysis [. . .] and that determines, in part, the dependent variable [. . .], then the OLS estimator will have omitted variable bias. Omitted variable bias occurs when two conditions are true: (1) the omitted variable is correlated with the included regressor and (2) the omitted variable is a determinant of the dependent variable." (p. 212 Stock & Watson, 2020).

[15]Hereinafter, preferred and final model synonymously refer to the model selected as the best model after conducting diagnostic and model validity tests as well as plausibility considerations.

(s. Appendix A4, Figure 22 and Figure 23), wherefore a transformation seems appropriate in order to better satisfy the assumptions of model linearity and normal distribution of the errors.

The linear additive model of the logs holds further advantages over other specifications. It can easily incorporate fixed effects to account for heterogeneity and instrumental variables to eliminate endogeneity (Gillingham & Tsvetanov, 2019a). Furthermore, its coefficients allow a straightforward interpretation, making it very popular with many researchers.

3.3.2. Deciding on the dependent variable

As described in section 3.1 above, most former research estimates demand models with installation count instead of system size as dependent variable (e.g., Cui, 2018; Gillingham & Tsvetanov, 2019a; Hughes & Podolefsky, 2015). Rather than determining to what extent price changes influence the system size installed they assess the propensity of adoption, meaning the decision to invest in solar PV, measured by the number of installed systems in each area. Unlike this study, those consequently also consider a zero-realisation with the decision not to invest by using a two-stage hurdle model. Thereby, they account for the fact that changes in demand factors can incite markets to change from zero to nonzero quantities, and vice versa. However, in order to use count data, a very broad and high-quality data coverage is necessary such that a count of zero installations can be attributed to the decision not to buy rather than missing data. Additionally, in California, subsidies were phased out in a way that depended on the total amount of installed PV capacity, allowing consumers and firms to reasonably anticipate the timing of subsidy declines and leading to bundled installations shortly beforehand which would have to be accounted for when using count data (Gillingham & Tsvetanov, 2019b).

Other research applies a dynamic discrete choice[16] approach (e.g., Bollinger & Gillingham, 2019; Burr, 2014), which might seem reasonable since most people would only install a system once. However, Gillingham and Tsvetanov (2019b) provide evidence that the investment in solar PV can more often be treated like normal purchases rather than a "buy-or-wait" decision. Furthermore, as more capacity can be added later, the decision on the size of the solar PV system is not necessarily a discrete choice. For those reasons, this study takes system size as the dependent variable to be explained.

3.3.3. Identifying suitable instrumental variables

Knowing of the endogeneity issue in the estimation, how can the IV approach be applied to consistently estimate the model elasticities β? The first step is to find suitable instruments for the installed price per watt. Generally, the choice of good instruments first and foremost relies on a profound understanding of the economic mechanism behind the relationship of interest (Angrist & Krueger, 2001; Angrist & Pischke, 2008). This leads to plausible and more intuitive results compared to abstract theoretical models based on hard-to-verify assumptions about certain distributions and relationships, so Angrist and Krueger (2001). To trace out the demand curve, one or more instruments that impact the supply of PV modules but not their demand have to be found. A typical supply-side instrument shifts costs of sales or production as these can be expected to affect supply without impacting demand (Angrist & Krueger, 2001; MacKay & Miller, 2019). Although statistical tests can assist to evaluate the relevance and exogeneity of instruments, it is useful to think about whether a chosen instrument plausibly satisfies these conditions (Stock & Watson, 2020). In complex demand models, especially the exogeneity of instruments can be challenging to assess. In this study, the preferred model employs two instrumental variables to approximate changes in the price for solar PV: Polycrystalline silicon prices and the sales tax rate on solar PV installations.

The first instrument, price quotes for polycrystalline silicon or 'polysilicon', aims to capture the variation in input factor costs. Polysilicon is the main raw material used for PV module production (Woodhouse, Smith, Ramdas, & Margolis, 2019). According to Woodhouse et al. (2019), crystalline-silicon made up about 90% of all PV production in 2014, increasing to over 95% in the subsequent years. Polysilicon is the basis for production of both multi- and monocrystalline silicon ingots of different purity and efficiency levels, which are then processed to wafers, manufactured into cells, and eventually fabricated into entire PV modules. Therefore, nearly all PV modules installed in the market in the past years used polysilicon as one factor of production. Moreover, the installed price of solar PV is to about 55% determined by hardware component costs (e.g., module, inverter) and to 45% by BoS and soft costs (e.g., installation labour, acquisition cost, system design, permit and inspection, installer margins, loan-related fees). As the following estimation considers only residential systems, for which the price decline in the past years was mainly driven by hardware cost reduction, instrumenting price changes via variation in those costs seems reasonable (Barbose & Darghouth, 2019). Therefore, input factor prices can be assumed to represent a *relevant* instrument. Further, as it seems plausible to assume that a residential customer's demand does not directly rely on the level of the price for polysilicon, these should be *exogenous* and influence demand only indirectly through the installed price. Even though there are possibly several relevant variables omitted from the model, these are similarly unlikely to be related to global polysilicon prices and will, therefore, not cause a correlation of this instrument with the error. Notwithstanding, a potential limitation of the power of the polysilicon prices as an instrument could be the fact that these are relevant only at the very beginning of a complex and costly production chain, making up less than 10% of the production costs of monocrystalline PV modules in 2018 (Woodhouse et

[16] A dynamic discrete choice model estimates the decision of a forward-looking agent over a finite number of options who is taking the utility of future alternatives into account (Heckman & Navarro, 2007).

al., 2019). Consequently, final module prices could differ significantly reflecting costs of subsequent production steps or further input factors even for initially equal polysilicon input costs, leading to a relatively weak instrument. This would imply that further production cost factors for PV modules need to be incorporated as instruments in the first stage estimation to approximate the price development more precisely. However, further data on hardware production costs are limited and soft costs are rather hard to quantify as they differ severely, depending on changing market and policy environments (Barbose & Darghouth, 2019). To account for a part of cost fluctuations for different modules and inverters, I include dummies indicating the type of module technology as well as MLPE in the model, where this information is available.

The second variable to instrumentalise installed prices is the sales tax rate levied on the hardware costs of installed systems, again assumed to be 55% of the total installed price. Tax rates are a popular instrument for price changes (Frondel & Vance, 2013; Stock & Watson, 2020). For one thing, they can be assumed to be *relevant*, as after-tax sales prices are adjusted to incorporate changes in taxes and often make up quite a noticeable part of the price for the final customer. We would assume prices to increase alongside rising sales tax rates, expecting producers to pass on at least part of the additional costs to consumers. The question whether tax rates can be considered *exogenous* is somewhat harder to answer. On the one hand, a change in tax rates can plausibly be assumed to impact consumer demand solely through the adjustment in prices and not directly through the mere fact that tax rates changed. However, unlike polysilicon prices, tax rates are more likely to be related to some of the omitted variables in the error term. While unobserved variables like hours of sunshine and irradiance are unrelated to setting tax rates, electricity prices may change alongside tax rates and simultaneously influence the amount of solar PV invested in. Also, incentive payments could be linked to tax policy, because general sales tax as well as many incentive programs are determined by state governments through political and financial considerations (Tax Policy Center, 2020). If these are not entirely captured by fixed effects included in the model, this might lead to inconsistent TSLS estimates. For this reason, the exogeneity test result will be particularly relevant here. In this study, the sales tax rate includes any potential tax exemptions granted to solar PV investors through the state government (Barbose & Darghouth, 2019; Shrimali & Jenner, 2013) which might lead to other-than-expected coefficients in the first stage linking tax rate and price per watt, if tax exemptions are primarily granted where prices are higher.

Apart from the instruments described in more detail above, I estimate specifications using further potential supply shifters given in section 3.4.2.2. However, mostly due to missing data and limited granularity, these did not yield meaningful estimates.

3.3.4. Defining relevant explanatory variables and controls

Other than price, further aspects determine the size of a solar photovoltaic system to be installed. In this estimation, I include information on the module efficiency assessing the energy conversion efficiency of the modules, and dummies indicating ground-mounting, tracking functionality, and the fact whether the system is installed as retrofit on an existing house or during the construction of a new building. Apart from these, data on the kind of module technology and MLPE are incorporated in the model. Module technologies are grouped into polysilicon, monocrystalline silicon, or other technologies. As MLPE categories I consider microinverters, DC optimizers, or no power electronics.

Naturally, there might be a large range of other variables that could turn out relevant determinants of the system size installed. However, as for the decision to invest in a good, especially a more complex one like a solar PV system, investors take many different variables into account, including the data on all these will probably be unattainable. This is no major cause for concern if the omitted factor is not correlated with any variable included in the model. However, if it is, this variable's coefficient estimate will be biased, reflecting not only its own effect but also that of the omitted variable (Stock & Watson, 2020). For this reason, control variables, though their coefficients are not of primary interest and might not have a sound causal interpretation, need to be incorporated in a model if their absence would otherwise cause OVB (Angrist & Pischke, 2008; Stock & Watson, 2020). In the final model, I include information on the number of households and the adjusted gross income per household on zip code level per year. A high level of income might positively relate to the propensity to invest in solar PV and its system size while the number of households, approximating the population and building density, could negatively impact the system size installed. In addition, they could in numerous ways be related to predictive variables included in the model, e.g., the efficiency, tracking equipment, or ground-mounting, assuming that richer people can afford to buy higher-class modules and higher population density requires the purchase of roof-mounted installations. By including the information in the model, I avoid confounding effects in the estimates of my coefficients of interest (Stock & Watson, 2020).

3.3.5. Accounting for heterogeneity through fixed effects

In a pooled data setting, fixed effects allow to eliminate OVB caused by factors that are not included in the model and which vary across states, but are constant over time (state fixed effects α_i), or which vary over time, but are constant across states (time fixed effects μ_t) (Borenstein, Hedges, Higgins, & Rothstein, 2010; Stock & Watson, 2020). They measure the residual difference across state and time, respectively, after accounting for all other factors in the model. Including an interaction term between state and time fixed effects allows the time effect to be different for individual states and vice versa (Stock & Watson, 2020). In this study, state fixed effects might capture e.g., weather conditions and hours of sunshine, established institutions, as well as prevalent fundamental culture and values. Time fixed effects, on the other hand, can account for aggregate time-varying demand shocks across states, U.S. economic and population growth, inflation, technological progress, and federal policy changes. One needs to bear in mind that time fixed effects might capture the impact of varying production factor input prices as well, if those are not included in the model. The interaction of both state and time fixed effects represents how state-specific factors change over time, e.g., state legislature and regulations, and subsidy and tax policies, as well as market conditions, prominent mindsets, trends, and acceptance of innovative technologies in a state. These interaction effects will only change estimations if enough data from different states and time frames is included, however. After data selection, this does not hold true for this study. The final model is estimated without interaction as there was no significant difference in the coefficient estimates.

3.4. Data

3.4.1. 'Tracking the Sun' data

Data set and structure

The original data of which a subset is used for estimation were collected and pre-processed in the 'Tracking the Sun' (TTS) data set by the Lawrence Berkeley National Laboratory (LBNL). Overall, the data cover PV systems installed in the U.S. from 1998 throughout 2018 with trends of the first half of 2019 (Barbose & Darghouth, 2019) (s. Appendix A2, Figure 5). The installation data were primarily reported to state agencies and utilities managing PV incentive programs, SREC registration, or interconnection processes in 30 states. The sample contains project-level information on nominal installed prices, system size, tax payment, financial incentives, module and inverter technology, efficiency, location, and further relevant characteristics of grid-connected, distributed solar PV systems. It excludes utility-scale systems. Barbose and Darghouth (2019) also dropped duplicate observations[17] and those where information on system size or installation date were missing. They corrected the data for obvious errors and standardised installer, module and inverter labels. Overall, 1,543,831 PV systems are included in the full sample, making up about 80% of all U.S. distributed PV systems installed throughout 2018 (Barbose & Darghouth, 2019). Most data stem from California, as solar PV is most prevalent here. In line with this, sample coverage tends to be weaker in small and mid-sized state markets.

Data selection

To improve accuracy, interpretability, and generalisability of the demand estimation, only a subset of the full TTS data sample[18] is used. First, I drop all observations with invalid price information. For this purpose, I exclude systems where installed prices are missing, which are about 23% of all observations. Among the remaining, I filter extreme outliers by including only installations with an installed price per watt between 1 and 10 USD. Those are prices between the 1^{st} and 99^{th} percentile in this data set, rounded inward to the nearest integer (s. Appendix A3, Figure 10). Likewise, all third-party owned systems (39.4% of full sample), systems with appraised price values (25.4% of full sample), self-installed systems (1.4% of full sample), and systems with battery backup (0.7% of full sample) are not considered in the estimation data set. All of these generally exhibit less representative price quotes (Barbose & Darghouth, 2019; Gillingham & Tsvetanov, 2019a). By excluding them, I prevent outliers and erroneous or unrepresentative values to distort the estimates. Thereafter, 613,157 observations (about 40% of full sample) remain in this price sample.

Apart from invalid price observations, further installations are removed from the estimation data set. To start with, I focus on residential systems only, because as shown above, installed prices and installation characteristics vary substantially across customer segments (Barbose & Darghouth, 2019) and residential systems make up almost 95% of the full sample and 93% of the price sample, respectively (s. Appendix A2, Figure 6). Also, price variations in residential PV systems are more likely to be captured by the instruments as their price decline was mainly driven by hardware cost reduction, whereas non-residential installers benefited to a major extent from lower BoS and soft costs (Barbose & Darghouth, 2019). Again, to account for outliers and in line with the customer categorisation by Barbose and Darghouth (2019), installations with system size bigger than 20,000 W are excluded (s. Appendix A3, Figure 11). As I assume efficiency to be non-negative, I also drop all observations with an efficiency of less than 0%. Those are less than 0.1% of the full sample and look like reporting errors rather than true values. Lastly, I consider only systems that were installed after 2009 (94% of full sample, 89% of price sample), as prior to this, data coverage in the sample is very limited and complementary instrument data on polysilicon prices is not available on a weekly basis.

All in all, these selection criteria lead to the estimation sample with 501,394 observations from 21 states and 9 years, representing 32.5% of the full TTS sample. Before being able to use this data for estimation, several data pre-processing steps need to be taken – namely the transformation and generation of predictors. Please refer to the Appendix B2 for a description of the most important steps.

[17] Few duplicate systems with redundant information were left in the sample. Those were deleted to ensure that one installation from one point in time is only included in the sample once.

[18] Hereinafter, full sample refers to the 'Tracking the Sun' data set as published by the LBNL (Barbose & Darghouth, 2019). Price sample refers to the sample left after applying all price-related selection criteria described. Estimation sample refers to the sample left after applying all selection criteria, also non-price related. The final sample is the sample left for model estimation after dropping all observations which have missing values in one or more of the included variables described in section 3.5.1.

Apart from the *Tracking the Sun* data set, further instrument as well as control data are included in the estimation. I elaborate on the more relevant in the following.

3.4.2. Complementary data

Polysilicon price data

Besides the tax rate, which is extracted directly from the TTS data set, I use the movement in price quotes for polysilicon as means to instrumentalise price changes in solar PV. Therefore, these need to be added to the estimation sample. There are different indices and data providers for polysilicon spot prices. Polysilicon production is concentrated almost entirely in China (Platzer, 2012; Woodhouse et al., 2019). Hence, figures referring to China commodity prices can be applied globally. I compare four different data sources providing weekly price quotes extracted from Bloomberg for China, international spot outside mainland China, companies regardless of region, and PVinsights poll prices for multiple contributors[19]. They are provided in USD per kg for comparable polysilicon purity from end of 2009 to mid-2019. However, only the PVinsights data are complete for the years 2010-2018. As those are correlated at over 99% with polysilicon spot prices from the other three indices and all but the data for China exhibit very similar median prices (s. Appendix A3, Figure 12 to Figure 14), I only use PVinsights poll prices in the subsequent estimations.

Despite the fact that in the estimation sample, polysilicon modules make up only about 35% of all observations, compared to 38% monocrystalline silicon, the price movement can be expected to be similar, as polysilicon is the fundamental input material for both module types (EIA, 2019d; Woodhouse et al., 2019).

Further instrument data

Although in the final estimation, only polysilicon price movement and changes in the sales tax rate are used to estimate installed PV price movements, I evaluated the quality of other instrument data, which I only outline briefly.

As next to hardware cost, BoS and soft costs for e.g., assembly, installation, and wiring make up about 45% of installed prices, PV costs can be expected to move in line with solar PV installer wages. Furthermore, changes are probably exogenous as those wages neither influence demand directly nor through other relevant variables, once accounting for income. Therefore, I consider the mean hourly and annual wage rates for U.S. solar PV installers estimated in the Occupational Employment Statistics Survey by the U.S. Bureau of Labor Statistics (BLS) for the years 2012 to 2018 in the estimation. Unfortunately, these data were not available on a more granular level regarding time frame and location, and did, therefore, not capture enough variation to serve as decent instruments for installed prices.

Another instrument I perceived promising is the number and scope of incentive programs, taken from the Database of State Incentives for Renewables and Efficiency (DSIRE). Because financial compensation reduces the effective installed price, it can be expected to provide a *relevant* instrument. However, those programs might reduce the marginal costs by also affecting other demand or supply conditions (Gillingham & Tsvetanov, 2019a), e.g., if there were a link between lower sales tax and offering incentives to promote consumption. Moreover, as the data available were highly incomplete with respect to start and end date as well as the size of financial benefit, I could only include the overall number of programs per state, which is by far not detailed enough to provide an accurate and valid approximation of the PV price changes.

Apart from these additional external data sources I also assess the absolute amount of sales tax paid per watt as well as the rebate or grant provided per watt contained in the TTS data set as instruments. I discard the former because it shows a substantially lower coefficient of determination[20] R^2 in the first stage compared to the tax rate and might additionally suffer from endogeneity due to the measurement in absolute dollar values. The latter could not be used because it did not provide enough data points for sufficient variation, with almost 10% missing values and over 50% zero rebate or grant. For those reasons, those instruments were not used in the final model.

Control variables

It is almost impossible to include all relevant predictors in a model, be it for the complexity and number of variables or for the unavailability of the required data. Nevertheless, some relevant information can be obtained from publicly available sources and included in the data in order to improve the estimation and prevent bias through omitted variables. An overview over the variables added to the estimation sample can be found in Appendix A3, Table 3. As outlined in section 3.3.4, I use yearly income and personal tax data gathered by the U.S. Internal Revenue Service (IRS) on zip code level for the years 2008 to 2017. More specifically, I join information on adjusted gross income, wages and salaries, the number of households approximated by the number of returns, the population approximated by the number of personal exemptions, the taxable income, and the income tax paid. As many of these variables showed an almost perfect correlation, I keep only the number of households and the adjusted gross income as controls. The data is available on a yearly basis for the whole time period considered in the estimation.

[19] Bloomberg indices: SSPSPSNC (BNEF survey), SSPSPSNI (BNEF survey), SSPFPSNO (BNEF survey), SOLRAPS (PVinsights poll).

[20] The fraction of sample variance of the dependent variable that is explained by the model, i.e., the variance in the regressors (Stock & Watson, 2020).

Moreover, I add the average electricity price for end customers by state and year provided by the U.S. Energy Information Administration (EIA). However, this did not add substantial quality to the estimation as electricity prices differ regionally on a smaller-than-state level and vary significantly throughout the year. Therefore, the data did not accurately picture the relevant electricity price movements and did not significantly impact the demanded system size which is why I do not include this variable in the final estimation.

Some data transformations are conducted to use the sample for the estimation of the log-log linear model. For detailed information, please refer to Appendix B3.

3.5. Estimation Results

3.5.1. Preferred econometric model

The model specification selected for the final estimation of the price elasticity of demand is given in equation (7). As already outlined above, it comprises several technology-related factors as well as economic control variables and fixed effects for state and installation year. The installed price is instrumentalised by the polysilicon spot price and the sales tax rate.

$$\begin{aligned}\log(\text{systemsize}) = {} & \beta_1 \log(\text{price}) + \beta_2 \log(efficiency) \\ & + \beta_3 D_{\text{new}} + \beta_4 D_{\text{tracking}} + \beta_5 D_{\text{ground}} \\ & + \beta_6 D_{\text{technology}} + \beta_7 D_{\text{mlpe}} \\ & + \gamma_1 \log(\text{households}) + \gamma_2 \log(\text{income}) \\ & + \alpha + \mu + u\end{aligned} \tag{7}$$

To estimate this model, 172,106 complete observations are used (11% of full TTS sample). It must be noted that these observations no longer represent the entire U.S. because by far not every state included in the initial sample has valid observations for all relevant variables. Thus, the data used for estimating the model is representative at most for the states included, namely California, Texas, Arizona, and other states initially grouped as there were not enough observations available at the individual state level. Since a major part (99%) of the remaining sample stems from California, the results mostly picture the situation present there rather than the whole U.S. Likewise, only the years 2010 to 2017 are still represented in the data and only a minor share of installations uses tracking or is ground-mounted. Summary statistics of the remaining data are provided in Appendix A5. The estimation results, models with different instruments and subsets of predictors, a comparison to alternative functional specifications to check the robustness of the results, as well as estimations for regional and economic subgroups is provided in the following to answer the three research questions introduced above.

3.5.2. Price elasticity of demand

The final model yields a constant long-term price elasticity of residential demand for solar PV of -0.443 (Table 1, model (3)). A 1% increase in the price per watt reduces the system size by about 0.44%. This means that investors are rather insensitive to price changes, as the system size declines less than proportional to the increase in price per watt. The finding is broadly in line with the slightly higher elasticity estimate of -0.65 obtained by Gillingham and Tsvetanov (2019a) for Connecticut and slightly lower elasticity of -0.4 found by Rogers and Sexton (2014) for California. However, this comparison of estimates is not entirely valid for two main reasons. Firstly, both Gillingham and Tsvetanov (2019a) and Rogers and Sexton (2014) estimate the effect of a price increase on the adoption, i.e., the number of installations, rather than the system size. Among those who purchase a solar PV system, the price elasticity of system size demand could be lower since the decision on how much capacity to buy might be less sensitive to changes in the price level than the decision to buy at all or not. Secondly, the final data set contains mainly Californian installations where price elasticities could be lower as found by Rogers and Sexton (2014). Thus, the estimate is not directly comparable to Connecticut.

3.5.3. Impact of further relevant variables

Explanatory variables

Next to prices, further variables turn out to be highly relevant in determining the system size to be purchased, all parameters being significant at a 0.1% level in the preferred model (3). Generally, for the interpretation of regression coefficients two aspects need to be considered: the significance of an effect and the effect size itself. Findings might be significant in terms of p-value, but this does not imply a practical significance in terms of effect size. Large data sets like the estimation data used here tend to produce highly significant estimates already for very minor effect sizes as very small differences can be detected as sample size rises (Lin, Lucas Jr, & Shmueli, 2013).

Increasing *module efficiency* by 1% results in a 0.41% increase in the installed system size. In absolute terms, this percentage change is almost as large as for installed price per watt. At first glance, the direction of the effect might seem somewhat counterintuitive: with increasing efficiency, less solar PV capacity should be sufficient to generate a given amount of energy, implying that, where there is little space and smaller systems are installed, higher efficiency modules are purchased, and vice versa. On the other hand, however, higher efficiency also makes the whole installation more profitable and worthwhile investing in, possibly inducing investors to purchase more and larger models. Furthermore, higher module efficiency results in relatively lower BoS and soft costs for installation per generated kWh of electrical energy. As these are not explicitly modelled here, the efficiency coefficient might incorporate this positive effect.

Strongly related to the module efficiency are both the type of *module technology* and *module-level power electronics*. As these factor variables are not log-transformed they cannot be interpreted as elasticities but rather as an expected

Table 1: IV regression results for the final model specification and different robustness checks

	Instrumental Variable Estimation Results					
	Dependent Variable: System Size (W)					
	IV: Polysallicon Price	IV: Polysilicoa Price, Tax Rate	IV: Polysilicon Price, Tax Rate	IV: Polysilicon Price, Tax Rate, Rebate/Grant per Watt	IV: Polysilicon Price, Tax Rate, Rebate/Grant per Watt	IV: Polysalicon Price, Tax Rate, Rebate/Grant per Watt Tax per Watt, Hourly Mean Installer Wage Incentive Programs
	(1)	(2)	(3)	(4)	(5)	(6)
Price per Watt	−0.600***	−0.452***	−0, 443***	−0, 447***	−0.454***	−0.431***
	(0.165)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Module Efficiency	0.175***	0.212***	0.408***	0.213***	0.212***	0.264***
	(0.046)	(0.018)	(0.012)	(0.017)	(0.018)	(0.018)
Dumuy: New Construction	−0.901***	−0.881***	−0.883***	−0.861***	−0.881***	−0.883***
	(0.024)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Dumzny: Tracking	−0.106***	−0.103***	−0.106***	−0.107***	−0.101***	−0.107***
	(0.015)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Dummy: Ground-mounted	0.372***	0.371***	0.359***	0.373***	0.361***	
	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)
Dummy: Premium Module	0.088***	0.066***		0.064***	0.067***	0.053***
	(0.025)	(0.004)		(0.004)	(0.004)	(0.004)
Module Technology: Mono	0.069***	0.056***	0.047***	0.056***	0.056***	0.059***
	(0.015)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Module Technology; Other	0.127***	0.108***	0.088***	0.106***	0.108***	0.107***
	(0.023)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
MLPE: DC Optimizer	0.055***	0.059***	0.053***	0.062***	0.059***	0.060***
	(0.006)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
ML.PE: None	−0.026***	−0.021***	−0.018***	−0.022***	−0.021***	−0.025***
	(0.006)	(0, 003)	(0.003)	(0, 003)	(0, 003)	(0, 003)
Electricity Price	0.850					
	(0.714)					
Households	−0.039***	−0.042***	−0.043***	−0, 012***	−0.042***	−0.012***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
AGH/Aousehold	0.060***	0.061***	0.629***	0.060***	0.633***	
	(0.005)	(0.002)	(0.002)	(0.033)	(0.002)	(0.033)
Wages/Household				−0.610***		−0.607***
				(0.014)		(0.014)

(Continued)

Table 1—continued

Taxable Income/Household				0.067		0.051
				(0.039)		(0.039)
Income Tax/Houshold				−0.074***		−0.067***
				(0.018)		(0.019)
Constant	6.730***	8.630***	8.090***	8.650***	8.630***	8.310***
	(1.860)	(0.056)	(0.043)	(0.096)	(0.056)	(0.098)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	172,106	172,106	172,106	172,096	172,096	165,478
R^2	0.327	0,332	0.331	0.342	0.332	0.343
Adjusted R^2	0.327	0.332	0.331	0.342	0.332	0.343
Resadual Std. Error	0.410(df−172082)	0.409(df−172083)	0.409(df−172084)	0.406(df−172070)	0.409(df−172073)	0.403(df−165455)

Note: * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$
Note: Second stage OLS regression results of explanatory variables on system size, using five different sets of instruments and four different sets of regressors. Model (3) is selected as the final model as it shows the highest R^2 and lowest standard errors among those models with no multicollinearity in the explanatory or control variables.
Source: Own analysis, estimation sample

percentage change from one group to the other (Benoit, 2011). A change in the respective variable changes the system size by $e^{\widehat{\beta}}$, or approximately by $\widehat{\beta}$ * 100%. Relative to polysilicon modules, using monocrystalline technology causes system size to increase by 4.7%. Likewise, the presence of both microinverters and DC optimizers is only weakly but positively related to system size installed. Intuitively, this might be related to the fact that installations with high-quality monocrystalline modules and MLPE technology became more and more prevalent, replacing less efficient systems (Barbose & Darghouth, 2019). Simultaneously, system size continuously increased (s. Appendix A2, Figure 7 and Figure 8). If the year fixed effects included in the model do not entirely capture this movement, the positive, though small coefficient might have resulted from this correlation.

Aside from efficiency-related aspects, the fact that a system is installed on a *new construction*, not as a retrofit on an existing house, strongly determines the system size installed, reducing it by 88.1%. The negative relationship could be due to the fact that new constructions are probably more space-constrained, offering less possibilities to install larger solar PV systems. Additionally, it should be taken into consideration that over 90% of the estimation sample are retrofits and that in the full sample, over 80% of the new construction installations were installed by a single company (Barbose & Darghouth, 2019). The estimate might, therefore, not offer an entirely valid representation.

Regarding the effect of *tracking technology*, the model states a 10.3% decrease in system size installed. As tracking is mainly used to improve the effective efficiency of a solar PV installation by maximising the amount of time the panels face the sun, this is probably particularly relevant for smaller systems with no space available to install further modules. On the contrary, installations that are *ground-mounted* show a 37.1% larger system size than rooftop systems. While the latter are constrained by a natural space limit of the house's roof, ground-mounted installations are likely to have a larger area available. Rooftop installations are also much more common among residential customers, accounting for over 95% of the estimation data.

For a percentage increase in the *electricity price* the model implies a 0.85% increase in system size (model (1)). The effect's direction is reasonable as higher electricity prices might induce customers to invest in more solar PV capacity in order to save money on the electricity bill or even earn some through net metering initiatives. However, this effect is not significant at a 5% level and the model exhibits high collinearity between numerous variables and electricity price, presumably because the information is aggregated and averaged on a state and year level. True electricity price movements may show substantial variation throughout the year or within a state.

Control variables

The effects' directions and sizes are consistent throughout the estimated models, using different sets of instruments and/ or different explanatory variables as regressors (s. Table 1). Some model specifications exhibit substantial multicollinearity, like for instance module efficiency and a dummy indicating a premium module in model (2) or several highly correlated control variables in models (4) and (6)[21]. As final model, I select the best specification in terms of a high coefficient of determination, a low residual standard error, low VIF values, and satisfactory diagnostic test results. Additionally, less complex models are preferable over more complex ones even if the bias of the estimates is lower for large models, because the variance increases in complexity (James et al., 2013). For these reasons, I refrain from including data on wages, population, taxable income, and income tax payment and limit the control variables to adjusted gross income per household and the number of households in a zip code area. Those are important to account for as they control for the average wealth and housing density in a given area. These aspects might influence the system size installed directly as well as through other factors included in the model. The preferred model estimates a small but negative elasticity of -0.04% upon a 1% increase in the *number of households* and a positive effect of 0.06% on system size resulting from a 1% increase in the *adjusted gross income per household*. Both estimate directions are plausible as higher income and more money available enable customers to invest in larger systems. Further, an increase in the number of households is often related to more urbanised regions where space is generally more limited, and installer costs might be higher due to larger market concentration. If more competition reduces installer experience and forces them to operate at smaller and less efficient scales (O'Shaughnessy, 2018), the coefficients might capture the effect of a resulting upward movement in prices compared to less densely populated regions, eventually decreasing the system size bought. Contrary to this, Gillingham et al. (2016) find that installer density is associated with substantially lower prices, likely due to reduced price mark-ups, which is more in line with common economic market theory (Mankiw, 2020; O'Shaughnessy, 2018).

Fixed effects

Besides explanatory and control variables the model includes fixed effects for the year and the U.S. state of installation. They are hard to interpret since fixed effects generally capture all influence over time and across states, respectively, that is not otherwise controlled for in the model. *Time fixed effects* are negative for all years throughout 2017 when compared to 2010 levels and only positive year-over-year for 2012 to 2013 (s. Appendix A6, Figure 24) despite the fact that median system size increased continuously (s. Appendix A6, Figure 7). One plausible reason for the negative

[21] I test all models on multicollinearity by computing the correlation matrices (s. Appendix A3, Figure 16 and Figure 17) and the VIF that indicates how much the variance of a coefficient is inflated (James et al., 2013) (s. Appendix A6, Table 13).

impact on installed capacity is that the fixed effects capture unobserved cost and price trends that are not included in the model. This might be the phase-out of incentive programs and rebates over the years, making systems less profitable (ceteris paribus). Cash rebates, for instance, are capacity-dependent, ending if a certain amount of solar PV has been installed in a given area (Consumer Energy Alliance, 2018; Shrimali & Jenner, 2013). On a federal level, the reduction of ITCs has been postponed to 2020, however. Furthermore, global negative demand shocks pushing up prices, policy changes and import tariffs, increasing wage rates, higher customer acquisition costs for more mature markets, and the price development of substitute sources of electricity like natural gas might be captured by the year coefficients (Burr, 2014; Newell & Raimi, 2014). As some of the factors named above change frequently, using more granular time effects on quarterly or weekly intervals might further improve the estimation. This will be part of the robustness checks in section 3.5.4.2 below.

State fixed effects capture differences across states, estimates showing a positive coefficient for Arizona and other states and a small but negative coefficient for Texas compared to California (s. Appendix A6, Figure 25). However, as there are only four distinct state groups in the model and California accounts for a major part of installations, these state fixed effects must be interpreted with caution. In terms of price-related aspects the effect sizes might be determined by the different implementation of various incentive programs like capacity- or performance-based compensation, local rebates, tax exemptions, and retail net metering compensation present in some U.S. states (Barbose & Darghouth, 2019; Burr, 2014). In California for instance, the California Solar Initiative offers capacity-based rebates, however with a declining absolute rebate amount per watt as system size increases (Burr, 2014). This may cause Californian investors to see a low marginal benefit in installing higher capacities, which is in line with the observation that California exhibits a systematically lower median system size than most other U.S. states (s. Appendix A2, Figure 8). Furthermore, BoS costs differ regionally as they depend more heavily on local wage rates, taxes and competition (Barbose & Darghouth, 2019). As they are not included as a model variable, these differences might be captured by the state coefficients. Apart from price and cost factors, local solar irradiance most likely plays another significant role as it varies across regions. Arizona is the sunniest state in the U.S. (NREL, 2020), increasing the profitability per kilowatts capacity installed compared to other states and, thus, incentivising customers to purchase larger systems. For California and Texas, the solar irradiance is similar on average though it differs within states and cities, suggesting that fixed effects on zip code or census block level might have further improved the estimation.

3.5.4. Robustness comparison to other model specifications

Varying the instruments

As already outlined above, I assess the predictive power of several instrument combinations to obtain the best possible approximation of the variation in solar PV installed prices. The results are displayed in Appendix A6. Particular importance is placed on the model quality measured by R^2 of the first stage as it is predictive for the accurate representation of price in the second stage estimation. Individual instruments as well as different instrument combinations are tested. For individual instruments, variation in the tax rate can explain 52.6% of the variation in price per watt, which is up to twice as much as all other instrument candidates (s. Appendix A6, Table 8). Adding polysilicon price movements as second instrument hardly improves the R^2 value (52.7%). Nevertheless, I use the combined set as instruments because polysilicon price movements capture the input factor cost side of the price development which might be more relevant and predictive in some model applications and time frames, also beyond the scope of this study. The first stage coefficient on polysilicon price is positive, price per watt increasing by 0.059% as polysilicon prices increase by 1%. Counterintuitively, with a percent-increase in the sales tax rate, the installed price per watt decreases. This can be explained by the fact that tax exemptions granted by the government are already regarded here. If more generous financial incentives like tax exemptions are granted in regions where prices are higher – which is in line with findings by Gillingham et al. (2016) – then the resulting effect of an increase in taxes can turn negative. For the data at hand this seems to be the case for some states (s. Appendix A6, Table 12). Further, it is not unreasonable to assume that there are numerous omitted variables taking effect here. Possible examples could be more subsidies granted to installers or lower BoS costs where tax rates are high. This needs to be investigated in more detail to identify the causation behind the estimated correlation.

Estimating the second stage as a simple linear model without further predictors the coefficient of determination[22] is low for all estimations, highlighting the need for further predictor variables (s. Appendix A6, Table 9). Therefore, numerous other model specifications are tested against the final model also in terms of instrumental variables used. Including the explanatory and control variables of the final model and accounting for year and state fixed effects strongly improves the predictive accuracy of the first stage, resulting in an R^2 of 0.992 for the final model (s. Appendix A6, Table 10). Nearly all the variation in the installed price per watt is now captured by the model. However, the polysilicon price coefficient in the first stage reduced form estimating the installed price is now very small and negative (-0.004), though still significant. This is counterintuitive as final prices are expected to move in the same direction as input factor

[22] Although R^2 is not a valid metric for IV regression estimates since "the actual values, not the instruments for the endogenous right-hand-side variables, are used to determine the model sum of squares" (Sribney, Wiggins, & Drukker, 2020), the value is also considered for evaluation as the correct manual computation of the second stage resulted in very similar R^2 values for all relevant model specifications.

costs. One possible explanation might be that time fixed effects now included in the model capture the effect of related changes in input costs that are correlated with the polysilicon price and had before been represented by its coefficient. This hypothesis is supported by a negligible Pearson correlation of price per watt and polysilicon price, only once year is held constant (s. Appendix A2, Figure 16 and Figure 17). Tax rate still has a negative and substantial impact on installed prices (-1.260). Adding further instruments improves the model only at the cost of high multicollinearity, if at all. Furthermore, diagnostic test results show endogeneity of instruments for some specifications[23]. On the other hand, using only the polysilicon price greatly reduces the R^2 to 0.188, in line with the findings from the simple linear first stage regression above. This shows that changes in polysilicon price do not suffice to capture the variation in the input factor costs. It might improve the predictive power to use variation in costs for intermediate products, as long as these variations can still be considered exogenous to the residential investment decision. Such production factors further down the value chain could be wafers or cells, and labour costs for technology development and system installation on a more granular distinction regarding region and time.

Fortunately, the predictive accuracy and good overall fit of the IV model – not the coefficients of the first stage – are of primary interest here. Therefore, a poor model specification in the first stage does not necessarily imply inconsistent coefficient estimates in the second stage as long as the fitted price values are accurate (Angrist & Krueger, 2001).

Varying the functional form

Next to instrument sets, I test four alternative model specifications against the final model described above as benchmark (Table 2, model (1)): A second-degree polynomial in the first stage, interaction of state and year fixed effects, quarterly time fixed effects, and no fixed effects.

The first alternative aims to obtain better prediction in the first stage. I estimate the price per watt with a *second-degree polynomial* for tax rate. Both coefficients are still highly significant, indicating that there might indeed be a non-linear relationship between prices and tax rate, also visible when investigating the regression plot of installed price on tax rate (s. Appendix A3, Figure 15). However, as the first stage coefficient of determination was already very high it could only be improved by 0.001 to 99.3% (s. Appendix A6, Table 11). Moreover, as mentioned above, the goodness of fit in the second stage does not depend on getting the first stage functional form exactly right (Angrist & Krueger, 2001; Kelejian, 1971), which is why the more complex first stage specification is not considered further.

Several legitimate reasons suggest that including an *interaction between state and year* might substantially improve the estimation by accounting for within-state differences over time, in effect allowing each state regression its own intercept and slope value (Borenstein et al., 2010). Contrary to expectations, the resulting elasticity estimates show hardly any difference, however. Presumably, the result would have turned out different for a broader range of data. With almost exclusively Californian installations from the years 2015 to 2017, the information available to detect state-specific differences over time is very limited. Notwithstanding, the first stage R^2 is slightly higher at 0.993 (s. Appendix A6, Table 11).

Quarterly fixed effects are tested against the baseline with yearly fixed effects to account for changes over shorter time intervals. The resulting estimates are displayed in Appendix A6, Figure 26. The second stage R^2 improves only marginally by 0.001, potentially because control data is only added on a yearly basis, offering no variation over quarters. I would generally expect more specific fixed effects to improve the estimation if there is reason to assume that there are differences between observational units. E.g., county, zip code, or even census block fixed effects could better capture constant regional differences in irradiance, community mindsets, and neighbourhood spillover.

Lastly, I use a model estimated *without fixed effects* as a simple benchmark to test whether they make a difference to the model fit. Surprisingly, the second stage R^2 is not much lower than for the other specifications, also when estimating it manually for the second stage. However, the first stage coefficient of determination diminishes (s. Appendix A6, Table 11), implying a worse fit of price per watt used in the subsequent step.

All in all, it is reassuring that the estimates do not differ much across most specifications which is why I choose the baseline, the simplest of the equally well performing models.

3.5.5. Impact of regional and economic differences

In the U.S., income inequality is a serious issue that is strongly linked to state welfare and policies (Jansa, 2020). I therefore group the observations in strata of income and population density as well as by state and estimate their price elasticities separately.

Income level impacts the price elasticity (Andreyeva, Long, & Brownell, 2010), although the effect has not been studied extensively, with most research rather focussing directly on income elasticities (Zhu, Li, Zhou, Zhang, & Yang, 2018). One expects consumers to be more price sensitive if they have little money to spend and vice versa (Mankiw, 2020). Running the preferred model regression for four different income groups[24] I find declining price elasticities as income level increases, meaning that households in zip codes with high average income have a substantially lower price elasticity (-0.195), i.e., are less price sensitive, than households with less

[23]Details on diagnostic tests are provided in section 3.5.6 'Tests on validity and model specification' for the final model.

[24]Observations are grouped by adjusted gross income per household. Equal intervals in thousands of USD are (1) low: (16.5,442], (2) low/medium: (442,867], (3) medium/high: (867,1.29e+03], (4) high: (1.29e+03,1.72e+03].

Table 2: IV regression results of robustness checks for the final model against alternative specifications

	Alternative specifications: Instrumental Variable Estimation Results				
	Dependent Variable: System Size (W)				
	IV: Baseline (1)	IV: Polynomial in first stage (2)	IV: Interaction state and year (3)	IV: Quarterly fixed effects (4)	IV: No fixed fixed effects (5)
Price per Watt	−0.443***	−0.446***	−0.443***	−0.443***	−0.385***
	(0.004)	(0, 004)	(0, 004)	(0.004)	(0.005)
Module Efficiency	0.408***	0.408***	0.409***	0.411***	0.243***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Dummy: New Construction	−0.883***	−0.884***	−0.884***	−0.870***	−0.858***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)
Dummy: Tracking	−0.106***	−0.106***	−0.104***	−0.105***	−0.111***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.013)
Dummy: Grotud-mounted	0.371***	0.371***	0.371***	0.371***	0.359***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Module Technology: Mono	0.047***	0.048***	0.047***	0.048***	0.052***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Module Technology: Other	0.088***	0.089***	0.088***	0.089***	0.101***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
MLPE: DC Optimizer	0.053***	0.053***	0.053***	0.054***	0.051***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
MLPE: None	−0.018***	−0.018***	−0.018***	−0.017***	0.010***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Households	−0.043***	−0.043***	−0.043***	−0.043***	−0.046***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
AGL/Household	0.061***	0.061***	0.061***	0.060***	0.062***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Constant	8.090***	8.090***	8.080***	8.100***	8.230***
	(0.043)	(0.043)	(0.043)	(0.046)	(0.043)
Time FE	Year	Year	Year	Quarter	None
State 11:	Yes	Yes	Yes	Yes	No
Interaction	No	No	Yes	No	No
Polynotaials	No	Yes	No	No	No
Cbservations	172,106	172,106	172,106	172,106	172,106
R^2	0.331	0.331	0.331	0.332	0.323
Adjusted R^2	0.331	0.331	0.331	0.332	0.323
Residual Std. Error	0.409 (df = 172084)	0.409 (df = 172084)	0.409 (df = 172070)	0.409 (df = 172060)	0.411 (df = 172094)

Note: Second stage OLS regression results of explanatory variables on system size. The final model as baseline estimation compared to alternative specifications including a second-degree polynomial in the first stage, interaction of the fixed effects, quarterly time fixed effects, and no fixed effects. All use the same instruments polysilicon price and tax rate for price per watt.
Source: Own analysis, estimation sample

money available (-0.521) (s. Appendix A6, Table 14). These insights, though maybe unsurprising, are highly relevant in the context of targeted marketing and sales for price discrimination strategies as well as policy measures for tailored state subsidies and rebate campaigns.

The relevance of *population density* on price sensitivity is worth investigating as respective insights could help to focus firms' and governments' attention on regions where it is most profitable, both in an economic and societal welfare sense. The regression results displayed in Appendix A6, Table 15 for three groups of population density[25] show small but consistent differences throughout. Low population density regions

[25]Observations are grouped by the number of households. Equal intervals

are associated with a slightly higher price sensitivity (-0.473) than areas with high population density (-0.338). I control for income which might be correlated with the number of households in a zip code area, if the suburbs of metropolitan areas show systematically higher average income levels (Pendall & Carruthers, 2003; Schuetz et al., 2018).

I also run the regression separately for the states California, Arizona, and Texas, which are still included in the sample data when estimating the full model. Notably, the estimates for California hardly change compared to the full model as those made up most of the final observations. Coefficient estimates for Arizona and Texas deviate substantially for some variables, based on relatively few observations and partly missing variation in instrument data (s. Appendix A6, Table 16). This strongly suggests that statements can be made about California, but that the generalisability of the results to other Southern states let alone the entire U.S. should be considered with great caution.

3.5.6. Tests on validity and model specification

In order to ensure the sound application of the IV estimation approach, the validity of the instruments needs to be tested (Stock & Watson, 2020). This comprises two main aspects already outlined in section 3.2.2: Relevance and endogeneity. The diagnostic tests results for the final model are shown in Appendix A7, Table 17.

Relevance of the instruments can be assessed in a straightforward way by calculating the correlation coefficient between the instrument and the endogenous regressor it replaces (s. Appendix A3, Figure 16). The higher the correlation the more information is kept and the better its suitability as an instrument. On the contrary, weak instruments tend to mirror the OLS estimate (Angrist & Krueger, 2001; Angrist & Pischke, 2008). I use the first stage F-statistic test on weak instruments for the preferred model, effectively testing if coefficients on the instruments are all zero in the first stage. The resulting p-value of the test statistic is <0.001%. Hence, I can reject the hypothesis that the instruments are irrelevant.

Exogeneity of instruments is somewhat more complex to determine. It is only possible for an overidentified model that has more instruments than endogenous variables to be replaced[26]. Fortunately, this is the case here. The Sargan test (or J-Statistic) can be used to determine whether the residuals $\hat{u}^{TSLS}$ of the IV model can be explained by the instruments. If they cannot, one can assume that the instruments are indeed exogenous and valid for estimation. For the preferred model, the Sargan test statistic is insignificant. I cannot reject the hypothesis that the instruments show coefficients different from zero to explain the model residuals (Stock & Watson, 2020). Therefore, I can assume exogeneity of the instruments chosen here.

Additionally, I use the Wu-Hausman test to evaluate whether endogeneity is actually present in the original model (6) as otherwise OLS would be preferable over TSLS (Stock & Watson, 2020). The test statistic is again highly significant, and I can reject the hypothesis that OLS is consistent, supporting the TSLS approach used for estimation.

Apart from IV-related tests, I assess some general regression assumptions, outlined in Appendix B1. Corresponding plots are shown in Appendix A7, Figure 27 and Figure 28. The *linearity assumption* of the model seems valid for the first but not for the second stage. This implies that potentially a different type of model could improve the fit. This is not unreasonable as the final model selected here can only explain a rather small share of the variance in the installed system size. Plotting the residuals against the predicted system size values suggests that the *i.i.d. assumption* is violated and *heteroskedasticity* is present in both the first and the second stage. This is at least partly accounted for by using heteroskedasticity-robust standard errors in order to obtain valid test statistics and p-values. The *normality assumption* does not hold in the first stage and indicates heavy left and right tails in the second stage regression. However, due to the high sample size this is a minor issue here as the estimates converge asymptotically towards the true parameter as the number of observations increases (Ghasemi & Zahediasl, 2012; Lumley, Diehr, Emerson, & Chen, 2002; Stock & Watson, 2020).

4. Insight Relevance

4.1. Economic and Policy Implications

The U.S. are still one of the world's major emitters of greenhouse gases (EPA, 2020). In order to sustainably reduce emissions conscious consumption, energy efficiency measures, public policy encouragement, and, most importantly, the increased deployment and implementation of renewable energy sources are necessary (Jaforullah & King, 2015). How can the insights obtained in the analysis be leveraged to move further towards a more sustainable energy landscape in the U.S.? First of all, the knowledge of price elasticities and a more differentiated view on the sensitivity for various subgroups of the population can help to build more targeted and, thus, more effective and cost-efficient incentive programs. If in line with the general political directions, subsidies could be set higher for those who react stronger to price changes. Here, final cost reductions will promote an increase in solar PV capacity most. On the other hand, where price sensitivity is low, benefits like tax reductions could be decreased, generating higher governmental tax income which could ultimately be invested into initiatives where it has a potentially bigger impact. Besides the environmental benefit of more energy generated from renewable resources, the U.S. economy could profit substantially, also from the creation of new job opportunities for higher- and lower-skilled workers (Wei, Patadia, & Kammen, 2010).

Likewise, the results can be useful as quantitative evidence and guidance in the ongoing discussions on reducing

in thousands are (1) low: (0.09,18.1], (2) medium: (18.1,36.2], (3) high: (36.2,54.3].

[26]For an exactly identified model it is not possible to test exogeneity formally, but one must rely on logical reasoning (Stock & Watson, 2020).

or completely phasing out financial incentives (Gillingham & Tsvetanov, 2019a). From relative changes in installed capacity, policymakers can derive absolute capacity added. This allows forecasting changes in the overall capacity upon policy modifications using simple policy simulations and deploying the prior estimated pass-through rate and price elasticity (Gillingham & Tsvetanov, 2019a). Moreover, Gillingham and Tsvetanov (2019a) show that the potential reduction of GHG emissions from certain programs can be quantified. Naturally, the assumed amount of avoided emissions depends on the expectations on the type of generation that is displaced by renewables, both today and in future (Gillingham & Tsvetanov, 2019a). Gillingham and Tsvetanov (2019a) estimate the cost of abatement through state incentives for solar PV to lie notably above the U.S. government social costs of carbon estimate. Directly comparing the costs of the program to the social benefit of pollution abated allows to evaluate the cost-effectiveness and social desirability in an economic welfare sense. This method could be used to rank policy programs according to a quantifiable impact on social welfare and invest funds where they create the greatest benefit, also going beyond solar PV. This would take into account that a new technology should always be assessed not only in the light of cost-effectiveness but also environmental friendliness and social acceptance (Khan & Arsalan, 2016).

4.2. Business Implications

While governments can clearly leverage the insights on demand elasticities for solar PV in numerous ways, firms can also capitalise on them. Seeing that residential consumers are not very price sensitive, pricing strategies can be optimised to maximise profits, assuming that they operate in an imperfect market and do not set prices equal to marginal costs (Gillingham & Tsvetanov, 2019a; Mankiw, 2020). Besides that, more data-driven and cost-optimal decisions on research and development efforts to reduce costs even further can be taken. As the study also indicates different price sensitivity for population groups and regions, firms could consider targeted price discrimination. Likewise, the possibility to forecast the shift in consumption upon price changes more accurately also in the short-term enables firms to improve their production planning and draw near actual demand for a better supply-demand-fit.

5. Concluding Remarks and Outlook

5.1. Critical Review

Several aspects need to be mentioned when assessing the quality of the estimation results, regarding both data and methodology.

Data quality. The estimation results can only ever be as good as the quality of the data they are based on. The data in the TTS sample are self-reported by installers, depending on reporting conventions which potentially vary significantly. The scope of installed prices can sometimes even include warranties, monitoring and maintenance, re-roofing costs, and loan-related fees (Barbose & Darghouth, 2019). Additionally, installed prices do not necessarily reflect actual costs as they include profit margins and other installer-related characteristics which cannot be captured by cost-related instruments. What is more, according to Barbose and Darghouth (2019) the data set likely contains all kinds of systems, not only turnkey solutions, which is not fully visible from the system information and, thus, not considered in the estimation.

Data coverage. The data were only collected for some states mainly through incentive administrators (Barbose & Darghouth, 2019) which might cover most but probably not all installed systems in the U.S. This objection is further supported by a weak or missing sample coverage for smaller, often lower-cost, state markets. Self-selection bias could be present here (Heckman, 1990). Discarding observations with missing predictor values in the final data sample and assuming them to be missing at random can likewise lead to biased estimates, especially if installations with missing values differ systematically from the completely observed cases (Gelman & Hill, 2006). Furthermore, a broader data coverage is necessary for potential instrument data in order to optimally capture the variation in prices.

Data granularity. The sample of installations collected by the LNBL contains information on the installation date, making it possible to identify the timeframe of the decision to invest quite accurately, assuming that buyers base their decision on the most up-to-date information available at that point in time. The same holds true for very detailed location data, provided at zip code level. Unfortunately, most of the complementary data joined to the TTS sample were available only for a much broader time frame and region. This prevented the model from identifying variation on a more granular level which might have otherwise added significant information to the estimation and greatly improved its accuracy.

Besides those data-related aspects, I shed light on drawbacks of the estimation methodology and model specification.

Omitted variables. The most evident and pressing issue is the fact that by far not all relevant variables could be considered in the final model, either due to unavailability or unobservability of the information. For one thing, I did not consider the actual performance potential of individual systems, which is to a major part determined by the weather conditions and the amount of sunshine received in a certain installation location. Adding solar irradiance data on zip code or census block level could greatly improve the estimation. This is relevant for the initial decision to invest but might also influence the size of the system installed. Equally important, the initial installed price is only one determining factor to assess the economics of a PV installation. To gain comprehensive insights on the profitability and benefits, aspects such as ongoing operating and maintenance costs, effective performance, later retrofitting costs, saved electricity costs and payback period could also be considered as they are possibly already taken into account in the decision on how much capacity to invest in. Furthermore, both federal, state, and utility support policies and regulations have hardly been taken into consideration but are most likely to have a tremendous effect on the buying decision. Capacity- and performance-based incentives will particularly impact the size of the system installed. Those data were not or only incompletely available and should be added for future estimations, if possible. Further, sociodemographic aspects, like investor age and education as well as idealistic values and mindset might impact the demand. The latter are unobservable but could be inferred from political party membership or voting behaviour (Iizuka, 2016; Matthew E. Kahn & Matsusaka, 1997). Additionally, Bollinger and Gillingham (2012), Graziano and Gillingham (2015), and Palm (2017) find significantly positive peer effects on the adoption of solar panels, suggesting that previous installations in the vicinity matter through neighbours attitudes and social influence. The omission of these variables, if they are correlated with any variable included in the model, will cause endogeneity and biased estimation results (Stock & Watson, 2020).

Model specification. The test on linearity in the second stage of estimating the log-log model suggests that the actual relationship might not be represented entirely accurately. It might be worth the effort to investigate which form better describes the relationship to improve the representation of the demand curve and obtain correct elasticity estimates. In case the price enters the model in a non-linear way or the model is even non-linear in its parameters (Imbens & Wooldridge, 2007; Wooldridge, 2015), control functions rather than standard IV methods should be applied, as done by Gillingham and Tsvetanov (2019a). Also, examining more profoundly which variables are highly relevant, differentiating between regions and population groups, would further back up the estimation. This could also help to identify even better instruments to further improve the first stage estimation.

5.2. Limitations

As with most research, the estimation results in this study are highly context-sensitive. Consequently, their application to other contexts is neither straight-forward nor generally possible. Although the goal of the study is to estimate the PV demand and price elasticity for the entire U.S., the final data sample contains almost exclusively Californian installations. Therefore, generalising these estimation results even to other states, let alone countries or cultures, is not unreservedly recommended, but must be done under consideration of numerous relevant aspects, if at all. Additionally, the estimation sample is limited to residential customers. There is good reason to assume that estimates would be substantially different in non-residential and utility-scale applications (Barbose & Darghouth, 2019). Furthermore, as pointed out several times before, the price sensitivity is estimated regarding the system size demanded rather than the discrete adoption decision. This needs to be taken into consideration not only when interpreting the coefficient estimates but also when generalising to a broader scope as the price elasticities regarding the system size are only accurate given that a consumer decided to invest in solar PV.

Lastly, one needs to be aware that correlation does not imply causation (Altman & Krzywinski, 2015; Holland, 1986). More than often, there are multiple ways to interpret coefficients. As a regression is unable to picture a causal direction, one needs to bear in mind that equations build on assumptions regarding the underlying causality (Altman & Krzywinski, 2015). If I do not entirely account for endogeneity in the model, variation in some variables might still bias coefficient estimates. This raises the question whether it is justifiable to report ceteris paribus effects.

5.3. Outlook and Conclusion

The study provides methodological insights as well as practical recommendations. Thoroughly and holistically addressing the concerns outlined above would be another big step forward towards applying the model not only to solar photovoltaics but also to many other recent and emerging technologies (Gillingham & Tsvetanov, 2019a, 2019b). This can yield valuable insights for future governmental policies and firm decisions alike, potentially also providing a way to assess the effectiveness of different types of programs. Further, extending the estimation to the technology adoption as explained variable in a two-stage model would enable more far-reaching and differentiated statements (e.g., Bollinger & Gillingham, 2012; Cui, 2018; Dong & Sigrin, 2019; Gillingham & Tsvetanov, 2019a; Palm, 2017; Rogers & Sexton, 2014).

Understanding the factors determining the propensity to invest in solar PV is particularly relevant when it comes to demand forecasting. The within-sample insights obtained in this study could be leveraged in order to make out-of-sample predictions. Numerous machine learning techniques are well-applicable to the estimation problem at hand. Especially non-parametric, tree-based learners like Random Forest and Gradient Boosting have proven highly beneficial for both classification and regression in many economic and business applications (James et al., 2013; Murdoch, Singh, Kumbier, Abbasi-Asl, & Yu, 2019). At the expense of interpretability (Orrù, Monaro, Conversano, Gemignani, & Sartori, 2020), they can use huge amounts of data and features while making no assumption on the functional form, which is convenient when the model is complex. Furthermore, non-parametric learners natively handle outliers and multicollinearity and are able to capture regional differences and non-linear relationships, which is highly relevant in case location matters (James et al., 2013). Using these methods on the estimation problem at hand could greatly supplement the interpretable insights obtained in this study and increase its relevance for researchers, businesses, as well as state and federal governments.

For now, the study successfully provides insights on the price elasticity for solar PV – if not for the entire U.S., at least for California. It shows that customers are generally rather insensitive to price changes. It also brings to light the relevance of other factors impacting the demand, module efficiency being almost as relevant as price per watt when it comes to the system size installed. The comparison of price elasticities within various population subsets highlights the potential of more targeted interventions to maximise the energy amount generated from renewable resources and promote the reduction of greenhouse gas emissions. Thereby, solar photovoltaics can indeed make a major contribution to the sustainable transformation of the energy and electricity generation landscape in the United States. Setting sound policies and incentive programs based on the findings described above, America could even attain to become what Barack Obama aspired to seven years ago: A leader in the global transition towards a sustainable economy.

References

Altman, N., & Krzywinski, M. (2015). Association, correlation and causation [Journal Article]. *Nature Methods*, *12*(10), 899-900. Retrieved from https://doi.org/10.1038/nmeth.3587 doi: 10.1038/nmeth.3587

Andreyeva, T., Long, M. W., & Brownell, K. D. (2010). The impact of food prices on consumption: A systematic review of research on the price elasticity of demand for food [Journal Article]. *American Journal of Public Health*, *100*(2), 216-222. Retrieved from https://ajph.aphapublications.org/doi/abs/10.2105/AJPH.2008.151415 doi: 10.2105/ajph.2008.151415

Angrist, J. D., & Krueger, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments [Journal Article]. *Journal of Economic perspectives*, *15*(4), 69-85.

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion* [Book]. Princeton University Press.

Baltagi, B. (2008). *Econometric analysis of panel data* [Book]. John Wiley & Sons.

Barbose, G., & Darghouth, N. (2019). *Tracking the sun: Pricing and design trends for distributed photovoltaic systems in the united states, 2019 edition* (Report). Lawrence Berkeley National Laboratory, Berkeley, USA. Retrieved from https://emp.lbl.gov/tracking-the-sun

Baum, C., Schaffer, M., & Stillman, S. (2007). *Ivendog: Stata module to calculate durbin-wu-hausman endogeneity test after ivreg* [Computer Program]. Retrieved from https://EconPapers.repec.org/RePEc:boc:bocode:s429401

Benoit, K. (2011). Linear regression models with logarithmic transformations [Journal Article]. *London School of Economics, London*, *22*(1), 23-36.

Bernstein, M. A., & Griffin, J. (2006). *Regional differences in the price-elasticity of demand for energy* (Report). National Renewable Energy Laboratory, Golden, CO, USA. Retrieved from https://www.nrel.gov/docs/fy06osti/39512.pdf

Blakers, A. (2019). Development of the perc solar cell [Journal Article]. *IEEE Journal of Photovoltaics*, *9*(3), 629-635. doi: 10.1109/JPHOTOV.2019.2899460

Bollinger, B., & Gillingham, K. (2012). Peer effects in the diffusion of solar photovoltaic panels [Journal Article]. *Marketing Science*, *31*(6), 900-912. Retrieved from https://EconPapers.repec.org/RePEc:inm:ormksc:v:31:y:2012:i:6:p:900-912

Bollinger, B., & Gillingham, K. (2019). Learning-by-doing in solar photovoltaic installations [Journal Article]. *Available at SSRN 2342406*. doi: http://dx.doi.org/10.2139/ssrn.2342406

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis [Journal Article]. *Research synthesis methods*, *1*(2), 97-111.

Burr, C. (2014). Subsidies, tariffs and investments in the solar power market [Journal Article]. *University of Colorado-Boulder Working Paper*.

Center for Climate and Energy Solutions. (2019). *Paris climate agreement q&a* (Report). Retrieved from https://www.c2es.org/site/assets/uploads/2017/09/paris-climate-agreement-qa.pdf

Center for Sustainable Systems. (2019). *U.s. energy system factsheet* (Report). University of Michigan, Ann Arbor, USA. Retrieved from http://css.umich.edu/sites/default/files/US%20Energy%20System_CSS03-11_e2019.pdf

Consumer Energy Alliance. (2018). *Incentivizing solar energy: An in-depth analysis of u.s. solar incentives* (Report). Retrieved from https://consumerenergyalliance.org/cms/wp-content/uploads/2018/06/Solar-incentive-report-060418.pdf

Cui, M. (2018). Demand and supply for rooftop solar [Conference Proceedings]. In *University of california, los angeles, usa.*

Deline, C., Meydbray, J., Donovan, M., & Forrest, J. (2012). *Photovoltaic shading testbed for module-level power electronics* (Report). National Renewable Energy Laboratory, Golden, CO, USA. Retrieved from https://www.osti.gov/servlets/purl/1045715 doi: https://dx.doi.org/10.2172/1045715

Denholm, P., & Margolis, R. (2007). *Regional per capita solar electric footprint for the united states* (Report). National Renewable Energy Laboratory, Golden, CO, USA. Retrieved from https://www.nrel.gov/docs/fy08osti/42463.pdf

Dennis, B. (2019, 2020, April 9). [Electronic Article]. Retrieved from https://www.washingtonpost.com/climate-environment/2019/11/04/trump-makes-it-official-us-will-withdraw-paris-climate-accord/

Dong, C., & Sigrin, B. (2019). Using willingness to pay to forecast the adoption of solar photovoltaics: A "parameterization + calibration" approach [Journal Article]. *Energy Policy*, *129*, 100-110.

Dong, C., Wiser, R., & Rai, V. (2018). Incentive pass-through for residential solar systems in california [Journal Article]. *Energy Economics*, *72*, 154-165. Retrieved from http://www.sciencedirect.com/science/article/pii/S0140988318301397 doi: https://doi.org/10.1016/j.eneco.2018.04.014

EIA. (2019a). *Annual energy outlook 2019* (Report). U.S. Energy Information Administration, Washington DC, USA. Retrieved from https://www.eia.gov/outlooks/aeo/

EIA. (2019b). *Monthly energy review (april 2019)* (Report). U.S. Energy Information Administration, Washington DC, USA. Retrieved from https://www.eia.gov/totalenergy/data/monthly/

EIA. (2019c). (Vol. 2020,) (Web Page No. April 4). Retrieved from https://www.eia.gov/todayinenergy/detail.php?id=39092

EIA. (2019d). (Vol. 2020,) (Web Page No. July 25). Retrieved from https://www.eia.gov/todayinenergy/detail.php?id=40873

EIA. (2020a). *Annual energy outlook 2020* (Report). U.S. Energy Information Administration, Washington DC, USA. Retrieved from https://www.eia.gov/outlooks/aeo/

EIA. (2020b). *Monthly energy review (june 2020)* (Report). U.S. Energy Information Administration, Washington DC, USA. Retrieved from https://www.eia.gov/totalenergy/data/monthly/

Elshurafa, A. M., Albardi, S. R., Bigerna, S., & Bollino, C. A. (2018). Estimating the learning curve of solar pv balance–of–system for over 20 countries: Implications and policy recommendations [Journal Article]. *Journal of Cleaner Production*, *196*, 122-134. Retrieved from http://www.sciencedirect.com/science/article/pii/S0959652618316652 doi: https://doi.org/10.1016/j.jclepro.2018.06.016

EPA. (2020). *Inventory of u.s. greenhouse gas emissions and sinks: 1990-2018* (Report). U.S. Environmental Protection Agency, Washington DC, USA. Retrieved from https://www.epa.gov/sites/production/files/2020-04/documents/us-ghg-inventory-2020-main-text.pdf

Feldman, D. J., & Margolis, R. M. (2020). *Q4 2019/q1 2020 solar industry update* (Report). National Renewable Energy Laboratory, Golden, CO, USA. Retrieved from https://www.nrel.gov/docs/fy20osti/77010.pdf

Friedman, L. (2019, 2020, April 9). [Electronic Article]. Retrieved from https://www.nytimes.com/2019/11/04/climate/trump-paris-agreement-climate.html

Frondel, M., & Vance, C. (2013). Fuel taxes versus efficiency standards–an instrumental variable approach [Journal Article]. *Ruhr Economic Paper*(445).

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models* [Book]. Cambridge: Cambridge University Press. Retrieved from https://www.cambridge.org/core/books/data-analysis-using-regression-and-multilevelhierarchical-models/32A29531C7FD730C3A68951A17C9D983 doi: DOI:10.1017/CBO9780511790942

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: a guide for non-statisticians [Journal Article]. *International journal of endocrinology and metabolism*, *10*(2), 486.

Gillingham, K., & Bollinger, B. (2019). Social learning and solar photovoltaic adoption [Journal Article]. *Available at SSRN 3500930*. doi: https://dx.doi.org/10.2139/ssrn.3500930

Gillingham, K., Deng, H., Wiser, R., Darghouth, N., Nemet, G., Barbose, G., ... Dong, C. (2016). Deconstructing solar photovoltaic pricing [Journal Article]. *The Energy Journal*, *37*(3).

Gillingham, K., & Tsvetanov, T. (2019a). Hurdles and steps: Estimating demand for solar photovoltaics [Journal Article]. *Quantitative Eco-*

nomics, *10*(1), 275-310.
Gillingham, K., & Tsvetanov, T. (2019b). Supplement to "hurdles and steps: Estimating demand for solar photovoltaics" [Journal Article].
Graziano, M., & Gillingham, K. (2015). Spatial patterns of solar photovoltaic system adoption: the influence of neighbors and the built environment [Journal Article]. *Journal of Economic Geography*, *15*(4), 815-839.
Heckman, J. J. (1990). Selection bias and self-selection [Book Section]. In J. Eatwell, M. Milgate, & P. Newman (Eds.), *Econometrics* (p. 201-224). London: Palgrave Macmillan UK. Retrieved from https://doi.org/10.1007/978-1-349-20570-7_29 doi: 10.1007/978-1-349-20570-7_29
Heckman, J. J., & Navarro, S. (2007). Dynamic discrete choice and dynamic treatment effects [Journal Article]. *Journal of Econometrics*, *136*(2), 341-396. Retrieved from http://www.sciencedirect.com/science/article/pii/S0304407605002186 doi: https://doi.org/10.1016/j.jeconom.2005.11.002
Holland, P. W. (1986). Statistics and causal inference [Journal Article]. *Journal of the American Statistical Association*, *81*(396), 945-960. Retrieved from https://www.tandfonline.com/doi/abs/10.1080/01621459.1986.10478354 doi: 10.1080/01621459.1986.10478354
Hughes, J., & Podolefsky, M. (2015). Getting green with solar subsidies: Evidence from the california solar initiative [Journal Article]. *Journal of the Association of Environmental and Resource Economists*, *2*(2), 235-275. Retrieved from https://EconPapers.repec.org/RePEc:ucp:jaerec:doi:10.1086/681131
IEA. (2019). *Renewables 2019* (Report). International Energy Agency, Paris, France. Retrieved from https://www.iea.org/reports/renewables-2019
Iizuka, M. (2016). Role of environmental awareness in achieving sustainable development [Journal Article].
Imbens, G., & Wooldridge, J. (2007). What's new in econometrics? lecture 6: Control functions and related methods [Journal Article]. *NBER*, *6*, 420-445. Retrieved from https://www.nber.org/WNE/Slides7-31-07/Slides7-31-07.pdf
IRENA. (2019). *Renewable power generation cost in 2018* (Report). International Renewable Energy Agency, Abu Dhabi, United Arab Emirates. Retrieved from https://www.irena.org/publications/2019/May/Renewable-power-generation-costs-in-2018
Jaforullah, M., & King, A. (2015). Does the use of renewable energy sources mitigate co2 emissions? a reassessment of the us evidence [Journal Article]. *Energy Economics*, *49*, 711-717. Retrieved from http://www.sciencedirect.com/science/article/pii/S0140988315001280 doi: https://doi.org/10.1016/j.eneco.2015.04.006
James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112) [Book]. Springer.
Jansa, J. M. (2020). Chasing disparity: Economic development incentives and income inequality in the u.s. states [Journal Article]. *State Politics & Policy Quarterly*, *0*(0), 1532440019900259. Retrieved from https://journals.sagepub.com/doi/abs/10.1177/1532440019900259 doi: 10.1177/1532440019900259
Kelejian, H. H. (1971). Two-stage least squares and econometric systems linear in parameters but nonlinear in the endogenous variables [Journal Article]. *Journal of the American Statistical Association*, *66*(334), 373-374.
Keyes, A. T., Lambert, K. F., Burtraw, D., Buonocore, J. J., Levy, J. I., & Driscoll, C. T. (2019). The affordable clean energy rule and the impact of emissions rebound on carbon dioxide and criteria air pollutant emissions [Journal Article]. *Environmental Research Letters*, *14*(4), 044018.
Khan, J., & Arsalan, M. H. (2016). Solar power technologies for sustainable electricity generation – a review [Journal Article]. *Renewable and Sustainable Energy Reviews*, *55*, 414-425. Retrieved from http://www.sciencedirect.com/science/article/pii/S1364032115012149 doi: https://doi.org/10.1016/j.rser.2015.10.135
Krupnick, A., Cropper, M., Kopp, R., Newell, R., Pizer, W. A., Rennert, K., ... Keyes, A. (2018). Comments to us epa on the proposed affordable clean energy rule [Journal Article]. *Washington, DC: Resources for the Future*.
Lin, M., Lucas Jr, H. C., & Shmueli, G. (2013). Too big to fail: Large samples and the p-value problem [Journal Article]. *Information Systems Research*, *24*(4), 906-917.
Lumley, T., Diehr, P., Emerson, S., & Chen, L. (2002). The importance of the normality assumption in large public health data sets [Journal Article]. *Annual review of public health*, *23*(1), 151-169.
MacKay, A., & Miller, N. (2019). Estimating models of supply and demand: Instruments and covariance restrictions [Journal Article]. *Available at SSRN 3025845*. doi: https://dx.doi.org/10.2139/ssrn.3025845
Mankiw, N. G. (2020). *Principles of economics* [Book]. Cengage Learning.
Matthew E. Kahn, B., & Matsusaka, J. G. (1997). Demand for environmental goods: Evidence from voting patterns on california initiatives [Journal Article]. *The Journal of Law and Economics*, *40*(1), 137-174. Retrieved from https://www.journals.uchicago.edu/doi/abs/10.1086/467369 doi: 10.1086/467369
Mewton, R. T., & Cacho, O. J. (2011). Green power voluntary purchases: Price elasticity and policy analysis [Journal Article]. *Energy Policy*, *39*(1), 377-385.
Miller, M., & Alberini, A. (2016). Sensitivity of price elasticity of demand to aggregation, unobserved heterogeneity, price trends, and price endogeneity: Evidence from us data [Journal Article]. *Energy Policy*, *97*, 235-249.
Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R., & Yu, B. (2019). Interpretable machine learning: definitions, methods, and applications [Journal Article]. *arXiv preprint arXiv:1901.04592*.
Newell, R. G., & Raimi, D. (2014). Implications of shale gas development for climate change [Journal Article]. *Environmental science & technology*, *48*(15), 8360-8368.
NREL. (2015, 2015-09-01). *Building america case study: Photovoltaic systems with module-level power electronics* (Report). National Renewable Energy Laboratory, Golden, CO, USA. Retrieved from https://www.osti.gov/servlets/purl/1225312
NREL. (2020, August 8, 2020). [Online Database]. Retrieved from https://www.nrel.gov/gis/solar.html
Nugent, D., & Sovacool, B. K. (2014). Assessing the lifecycle greenhouse gas emissions from solar pv and wind energy: A critical meta-survey [Journal Article]. *Energy Policy*, *65*, 229-244.
Orrù, G., Monaro, M., Conversano, C., Gemignani, A., & Sartori, G. (2020). Machine learning in psychometrics and psychological research [Journal Article]. *Frontiers in Psychology*, *10*(2970). Retrieved from https://www.frontiersin.org/article/10.3389/fpsyg.2019.02970 doi: 10.3389/fpsyg.2019.02970
O'Shaughnessy, E. J. (2018). *The effects of market concentration on residential solar pv prices: Competition, installer scale, and soft costs* (Report). National Renewable Energy Laboratory, Golden, CO, USA. Retrieved from https://www.osti.gov/biblio/1452704 doi: 10.2172/1452704
Palm, A. (2017). Peer effects in residential solar photovoltaics adoption—a mixed methods study of swedish users [Journal Article]. *Energy Research & Social Science*, *26*, 1-10.
Parida, B., Iniyan, S., & Goic, R. (2011). A review of solar photovoltaic technologies [Journal Article]. *Renewable and Sustainable Energy Reviews*, *15*(3), 1625-1636. Retrieved from http://www.sciencedirect.com/science/article/pii/S1364032110004016 doi: https://doi.org/10.1016/j.rser.2010.11.032
Pendall, R., & Carruthers, J. I. (2003). Does density exacerbate income segregation? evidence from u.s. metropolitan areas, 1980 to 2000 [Journal Article]. *Housing Policy Debate*, *14*(4), 541-589. Retrieved from https://doi.org/10.1080/10511482.2003.9521487 doi: 10.1080/10511482.2003.9521487
Perea, A., Smith, C., Davis, M., Sun, X., White, B., Cox, M., ... Baca, J. (2020a). *U.s. solar market insight report 2019 year in review* (Report). Solar Energy Industries Association, Washington DC, USA. Retrieved from https://www.seia.org/research-resources/solar-market-insight-report-2019-year-review
Perea, A., Smith, C., Davis, M., Sun, X., White, B., Cox, M., ... Baca, J. (2020b). *U.s. solar market insight report q2 2020* (Report). Solar Energy Industries Association, Washington DC, USA. Retrieved from https://www.woodmac.com/research/products/power-and-renewables/us-solar-market-insight/

Platzer, M. D. (2012). Us solar photovoltaic manufacturing: Industry trends, global competition, federal support [Conference Proceedings]. In *Library of congress, congressional research service.* Library of Congress, Congressional Research Service.

Raman, S. (2013). Fossilizing renewable energies [Journal Article]. *Science as Culture*, *22*(2), 172-180.

Rogers, E., & Sexton, S. (2014). Effectiveness of subsidies for residential rooftop solar adoption: Lessons from california [Journal Article]. *North Carolina State University Working Paper*.

Schuetz, J., Larrimore, J., Merry, E. A., Robles, B. J., Tranfaglia, A., & Gonzalez, A. (2018). Are central cities poor and non-white? [Journal Article]. *Journal of Housing Economics*, *40*, 83-94.

SEIA. (2019). *Section 201 solar tariffs* (Report). Solar Energy Industries Association, Washington DC, USA. Retrieved from https://www.seia.org/sites/default/files/2019-12/SEIA-Section-201-Factsheet-Dec2019.pdf

SEIA. (2020). *Solar investment tax credit* (Report). Solar Energy Industries Association, Washington DC, USA. Retrieved from https://www.seia.org/sites/default/files/2020-01/SEIA-ITC-Factsheet-2020-Jan_1.pdf

Shrimali, G., & Jenner, S. (2013). The impact of state policy on deployment and cost of solar photovoltaic technology in the u.s.: A sector-specific empirical analysis [Journal Article]. *Renewable Energy*, *60*, 679-690. Retrieved from http://www.sciencedirect.com/science/article/pii/S0960148113003169 doi: https://doi.org/10.1016/j.renene.2013.06.023

Singh, G. K. (2013). Solar power generation by pv (photovoltaic) technology: A review [Journal Article]. *Energy*, *53*, 1-13. Retrieved from http://www.sciencedirect.com/science/article/pii/S0360544213001758 doi: https://doi.org/10.1016/j.energy.2013.02.057

Sribney, W., Wiggins, V., & Drukker, D. (2020). (Vol. 2020,) (Web Page No. August 2). Retrieved from https://www.stata.com/support/faqs/statistics/two-stage-least-squares/

Stock, J. H., & Watson, M. W. (2020). *Introduction to econometrics* [Book]. Pearson.

Tax Policy Center. (2020). *The state of state (and local) tax policy* (Report). Urban-Brookings Tax Policy Center, Washington DC, USA. Retrieved from https://www.taxpolicycenter.org/sites/default/files/briefing-book/how_do_state_and_local_sales_taxes_work_1.pdf

United Nations. (2016, 2020, April 16). [Online Database]. Retrieved from https://www4.unfccc.int/sites/NDCStaging/Pages/All.aspx

van der Hulst, M. K., Huijbregts, M. A., van Loon, N., Theelen, M., Kootstra, L., Bergesen, J. D., & Hauck, M. (2020). A systematic approach to assess the environmental impact of emerging technologies: A case study for the ghg footprint of cigs solar photovoltaic laminate [Journal Article]. *Journal of Industrial Ecology*.

Wei, M., Patadia, S., & Kammen, D. M. (2010). Putting renewables and energy efficiency to work: How many jobs can the clean energy industry generate in the us? [Journal Article]. *Energy Policy*, *38*(2), 919-931. Retrieved from http://www.sciencedirect.com/science/article/pii/S0301421509007915 doi: https://doi.org/10.1016/j.enpol.2009.10.044

Woodhouse, M., Smith, B., Ramdas, A., & Margolis, R. (2019). *Crystalline silicon photovoltaic module manufacturing costs and sustainable pricing: 1h 2018 benchmark and cost reduction road map* (Report). National Renewable Energy Laboratory, Golden, CO, USA. Retrieved from https://www.nrel.gov/docs/fy19osti/72134.pdf

Wooldridge, J. M. (2005). Instrumental variables estimation with panel data [Journal Article]. *Econometric Theory*, *21*(4), 865-869. Retrieved from www.jstor.org/stable/3533399

Wooldridge, J. M. (2015). Control function methods in applied econometrics [Journal Article]. *Journal of Human Resources*, *50*(2), 420-445.

Wright, P. G. (1928). *Tariff on animal and vegetable oils* [Book]. Macmillan Company, New York.

Yin, H., & Powers, N. (2010). Do state renewable portfolio standards promote in-state renewable generation? [Journal Article]. *Energy Policy*, *38*(2), 1140-1149.

Zainudin, H. N., & Mekhilef, S. (2010). Comparison study of maximum power point tracker techniques for pv systems [Conference Proceedings]. In *14th international middle east power systems conference, cairo university, egypt.*

Zhang, H.-B., Dai, H.-C., Lai, H.-X., & Wang, W.-T. (2017). Us withdrawal from the paris agreement: Reasons, impacts, and china's response [Journal Article]. *Advances in Climate Change Research*, *8*(4), 220-225.

Zhu, X., Li, L., Zhou, K., Zhang, X., & Yang, S. (2018). A meta-analysis on the price elasticity and income elasticity of residential electricity demand [Journal Article]. *Journal of Cleaner Production*, *201*, 169-177. Retrieved from http://www.sciencedirect.com/science/article/pii/S0959652618323588 doi: https://doi.org/10.1016/j.jclepro.2018.08.027