Online-Appendix zu

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

Veronique Clara Bukow
Technische Universität München

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

Appendix

Note:

Full sample refers to the ‘Tracking the Sun’ data set as published by the LBNL (Barbose & Darghouth, 2019) without changes unless specified otherwise.

Price sample refers to the sample left after applying all price-related selection criteria described in section 3.4.1.2 ‘Data selection’.

Estimation sample refers to the sample left after applying all selection criteria, also non-price related, described in section 3.4.1.2 ‘Data selection’.

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 ‘Preferred econometric model’.

A1. Solar Photovoltaic Characteristics

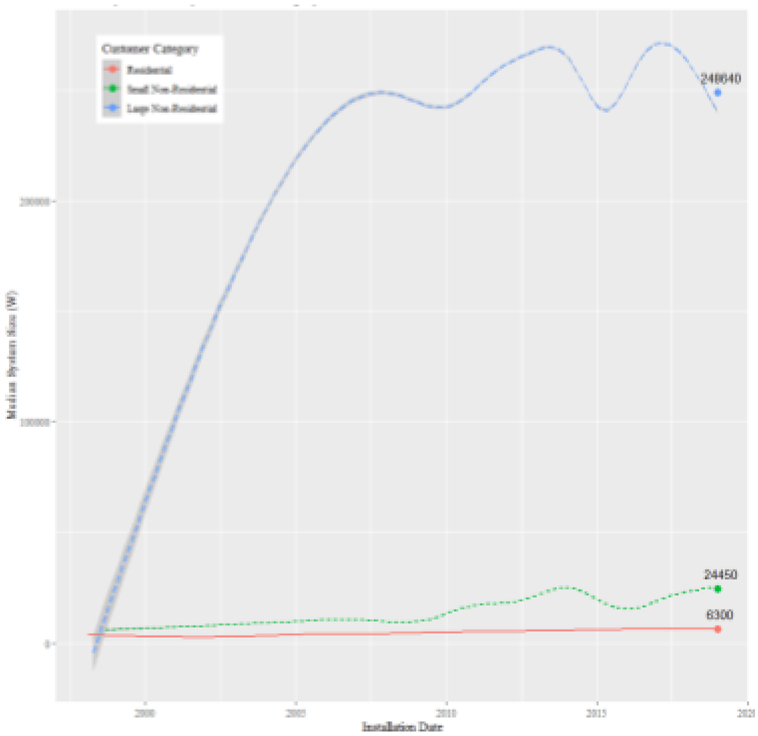

Figure SEQ Figure * ARABIC 1. Median system size by customer category over time, 1998 to 2008
Note: Lines fitted using generalised additive model (GAM) smoother, displaying .95 confidence interval.
Source: LBNL Tracking the Sun data (full sample)

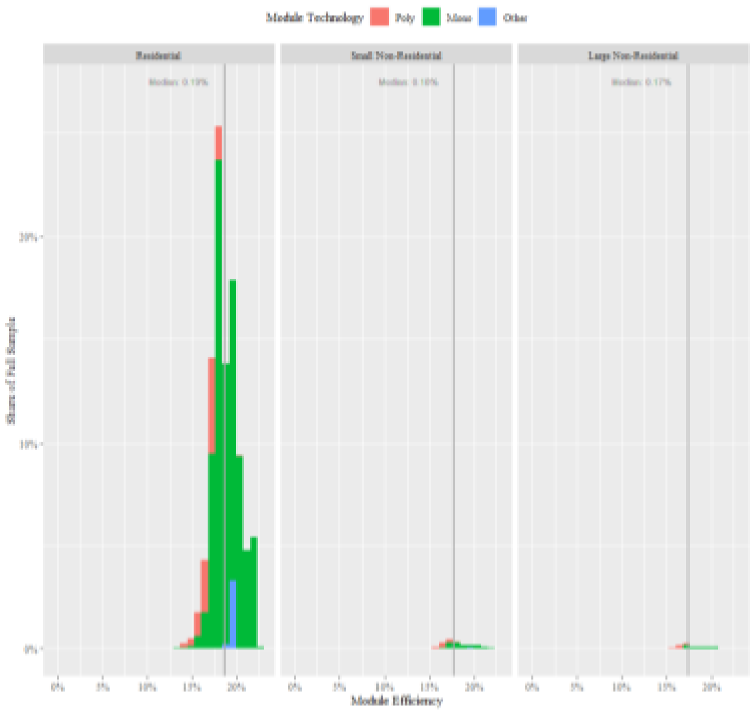

Figure SEQ Figure * ARABIC 2. Module efficiency by module technology and customer category, 2018
Source: LBNL Tracking the Sun data (full sample)

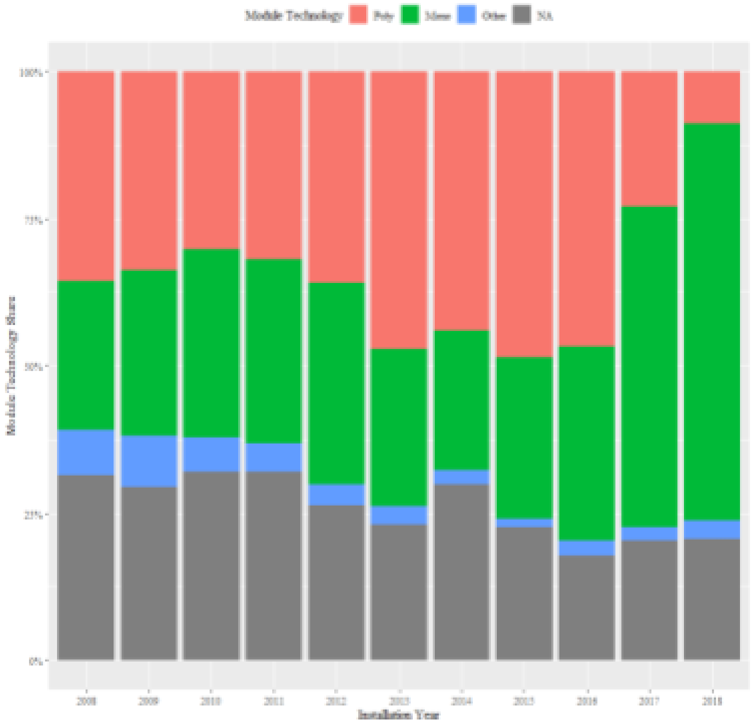

Figure SEQ Figure * ARABIC 3. Share of module technology over time
Source: LBNL Tracking the Sun data (full sample)

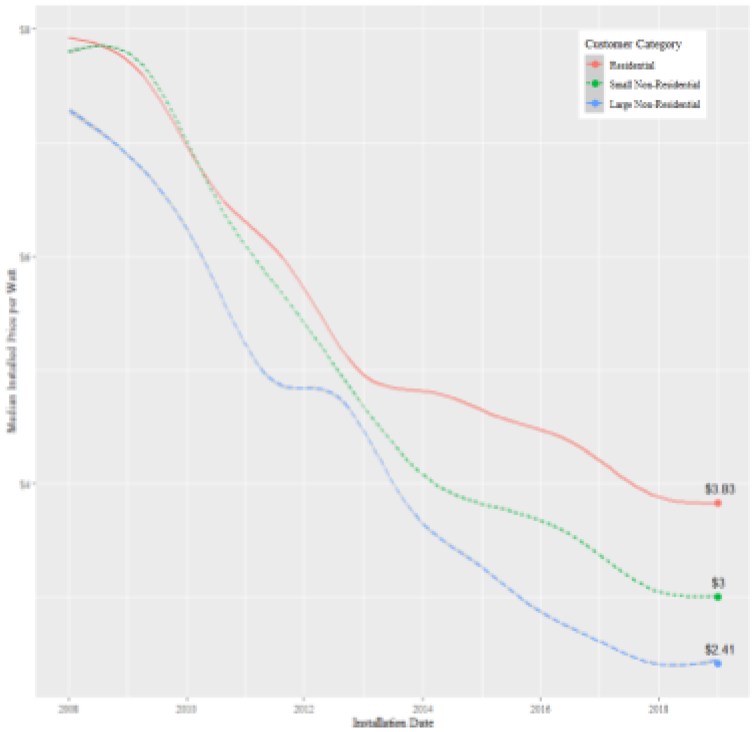

Figure SEQ Figure * ARABIC 4. Median installed price per watt over time by customer category, 2008 to 2018
Note: Lines fitted using generalised additive model (GAM) smoother, displaying .95 confidence interval.
Source: LBNL Tracking the Sun data (full sample)

A2. TTS Data: Exploratory Data Analysis

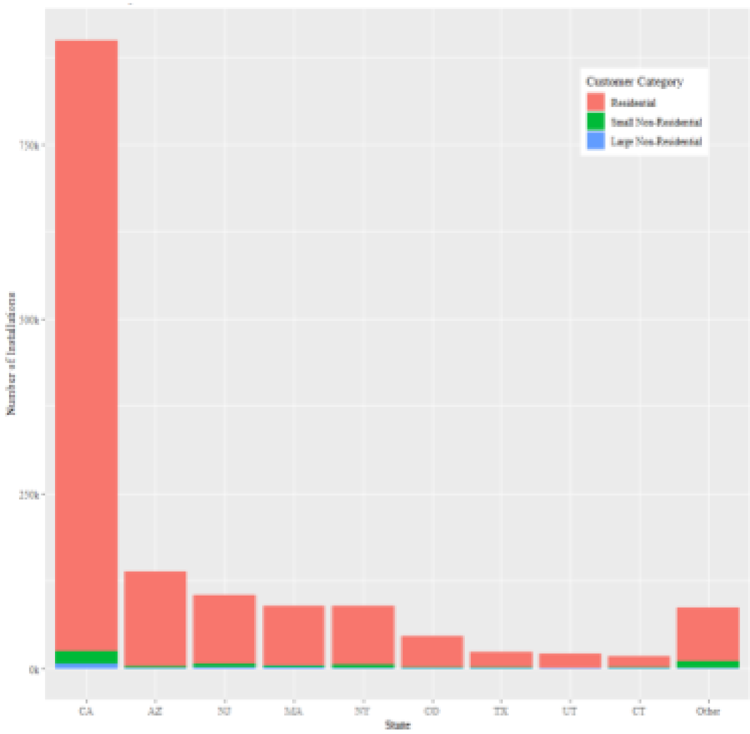

Figure SEQ Figure * ARABIC 5. PV installations by state and customer category in the full sample
Source: LBNL Tracking the Sun data (full sample)

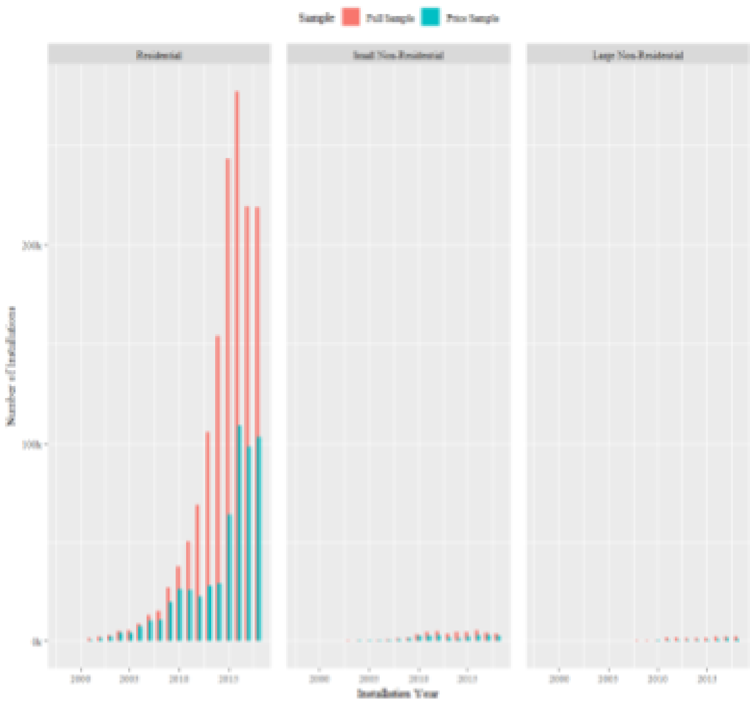

Figure SEQ Figure * ARABIC 6. PV installations by sample and customer category, 1998 to 2008
Source: LBNL Tracking the Sun data (full sample)

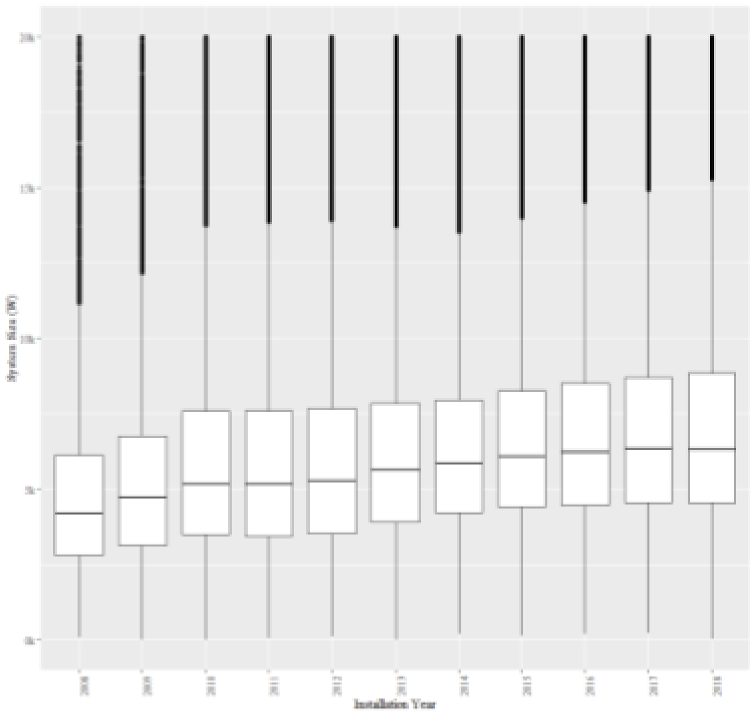

Figure SEQ Figure * ARABIC 7. System size development over time
Note: Only observations with system size (W) in [0, 20000] and installed after 2007 included. System size increases over time from about 4kW median size in 2008 to over 6kW in 2018. There are several outliers, particularly at the upper bound, most likely from non-residential systems.
Source: LBNL Tracking the Sun data (full sample)

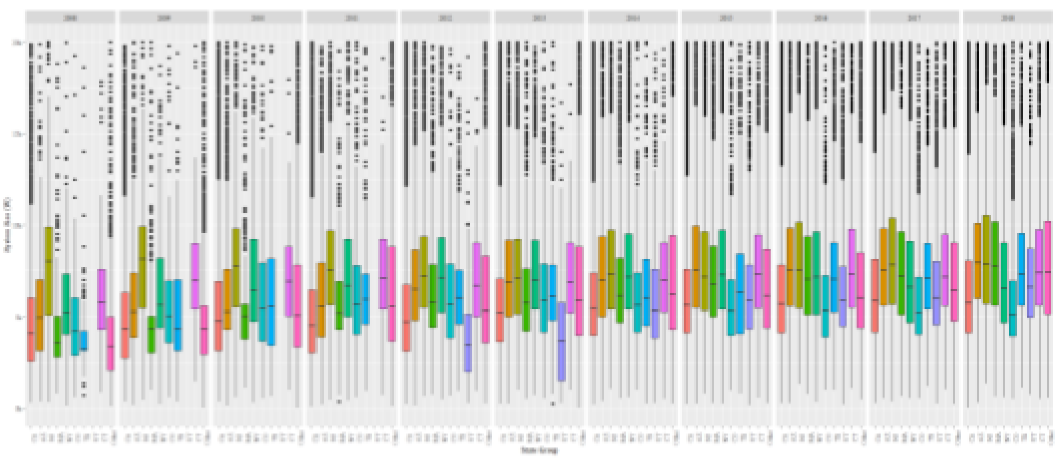

Figure SEQ Figure * ARABIC 8. System size by state over time, 2008 to 2018
Note: Only observations with system size (W) in [0, 20000] and installed after 2007 included. System size varies across states, especially in early years with fewer installations. Californian systems have a systematically lower median system size than most other states.
Source: LBNL Tracking the Sun data (full sample)

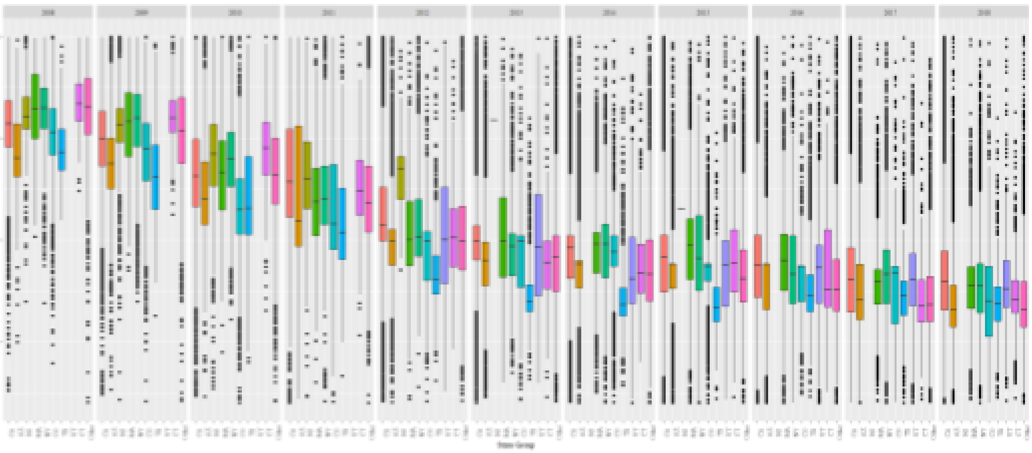

Figure SEQ Figure * ARABIC 9. Price per watt by state over time, 2008 to 2018 (w/o outliers)
Note: Only observations with price per watt in [0, 10] included.
Source: LBNL Tracking the Sun data (full sample)

A3. Estimation Data: Selection

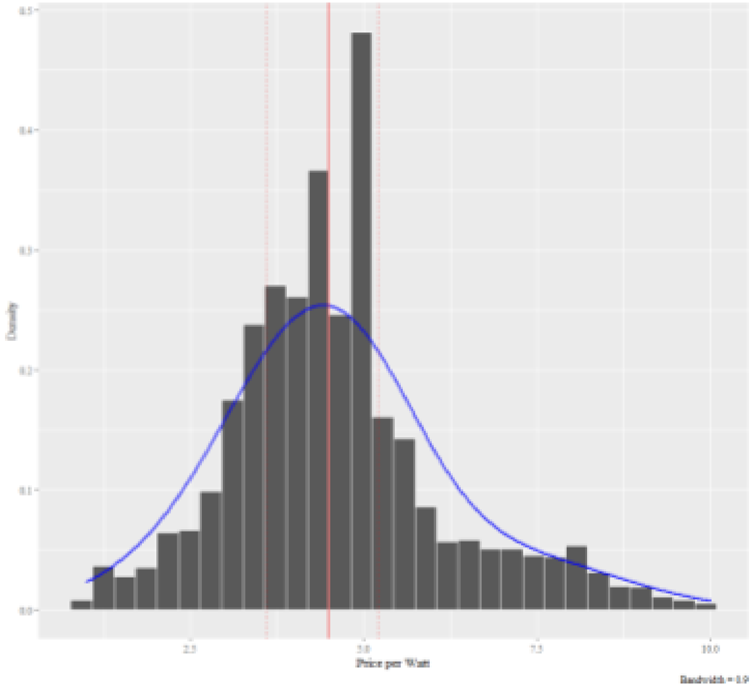

Figure SEQ Figure * ARABIC 10. Density plot of price per watt (w/o outliers)
Note: Only observations with price per watt in [0, 10] included. Density curve plotted with bandwidth .9 in blue. Median price shown by red solid line, .25 and .75 quantiles shown by red dotted lines, respectively.
Source: LBNL Tracking the Sun data (full sample)

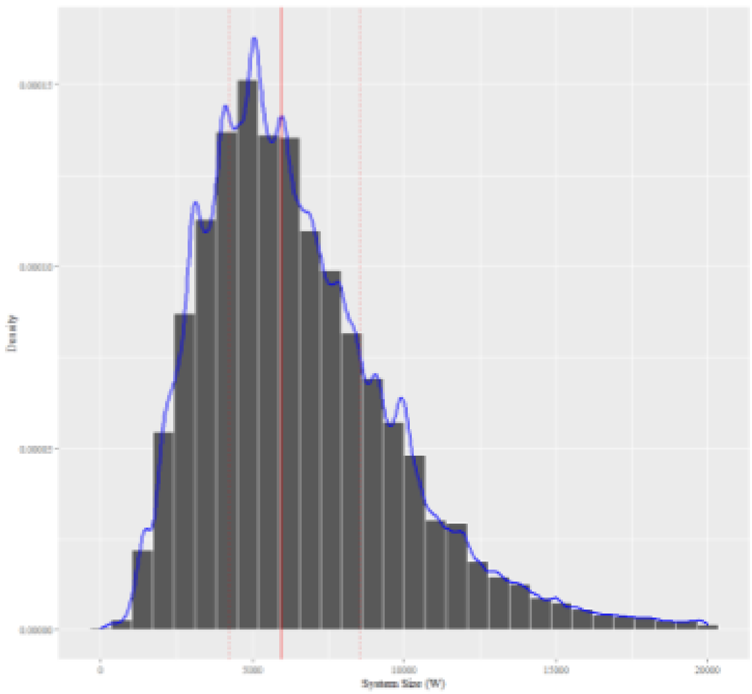

Figure SEQ Figure * ARABIC 11. Density plot of system size (w/o outliers)
Note: Only observations with system size (W) in [0, 20000] included. Density curve plotted in blue. Median size shown by red solid line, .25 and .75 quantiles shown by red dotted lines, respectively.
Source: LBNL Tracking the Sun data (full sample)

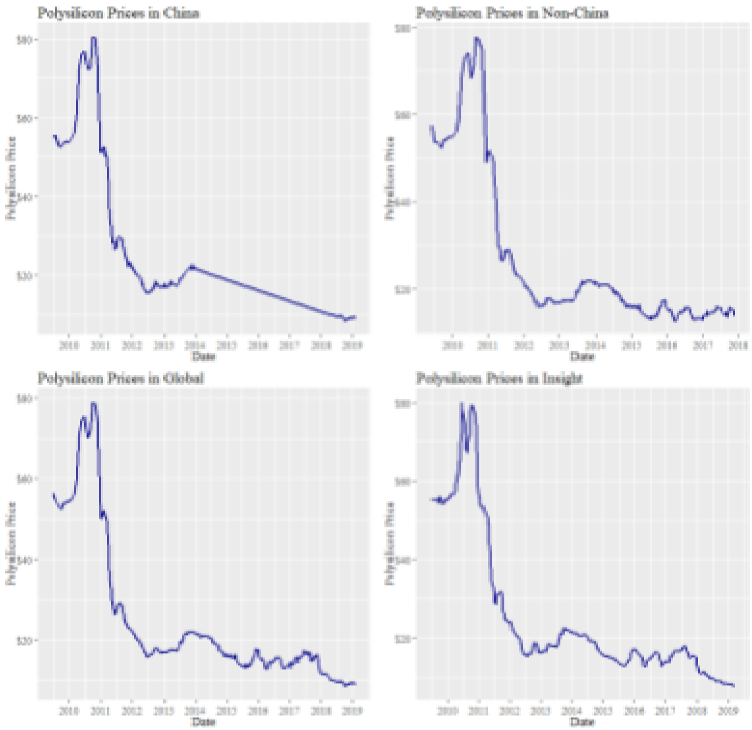

Figure SEQ Figure * ARABIC 12. Polysilicon prices over time for four data sets, 2010 to 2019
Source: Bloomberg, for indices SSPSPSNC (China), SSPSPSNI (Non-China), SSPFPSNO (Global), SOLRAPS (Insight)

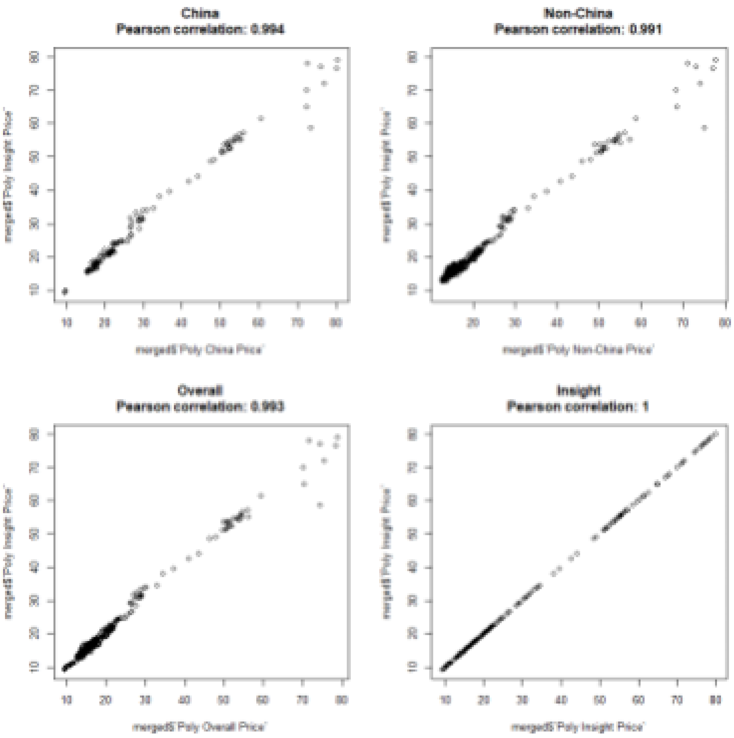

Figure SEQ Figure * ARABIC 13. Pearson correlation of polysilicon price data sets with PVinsights data
Source: Bloomberg, for indices SSPSPSNC (China), SSPSPSNI (Non-China), SSPFPSNO (Overall), SOLRAPS (Insight)

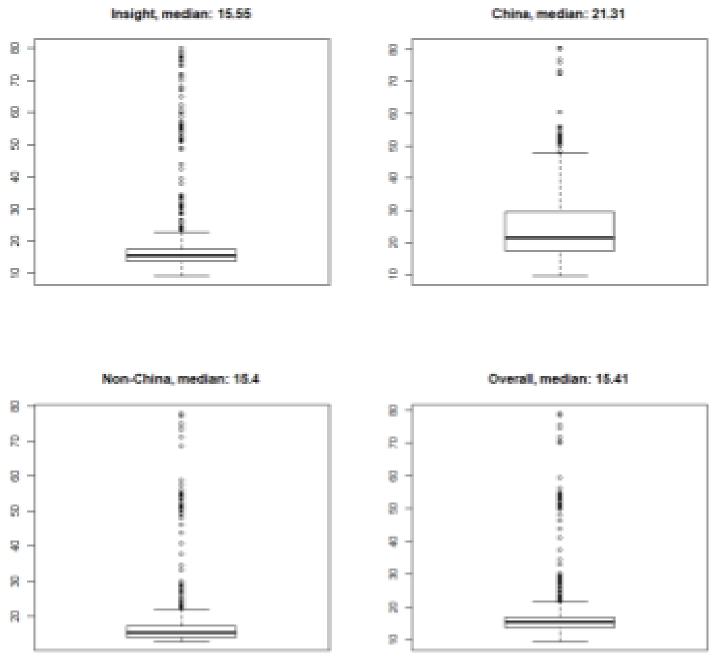

Figure SEQ Figure * ARABIC 14. Median polysilicon prices for four data sets
Source: Bloomberg, for indices SSPSPSNC (China), SSPSPSNI (Non-China), SSPFPSNO (Overall), SOLRAPS (Insight)

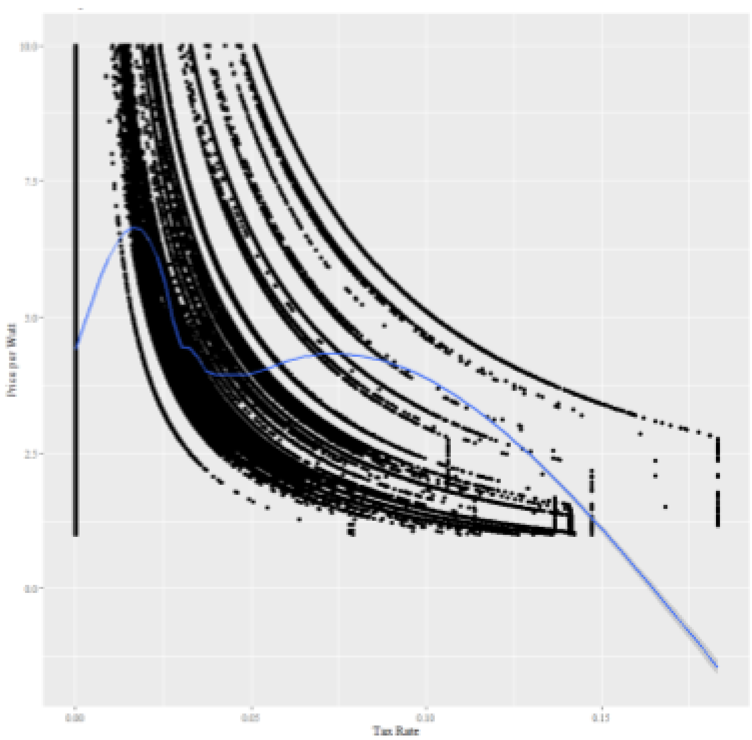

Figure SEQ Figure * ARABIC 15. Price per watt by tax rate
Note: Regression line plotted in blue, fitted using generalised additive model (GAM) smoother and displaying .95 confidence interval.
Source: LBNL Tracking the Sun data (estimation sample)

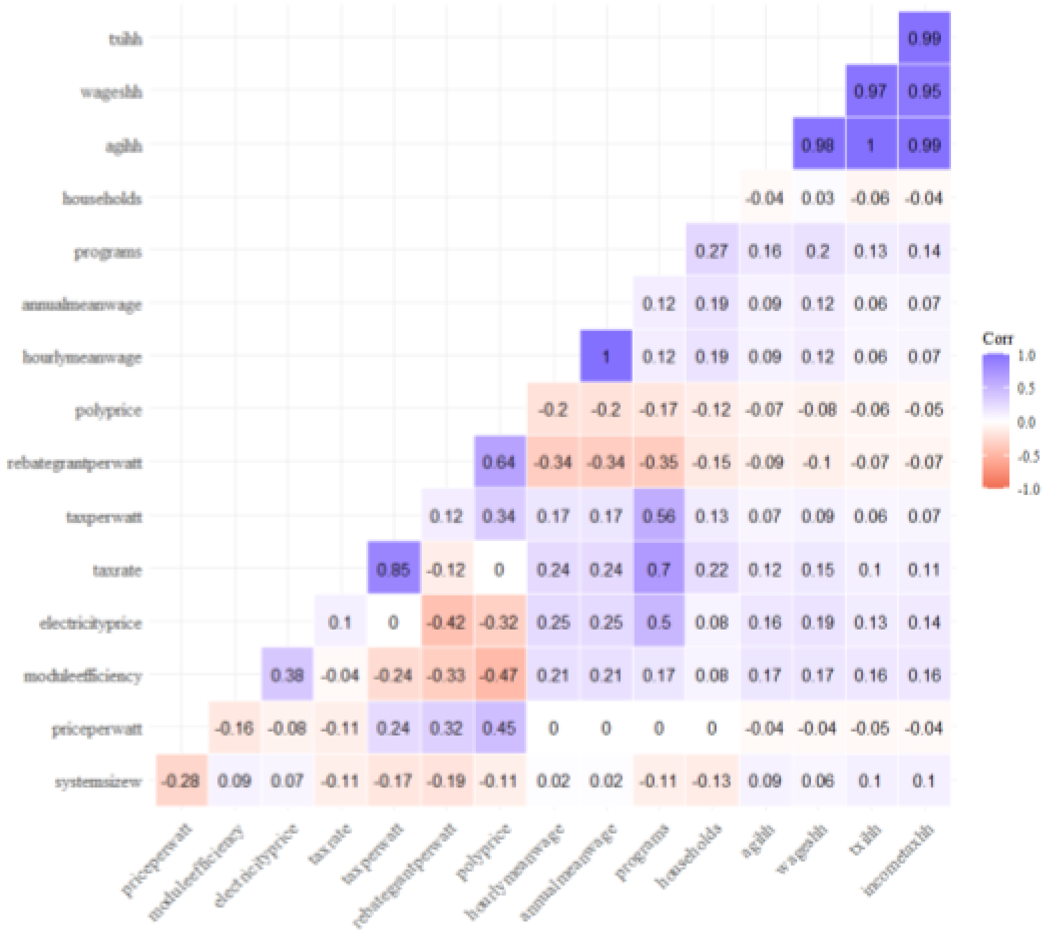

Figure SEQ Figure * ARABIC 16. Correlation matrix of continuous variables
Note: Pairwise Pearson correlation coefficients are calculated for relevant continuous variables. Some control variables as well as instrument data show high correlation, indicating that including all of them in the regression might result in collinearity problems.
Source: Own analysis, estimation sample

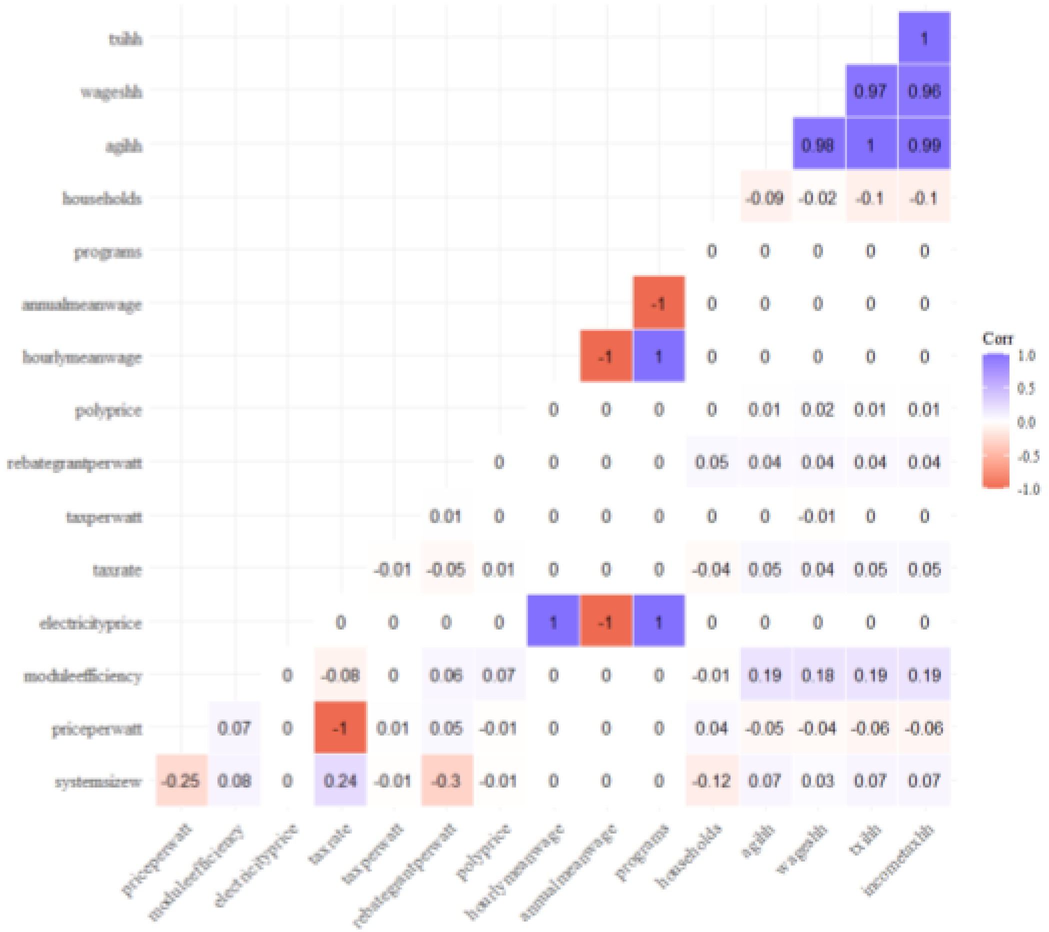

Figure SEQ Figure * ARABIC 17. Correlation matrix for continuous variables for California 2017
Note: Pairwise Pearson correlation coefficients are calculated for relevant continuous variables, holding state and year fixed. High correlation between some control variables and instruments becomes even more apparent, indicating that including all of them in the regression might result in collinearity problems. Some correlation coefficients are zero because there is no within-year variation.
Source: Own analysis, estimation sample

Data summary of estimation sample

	class	% missing	% zero	distinct	observations
systemsizew	numeric	0	0	12,031	501,394
priceperwatt	numeric	0	0	281,111	501,394
salestaxcost	numeric	0	28.50	35,421	501,390
taxperwatt	numeric	0	28.50	27,172	501,390
taxrate	numeric	0	28.50	279,052	501,390
rebateorgrant	numeric	7.86	53.40	26,904	462,001
rebategrantperwatt	numeric	7.86	53.40	75,477	462,001
performancebasedincentiveannualpayment	numeric	1.11	98.60	1,581	495,840
performancebasedincentivesduration	integer	1.11	98.60	6	495,840
feedintariffannualpayment	numeric	0	100	1	501,394
feedintariffduration	integer	0	100.00	2	501,394
moduleefficiency	numeric	23.40	0	1,380	384,229
polyinsightprice	numeric	0	0	334	501,394
polychinaprice	numeric	83.90	0	145	80,758
polynonchinaprice	numeric	21.60	0	305	393,244
polyoverallprice	numeric	7.03	0	341	466,167
polyprice	numeric	0	0	334	501,394
programs	integer	0	0	20	501,394
hourlymeanwage	numeric	16.50	0	72	418,809
annualmeanwage	numeric	16.50	0	75	418,809
households	numeric	35.70	0	8,320	322,524
population	numeric	35.70	0	10,896	322,524
agi	numeric	35.70	0	23,483	322,524
agihh	numeric	35.70	0	23,678	322,524
wages	numeric	35.70	0	23,424	322,524
wageshh	numeric	35.70	0	23,658	322,524
txi	numeric	35.70	0	23,406	322,524
txihh	numeric	35.70	0	23,656	322,524
incometax	numeric	35.70	0	22,029	322,524
incometaxhh	numeric	35.70	0	23,557	322,524
electricityprice	numeric	0	0	149	501,394

Table SEQ Table * ARABIC 3. Data summary of the estimation sample
Note: Some variables related to incentive programs, e.g., rebate, PBI, and feed-in tariff payments show a large share of zero-values. As these most likely stem from incomplete data rather than true zeros, these variables will probably be of little use for the estimations. The control variables exhibit over 35% of missing values, which will cause a substantial reduction in observations used for the estimation.
Source: Own analysis, estimation sample

A4. Estimation Data: Transformation

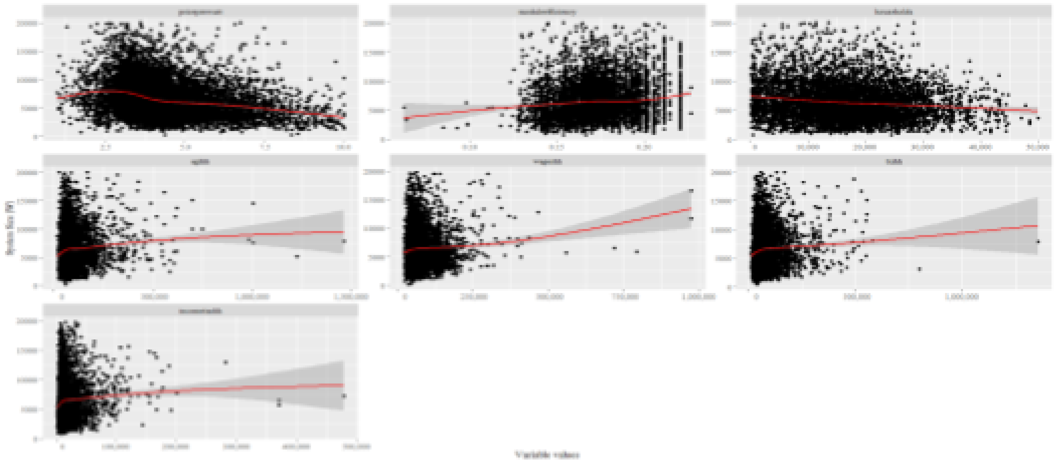

Figure SEQ Figure * ARABIC 18. Regression plots of system size (W) against selected variables
Note: Regression lines fitted using generalised additive model (GAM) smoother, displaying .95 confidence interval.
Source: Own analysis, 50,000 observations randomly sampled without replacement from estimation sample

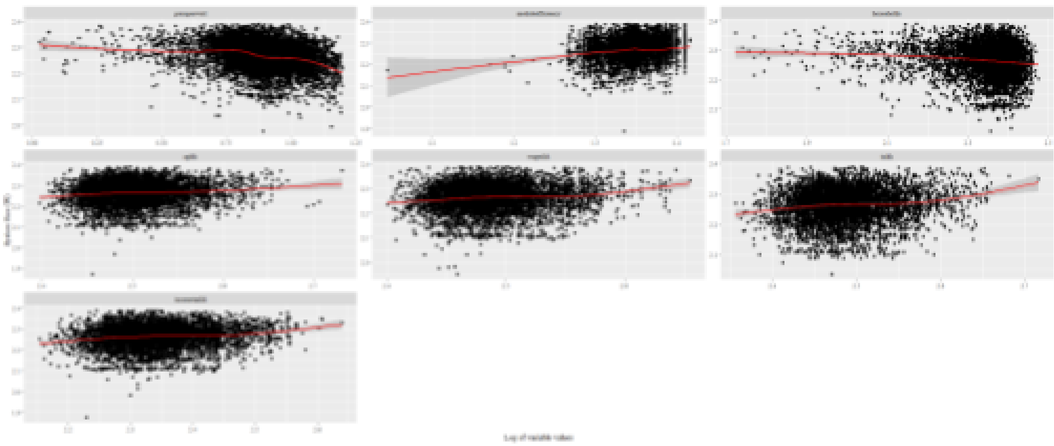

Figure SEQ Figure * ARABIC 19. Regression plots of system size (W) against selected variables after log-transformation
Note: Regression lines fitted using generalised additive model (GAM) smoother, displaying .95 confidence interval.
Source: Own analysis, 50,000 observations randomly sampled without replacement from estimation sample

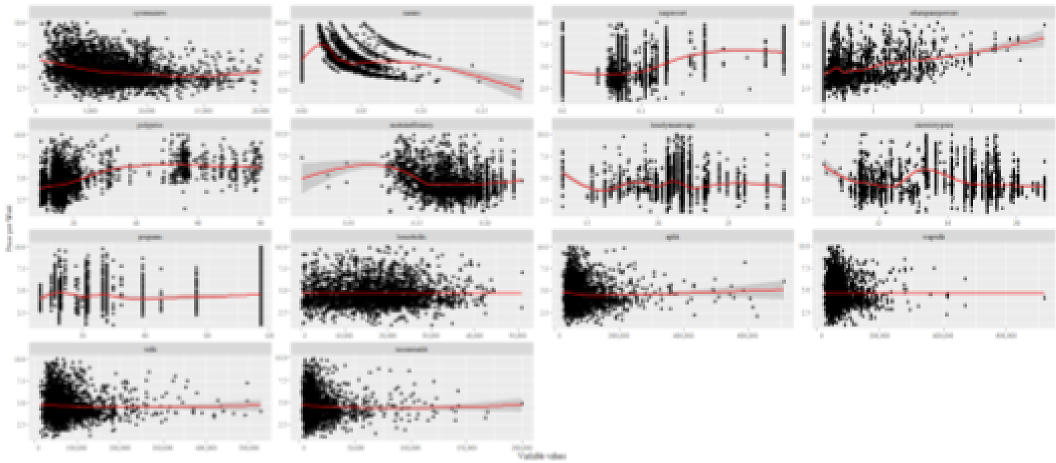

Figure SEQ Figure * ARABIC 20. Regression plots of price per watt against selected variables
Note: Regression lines fitted using generalised additive model (GAM) smoother, displaying .95 confidence interval.
Source: Own analysis, 50,000 observations randomly sampled without replacement from estimation sample

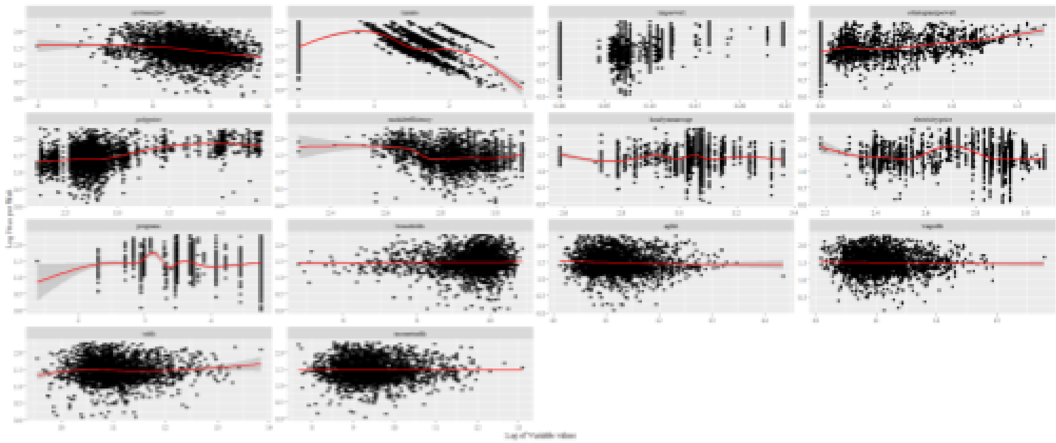

Figure SEQ Figure * ARABIC 21. Regression plots of price per watt against selected variables after log-transformation
Note: Regression lines fitted using generalised additive model (GAM) smoother, displaying .95 confidence interval.
Source: Own analysis, 50,000 observations randomly sampled without replacement from estimation sample

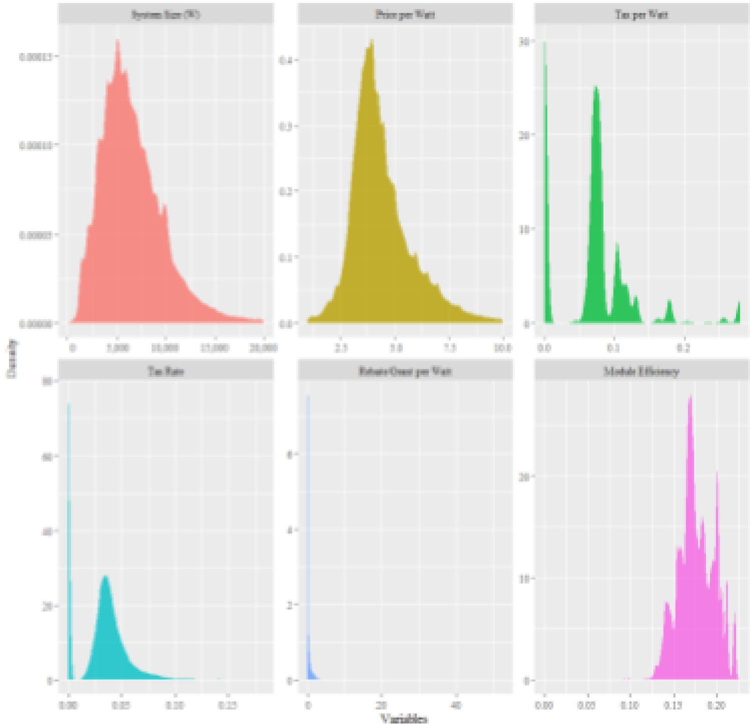

Figure SEQ Figure * ARABIC 22. Density plots of relevant variables
Note: System Size (W), Price per Watt, Tax Rate, Tax per Watt, and Rebate/Grant per Watt are right skewed. Module Efficiency is left skewed.
Source: Own analysis, estimation sample

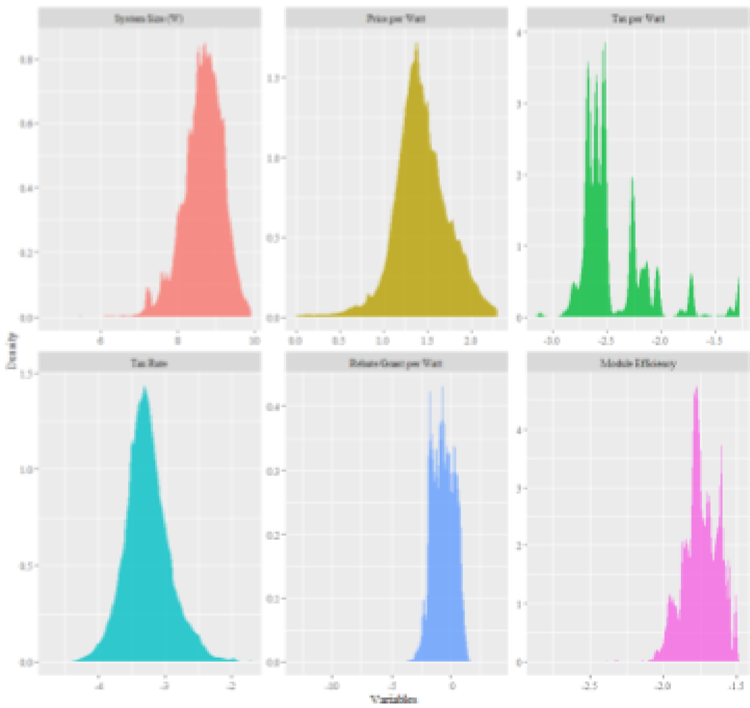

Figure SEQ Figure * ARABIC 23. Density plots of the log of relevant variables
Note: Taking the log relieves skewness in most of the variables.
Source: Own analysis, estimation sample

A5. Estimation Data: Summary

Data summary of final sample

	class	% missing	% zero	distinct	observations
systemsizew	numeric	0	0	8,447	172,106
priceperwatt	numeric	0	0.01	103,719	172,106
moduleefficiency	numeric	0	0	956	172,106
newconstruction	factor	0	91.70	2	172,106
tracking	factor	0	99.30	2	172,106
groundmounted	factor	0	96.60	2	172,106
moduletechnology	factor	0	0	3	172,106
mlpe	factor	0	0	3	172,106
installationyear	factor	0	0	8	172,106
stategroup	factor	0	0	4	172,106
polyprice	numeric	0	0	298	172,106
taxrate	numeric	0	0.51	131,655	172,106
households	numeric	0	0	3,496	172,106
agihh	numeric	0	0	6,332	172,106

Table SEQ Table * ARABIC 4. Data summary of final sample
Source: Own analysis, final sample

Summary statistics of final sample

	mean	median	sd	min	max
systemsizew	6293.55	5865.00	2981.76	265.00	20000.00
priceperwatt	4.32	4.10	1.28	1.00	10.00
moduleefficiency	0.18	0.17	0.02	0.06	0.23
polyprice	16.81	15.24	8.24	12.65	80.00
taxrate	0.04	0.04	0.01	0.00	0.18
households	18231.24	17990.00	8514.05	100.00	49920.00
agihh	92366.61	74207.27	68125.12	22215.54	1718304.63

Table SEQ Table * ARABIC 5. Summary statistics for numeric variables of the final sample
Note: Original data, not log-transformed.
Source: Own analysis, final sample

Unique factor values of final sample

	values
newconstruction	0, 1
tracking	0, 1
groundmounted	0, 1
moduletechnology	Poly, Mono, Other
mlpe	None, Microinverter, DCOptimizer
installationyear	2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017
stategroup	AZ, CA, TX, Other

Table SEQ Table * ARABIC 6. Unique values for factor variables of the final sample
Source: Own analysis, final sample

Observations in final sample by Installation Year

	Installation Year	Installations	Share of total (%)
1	2017	58,099	33.80
2	2016	64,376	37.40
3	2015	30,684	17.80
4	2014	4,158	2.42
5	2013	4,977	2.89
6	2012	3,981	2.31
7	2011	3,054	1.77
8	2010	2,777	1.61

Observations in final sample by State

	State Group	Installations	Share of total (%)
1	CA	170,900	99.30
2	Other	721	0.42
3	TX	326	0.19
4	AZ	159	0.09

Observations in final sample by New Construction

	New Construction	Installations	Share of total (%)
1	0	157,810	91.70
2	1	14,296	8.31

Observations in final sample by Tracking

	Tracking	Installations	Share of total (%)
1	0	170,889	99.30
2	1	1,217	0.71

Observations in final sample by Ground-Mounting

	Ground-Mounted	Installations	Share of total (%)
1	0	166,286	96.60
2	1	5,820	3.38

Observations in final sample by Module Technology

	Module Technology	Installations	Share of total (%)
1	Mono	126,015	73.20
2	Poly	42,854	24.90
3	Other	3,237	1.88

Observations in final sample by MLPE

	MLPE	Installations	Share of total (%)
1	Microinverter	77,171	44.80
2	None	47,925	27.90
3	DCOptimizer	47,010	27.30

Table SEQ Table * ARABIC 7. Distribution and share of installations by factor variables in the final sample

A6. Regression Results

	First Stage Estimation Results						
	Dependent Variable: Price per Watt						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Polysilicon Price	0.052***						0.059***
	(0.002)						(0.002)
Tax Rate		-0.597***					-0.597***
		(0.003)					(0.003)
Tax per Watt			0.115***				
			(0.014)				
Rebate/Grant per Watt					0.022***		
					(0.002)		
Incentive Programs						0.023***	
						(0.003)	
Hourly Mean Installer Wage				0.045***			
				(0.008)			
Constant	1.690***	3.160***	1.880***	1.560***	1.890***	1.790***	2.920***
	(0.010)	(0.006)	(0.003)	(0.023)	(0.002)	(0.012)	(0.011)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	501,394	501,390	501,390	418,809	462,001	501,394	501,390
R^2	0.270	0.526	0.269	0.106	0.281	0.269	0.527
Adjusted R^2	0.270	0.526	0.269	0.106	0.281	0.269	0.527
Residual Std. Error	0.262	0.211	0.262	0.265	0.260	0.262	0.211
F Statistic	10,288.000***	30,942.000***	10,258.000***	3,115.000***	10,020.000***	10,261.000***	29,411.000***
Note:							*p<0.05; **p<0.01; ***p<0.001

Table SEQ Table * ARABIC 8. First stage results for simple linear regression of selected instruments on Price per Watt
Note: Individual OLS regressions of potential instruments to evaluate their predictive power. Tax rate shows by far the highest coefficient of determination (.526), the combination with polysilicon price improves it very slightly.
Source: Own analysis, estimation sample

	Instrumental Variable Estimation Results						
	Dependent Variable: System Size (W)						
	IV: Polysilicon Price	IV: Tax Rate	IV: Tax per Watt	IV: Hourly Mean Installer Wage	IV: Rebate/Grant per Watt	IV: Incentive Programs	IVs: Polysilicon Price, Tax Rate
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Price per Watt	-0.605***	-0.478***	5.500***	-3.130***	-28.800***	-3.560***	-0.479***
	(0.088)	(0.005)	(0.789)	(0.552)	(2.440)	(0.389)	(0.005)
Constant	9.620***	9.380***	-1.970	13.700***	63.400***	15.200***	9.380***
	(0.167)	(0.010)	(1.500)	(0.935)	(4.660)	(0.739)	(0.009)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	501,394	501,390	501,390	418,809	462,001	501,394	501,390
R^2	0.119	0.126	-9.290	-1.940	-212.000	-2.480	0.126
Adjusted R^2	0.119	0.126	-9.290	-1.940	-212.000	-2.480	0.126
Residual Std. Error	0.476	0.474	1.630	0.860	7.380	0.946	0.474
Note:							*p<0.05; **p<0.01; ***p<0.001

Table SEQ Table * ARABIC 9. IV regression results for simple linear regression of selected instruments on System Size
Note: Individual TSLS regressions using potential instruments. The best results in terms of R^2 and low residual standard error provide the regressions using tax rate as well as tax rate and polysilicon price as instruments.
Source: Own analysis, estimation sample

	First Stage Estimation Results					
	Dependent Variable: Price per Watt					
	(1)	(2)	(3)	(4)	(5)	(6)
Polysilicon Price	0.060***	-0.004***	-0.004***	-0.009***	-0.009***	-0.004***
	(0.006)	(0.001)	(0.001)	(0.001)	(0.001)	(0.0004)
Tax Rate		-1.260***	-1.260***	-1.260***	-1.260***	-1.260***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Tax per Watt						6.680***
						(0.148)
Rebate/Grant per Watt				0.075***	0.075***	0.010***
				(0.005)	(0.005)	(0.001)
Hourly Mean Installer Wage						-0.044
						(0.088)
Incentive Programs						0.013
						(0.033)
Module Efficiency	-0.254***	-0.008***	-0.015***	-0.021***	-0.021***	-0.025***
	(0.012)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Dummy: New Construction	-0.141***	-0.006***	-0.006***	-0.049***	-0.048***	-0.006***
	(0.003)	(0.0005)	(0.0005)	(0.003)	(0.003)	(0.0005)
Dummy: Tracking	-0.023*	0.005***	0.005***	0.005***	0.005***	0.003***
	(0.010)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Dummy: Ground-mounted	0.006	-0.001*	-0.001*	-0.001	-0.001	-0.001***
	(0.004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0002)
Dummy: Premium Module	0.147***	-0.002***		0.002***	0.002***	0.002***
	(0.002)	(0.0004)		(0.0003)	(0.0003)	(0.0002)
Module Technology: Mono	0.088***	0.001***	0.002***	0.001***	0.001***	0.0003**
	(0.002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0001)
Module Technology: Other	0.131***	-0.0004	0.0002	-0.0002	-0.0002	-0.0001
	(0.005)	(0.001)	(0.001)	(0.001)	(0.001)	(0.0003)
MLPE: DC Optimizer	-0.033***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.002)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
MLPE: None	-0.034***	0.003***	0.003***	0.001***	0.001***	0.002***
	(0.002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0001)
Electricity Price	1.150***					
	(0.326)					
Households	0.017***	0.001***	0.001***	0.001***	0.001***	0.0003***
	(0.001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
AGI/Household	-0.024***	-0.002***	-0.002***	0.002	-0.001***	0.006***
	(0.001)	(0.0001)	(0.0001)	(0.002)	(0.0001)	(0.001)
Wages/Household				0.004***		0.004***
				(0.001)		(0.0005)
Taxable Income/Household				-0.013***		-0.016***
				(0.002)		(0.001)
Income Tax/Houshold				0.006***		0.005***
				(0.001)		(0.001)
Constant	-0.646	4.640***	4.660***	4.640***	4.620***	3.160***
	(0.879)	(0.006)	(0.005)	(0.008)	(0.007)	(0.115)
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	0.188	0.992	0.992	0.992	0.992	0.997
Adjusted R^2	0.188	0.992	0.992	0.992	0.992	0.997
Residual Std. Error	0.266	0.026	0.026	0.026	0.026	0.015
F Statistic	1,737.000***	930,595.000***	972,478.000***	832,343.000***	936,141.000***	2,157,171.000***
Note:						$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

Table SEQ Table * ARABIC 10. First stage regression results for the final model specification and different robustness checks

Note: First stage OLS regression results of instruments and explanatory variables on price per watt, using five different sets of instruments and four different sets of regressors. The combination of model (3) is selected for the final model as it shows the highest R^2 (.992) and lowest standard errors while there is no multicollinearity in the explanatory or control variables

	Alternative specifications: First Stage Estimation Results				
	Dependent Variable: Price per Watt				
	IV: Baseline (1)	IV: Polynomial in first stage (2)	IV: Interaction state and year (3)	IV: Quarterly fixed effects (4)	IV: No fixed effects (5)
Polysilicon Price	-0.004^{***}	-0.003^{**}	-0.004^{***}	-0.007^{***}	0.457^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)
Tax Rate	-1.260^{***}	-1.520^{***}	-1.270^{***}	-1.260^{***}	-0.846^{***}
	(0.001)	(0.014)	(0.001)	(0.001)	(0.006)
Tax Rate^2		0.078^{***}			
		(0.004)			
Module Efficiency	-0.015^{***}	-0.006^{***}	-0.015^{***}	-0.015^{***}	-0.236^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.006)
Dummy: New Construction	-0.006^{***}	-0.006^{***}	-0.005^{***}	-0.006^{***}	-0.042^{***}
	(0.0005)	(0.001)	(0.0005)	(0.001)	(0.002)
Dummy: Tracking	0.005^{***}	0.003^{***}	0.005^{***}	0.005^{***}	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.005)
Dummy: Ground-mounted	-0.001^{*}	-0.0005	-0.001^{*}	-0.001^{*}	-0.031^{***}
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.003)
Module Technology: Mono	0.002^{***}	0.002^{***}	0.001^{***}	0.002^{***}	0.038^{***}
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.001)
Module Technology: Other	0.0002	0.001	0.0001	0.001	0.025^{***}
	(0.001)	(0.0005)	(0.0004)	(0.001)	(0.003)
MLPE: DC Optimizer	-0.001^{***}	0.0001	-0.001^{***}	-0.001^{***}	-0.043^{***}
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.001)
MLPE: None	0.003^{***}	0.002^{***}	0.002^{***}	0.003^{***}	0.082^{***}
	(0.0002)	(0.0002)	(0.0001)	(0.0002)	(0.001)
Households	0.001^{***}	0.001^{***}	0.001^{***}	0.001^{***}	0.011^{***}
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.001)
AGI/Household	-0.002^{***}	-0.001^{***}	-0.001^{***}	-0.001^{***}	-0.002^{*}
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.001)
Constant	4.660^{***}	4.820^{***}	4.680^{***}	4.660^{***}	2.060^{***}
	(0.005)	(0.011)	(0.004)	(0.007)	(0.024)
Time FE	Year	Year	Year	Quarter	None
State FE	Yes	Yes	Yes	Yes	No
Interaction	No	No	Yes	No	No
Polynomials	No	Yes	No	No	No
Instruments	Polysilicon Price, Tax Rate	Polysilicon Price, Tax Rate	Polysilicon Price, Tax Rate	Polysilicon Price, Tax Rate	Polysilicon Price, Tax Rate
Observations	172,106	172,106	172,106	172,106	172,106
R^2	0.992	0.993	0.993	0.992	0.653
Adjusted R^2	0.992	0.993	0.993	0.992	0.653
Residual Std. Error	0.026	0.025	0.024	0.026	0.174
F Statistic	$972,478.000^{***}$	$1,041,409.000^{***}$	$726,126.000^{***}$	$467,302.000^{***}$	$26,986.000^{***}$
Note:					$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

Table SEQ Table * ARABIC 11. First stage results of robustness checks for the final model against alternative specifications

Note: First stage OLS regression results of instruments and explanatory variables on price per watt. 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

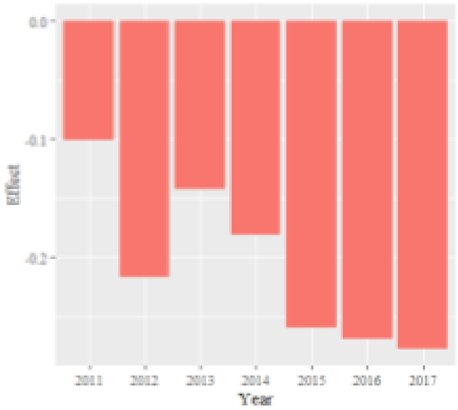

Figure SEQ Figure * ARABIC 24. Year fixed effects of final model
Note: Effects are provided relative to the baseline year 2010. They capture changes over time that are common for all states.
Source: Own analysis, final sample

Figure SEQ Figure * ARABIC 25. State fixed effects of final model
Note: Effects are provided relative to the baseline state California. They capture differences across states that are constant over time.
Source: Own analysis, final sample

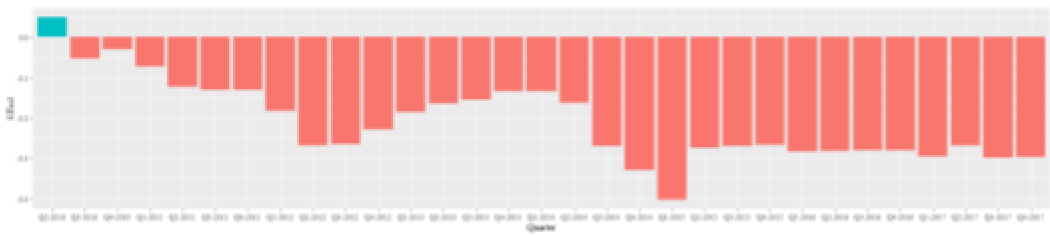

Figure SEQ Figure * ARABIC 26. Quarterly fixed effects of alternative specification
Note: Effects are provided relative to the baseline quarter Q1 in 2010. They capture changes over time that are common for all states.
Source: Own analysis, final sample

Median installed price per watt and tax rate by state

	State	Median price per watt	Median tax rate
1	AZ	6.31	0.00
2	TX	4.99	0.06
3	Other	4.71	0.00
4	CA	4.09	0.04

Table SEQ Table * ARABIC 12. Median price per watt and corresponding median tax rate by state
Source: Own analysis, estimation sample

Variance inflation factor

	GVIF	Df	GVIF^(1/(2*Df))
priceperwatt	1.208	1	1.099
moduleefficiency	1.649	1	1.284
newconstruction	1.227	1	1.108
tracking	1.005	1	1.003
groundmounted	1.046	1	1.023
moduletechnology	1.523	2	1.111
mlpe	1.432	2	1.094
installationyear	1.841	7	1.045
stategroup	1.080	3	1.013
households	1.056	1	1.028
agihh	1.057	1	1.028

Instruments: Polysilicon Price, Tax Rate

Table SEQ Table * ARABIC 13. Variance inflation factor for the final model
Note: VIF values below 5 indicate absence of severe multicollinearity problems, values exceeding 5 or 10 show a collinearity issue (p. 101, ADDIN EN.CITE <EndNote><Cite AuthorYear="1"><Author>James</Author><Year>2013</Year><RecNum>88</RecNum><DisplayText>James, Witten, Hastie, and Tibshirani (2013)</DisplayText><record><rec-number>88</rec-number><foreign-keys><key app="EN" db-id="vt2tf2wr5v0senewzsapxtf3zr0eavfxvfds" timestamp="1594985146">88</key></foreign-keys><ref-type name="Book">6</ref-type><contributors><authors><author>James, Gareth</author><author>Witten, Daniela</author><author>Hastie, Trevor</author><author>Tibshirani, Robert</author></authors></contributors><titles><title>An introduction to statistical learning</title></titles><volume>112</volume><dates><year>2013</year></dates><publisher>Springer</publisher><urls></urls></record></Cite>

	Income groups: Instrumental Variable Estimation Results			
	Dependent Variable: System Size (W)			
	IV: Low income	IV: Low/Medium income	IV: Medium/High income	IV: High income
	(1)	(2)	(3)	(4)
Price per Watt	-0.521***	-0.455***	-0.327***	-0.195**
	(0.011)	(0.006)	(0.016)	(0.071)
Module Efficiency	0.022	0.222***	0.744***	0.776***
	(0.026)	(0.013)	(0.035)	(0.157)
Dummy: Tracking	-0.182***	-0.003	-0.047	0.519***
	(0.035)	(0.016)	(0.051)	(0.110)
MLPE: DC Optimizer	0.081***	0.096***	0.159***	0.088*
	(0.005)	(0.003)	(0.008)	(0.036)
MLPE: None	-0.059***	-0.103***	0.046***	0.118*
	(0.006)	(0.003)	(0.009)	(0.046)
Households	-0.070***	-0.076***	-0.125***	-0.088*
	(0.003)	(0.002)	(0.006)	(0.044)
AGI/Household	0.364***	-0.031***	0.190***	-0.173**
	(0.014)	(0.004)	(0.013)	(0.063)
Constant	6.260***	9.970***	6.040***	10.400***
	(0.167)	(0.060)	(0.195)	(1.150)
Year FE	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes
Instruments	Polysilicon Price, Tax Rate	Polysilicon Price, Tax Rate	Polysilicon Price, Tax Rate	Polysilicon Price, Tax Rate
Observations	46,552	151,378	16,949	745
R^2	0.143	0.101	0.111	0.112
Adjusted R^2	0.143	0.101	0.110	0.094
Residual Std. Error	0.478 (df = 46533)	0.479 (df = 151359)	0.451 (df = 16930)	0.431 (df = 729)
Note:				*p<0.05; **p<0.01; ***p<0.001

Table SEQ Table * ARABIC 14. IV regression results for four different income groups
Note: Observations are grouped by adjusted gross income per household (AGI/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].
Source: Own analysis, estimation sample

	Population density groups: Instrumental Variable Estimation Results		
	Dependent Variable: System Size (W)		
	IV: Low population density	IV: Medium population density	IV: High population density
	(1)	(2)	(3)
Price per Watt	-0.473^{***}	-0.443^{***}	-0.338^{***}
	(0.072)	(0.018)	(0.005)
Module Efficiency	0.677^{***}	0.417^{***}	0.227^{***}
	(0.135)	(0.043)	(0.013)
Dummy: Tracking	0.141	-0.040	-0.085^{***}
	(0.098)	(0.055)	(0.014)
Dummy: Ground-mounted	0.284^{***}	0.341^{***}	0.450^{***}
	(0.027)	(0.011)	(0.007)
MLPE: DC Optimizer	0.051	0.084^{***}	0.084^{***}
	(0.030)	(0.009)	(0.003)
MLPE: None	-0.011	-0.026^{**}	-0.156^{***}
	(0.032)	(0.010)	(0.003)
Households	-0.043	-0.055^{***}	-0.089^{***}
	(0.029)	(0.007)	(0.003)
AGI/Household	0.134^{***}	0.109^{***}	0.043^{***}
	(0.031)	(0.006)	(0.003)
Constant	6.338^{***}	7.502^{***}	9.079^{***}
	(0.486)	(0.137)	(0.053)
Year FE	Yes	Yes	Yes
State FE	Yes	Yes	Yes
Instruments	Polysilicon Price, Tax Rate	Polysilicon Price, Tax Rate	Polysilicon Price, Tax Rate
Observations	1,182	14,185	158,243
R^2	0.178	0.157	0.131
Adjusted R^2	0.166	0.156	0.131
Residual Std. Error	0.426 (df = 1164)	0.452 (df = 14165)	0.466 (df = 158223)
Note:			*p<0.05; **p<0.01; ***p<0.001

Table SEQ Table * ARABIC 15. IV regression results for three different groups of population density
Note: Observations are grouped by the number of households. Equal intervals in thousands are (1) low: (0.09,18.1], (2) medium: (18.1,36.2], (3) high: (36.2,54.3].
Source: Own analysis, estimation sample

	State groups: Instrumental Variable Estimation Results		
	Dependent Variable: System Size (W)		
	IV: California	IV: Arizona	IV: Texas
	(1)	(2)	(3)
Price per Watt	-0.443***	-1.270	-0.481***
	(0.004)	(0.726)	(0.021)
Module Efficiency	0.407***	-0.081	1.520***
	(0.012)	(0.087)	(0.054)
Dummy: New Construction	-0.889***	-0.729***	
	(0.005)	(0.024)	
Dummy: Tracking	-0.104***		
	(0.014)		
Dummy: Ground-mounted	0.372***		
	(0.006)		
Module Technology: Mono	0.049***	-0.027	-0.169***
	(0.003)	(0.079)	(0.011)
Module Technology: Other	0.091***	0.008	-0.321***
	(0.007)	(0.185)	(0.032)
MLPE: DC Optimizer	0.053***		0.189***
	(0.002)		(0.011)
MLPE: None	-0.018***		0.184***
	(0.003)		(0.018)
Households	-0.043***	0.042*	-0.048***
	(0.002)	(0.018)	(0.008)
AGI/Household	0.061***	0.086*	0.147***
	(0.002)	(0.036)	(0.011)
Constant	8.090***	9.640***	4.080***
	(0.043)	(1.800)	(0.220)
Year FE	Yes	Yes	Yes
State FE	No	No	No
Instruments	Polysilicon Price, Tax Rate	Polysilicon Price, Tax Rate	Polysilicon Price, Tax Rate
Observations	170,900	3,722	9,167
R^2	0.332	0.500	0.285
Adjusted R^2	0.332	0.498	0.284
Residual Std. Error	0.409 (df = 170881)	0.358 (df = 3708)	0.400 (df = 9151)
Note:			*p<0.05; **p<0.01; ***p<0.001

Table SEQ Table * ARABIC 16. IV regression results by state

Note: For Arizona and Texas, some explanatory factor variables did not provide any variation, such that they were excluded from the estimation for the respective subsamples.

Source: Own analysis, estimation sample

A7. Model Evaluation

Diagnostic tests

	df1	df2	statistic	p-value
Weak instruments	2	172,083	8,855,961	<0.001
Wu-Hausman	1	172,083	1,929	<0.001
Sargan	1		0.510	0.475

Instruments: Polysilicon Price, Tax Rate

Table SEQ Table * ARABIC 17. Diagnostic test results for final model
Note: Test on instrument relevance (Weak instruments), exogeneity of regressors (Wu-Hausman), and exogeneity of instruments (Sargan), including the test statistics and p-values. The instruments used are not weak and the regressors are endogenous while there is no evidence to assume endogeneity in the instruments.
Source: Own analysis, final sample

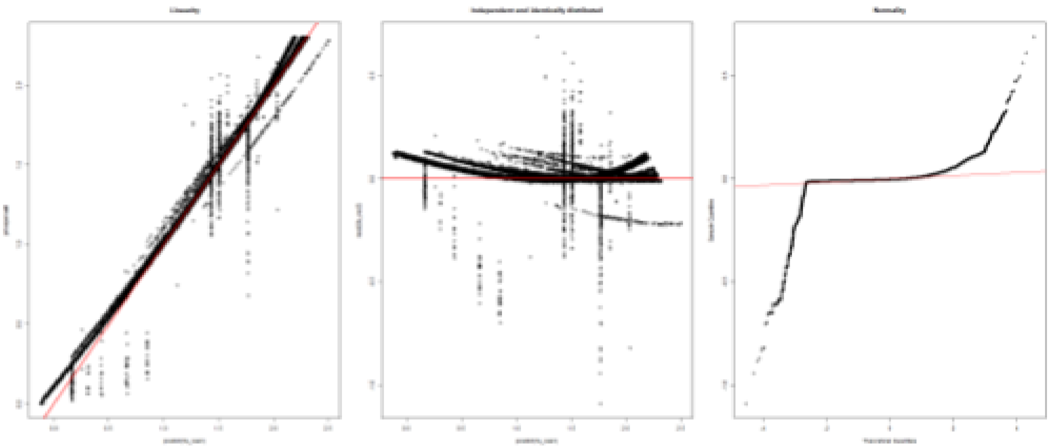

Figure SEQ Figure * ARABIC 27. Test of OLS assumptions for the first stage of the final model
Note: Plots to evaluate linearity assumption (observed vs. predicted values), i.i.d. assumption/ homoskedasticity (residuals vs. predicted values), and normality assumption (sample vs. theoretical quantiles) in the first stage.
Source: Own analysis, final sample

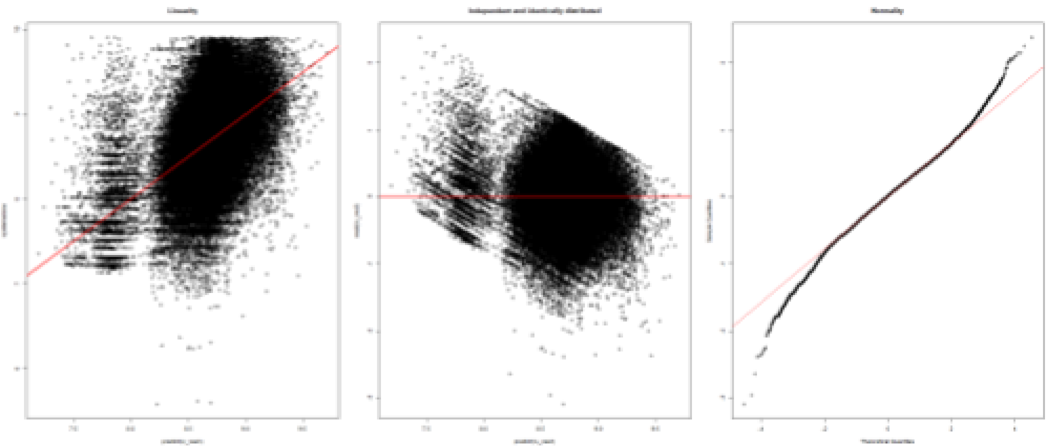

Figure SEQ Figure * ARABIC 28. Test of OLS assumptions for the second stage of the final model
Note: Plots to evaluate linearity assumption (observed vs. predicted values), i.i.d. assumption/ homoskedasticity (residuals vs. predicted values), and normality assumption (sample vs. theoretical quantiles) in the second stage.
Source: Own analysis, final sample

Variable overview from combined estimation data set

Name	Abbreviation	Description	Data source
System size (W)	systemsizew	The total rated direct-current (DC) output of the module arrays at standard test conditions.	LNBL
Price per watt	priceperwatt	The installed price per watt installed, prior to receipt of any incentives.	LNBL
Installation date	installationdate	The day the system was installed. For some data providers, the installation date may be based on the best available proxy, such as the date that an incentive claim was submitted or when the inspection was performed.	LNBL
Installation year	installationyear	The year of installation, extracted from the installation date.	LNBL
Installation week	installationweek	The week of installation, extracted from the installation date.	LNBL
State	stategroup	The U.S. state the system is installed in. Rare states with less than 20,001 observations were grouped to ‘other’ states.	LNBL
Zip code	zip	The 5-digit zip code the system is installed in.	LNBL
Sales tax cost	salestaxcost	The calculated cost of sales taxes. This is estimated based on average sales tax rates for the given state and year, accounting for any sales tax exemptions that may exist for PV systems. Sales taxes, if applicable, are assumed to be levied only on hardware costs, which are assumed to represent 55% of the total installed price.	LNBL
Tax per watt	taxperwatt	The sales tax cost per watt.	LNBL
Tax rate	taxrate	The sales tax rate applied on the installed price.	LNBL
Rebate or grant	rebateorgrant	The pre-tax value of any up-front rebate or grant provided by the entity supplying the data.	LNBL
Rebate or grant per watt	rebateorgrantperwatt	The rebate or grant provided per watt.	LNBL
New construction	newconstruction	Indicates if the system was installed at the time of building construction.	LNBL
Tracking	tracking	Indicates if the system includes tracking equipment.	LNBL

Name	Abbreviation	Description	Data source
Ground-mounted	groundmounted	Indicates if the system is ground-mounted (which may include pole-mounted systems). PV systems consisting of a combination of rooftop and ground-mounted arrays are coded as ground-mounted.	LNBL
Module efficiency	moduleefficiency	Identifies the energy conversion efficiency of the modules.	LNBL
Premium module	premiummodule	Indicates if the system has a premium module (efficiency of at least 20%).	LNBL
Module technology	moduletechnology	Identifies the module technology type.	LNBL
MLPE	mlpe	Identifies the MLPE type.	LNBL
Electricity price	electricityprice	The average electricity price for end customers by state and year.	EIA
Polysilicon price	polyprice	The weekly polysilicon spot prices.	Bloomberg
Programs	programs	The number of incentive programs by state and year.	DSIRE
Hourly mean wage	hourlymeanwage	The hourly mean wage for solar PV installers in the U.S. by year.	BLS
Annual mean wage	annualmeanwage	The annual mean wage for solar PV installers in the U.S. by year.	BLS
Households	households	The number of households, approximated by the number of returns.	IRS
Population	population	The population, approximated by the number of personal exemptions.	IRS
AGI per household	agihh	The adjusted gross income per household by zip code and year.	IRS
Wages per household	wageshh	The wages and salaries per household by zip code and year.	IRS
Taxable income per household	txihh	The taxable income per household by zip code and year.	IRS
Income tax per household	incometaxhh	The income tax paid per household by zip code and year.	IRS

Table 18. Overview of variables used in or considered for the estimation

Note: This is not an exhaustive overview of the variables contained in the data set and investigated in the study, but a selection of the most relevant ones. Some variables were calculated or derived from the original data sets.

B1. Discussion of IV model assumptions

(1) In every linear regression, the number one assumption is that the conditional distribution of the error term given the exogenous regressors has a mean of zero ($E\left(u_i|W_i\right) = 0$), i.e., error terms and the exogenous regressors are uncorrelated (Stock & Watson, 2020). This assumption is key as it makes the estimators unbiased. Omitted variables can cause the assumption to be violated. If there are further unobserved factors that influence the PV demand and that are simultaneously correlated with the regressors included in the model, then estimation results will be wrong in case these missing variables cannot be included in the model. As some data are not available or unobservable, this will be a relevant issue to be discussed in more detail.

(2) The second assumption for a valid IV estimation is that all variables are independent and identically distributed ($i.i.d.$), i.e., all observation samples (Y, X, W, Z) must be $i.i.d.$ draws from their joint distribution (Stock & Watson, 2020). This assumption holds if the data are collected as random samples from a representative pool of observations. In many real-life applications this assumption is violated due to limited representativeness or time-series relationships. It shows that the way of collecting data is crucial. In this study, data used for estimation were reported mainly by incentive administrators in 30 states (Barbose & Darghouth, 2019). The representativeness is questionable as there might be a selection bias towards systems eligible for incentive payments, although this holds true for most. Furthermore, many observations stem from rather high-cost locations. That implies that the estimated effects might be valid for a reasonable subset of the installations rather than all installations in the U.S. This needs to be kept in mind for the interpretation and generalisability of the results.

(3) The third assumption is that there are no extreme outliers in the data, implying that the fourth moments measuring the kurtosis, i.e., the tail of the distribution, are finite. Luckily, the data at hand has been cleaned and pre-processed and obvious outliers have been removed[27], wherefore one can assume this assumption to hold.

(4) The last and probably most important requirement is that all instruments are valid. This means that all instruments must be both *relevant* (not weak) and *exogenous* (not correlated to the error term), as already outlined in 3.2.2 above

[27] See section 3.4.1.2 for details on data selection.

(Stock & Watson, 2020). If this is not the case, the resulting estimates might be biased much more than with an OLS estimation (Angrist & Krueger, 2001). Fortunately, there are ways to test relevance and exogeneity, at least for overidentified IV regressions. One solution to the problem of weak instruments is to estimate the reduced form equation, i.e., conducting an OLS regression of the dependent variable Y directly on the instruments Z and the exogenous variables W (Angrist & Pischke, 2008). The resulting estimates are unbiased and proportional to the coefficients of interest which can be inferred by rescaling (Angrist & Krueger, 2001). Further, the relevance condition includes that there is no perfect multicollinearity between regressors. In practice, regressors are often partly correlated, e.g., module type and module efficiency might be related in this data set. However, the estimations are still valid as long as the collinearity is not perfect. Multicollinearity will be investigated by computing pairwise correlations as well as the variance inflation factor[28] (VIF) (James et al., 2013).

[28] "The VIF is the ratio of the variance of j when fitting the full model divided by the variance of j if fit on its own." The smallest possible value indicating absence of multicollinearity is 1, a VIF value exceeding 5 or 10 shows a collinearity problem (p. 101, James et al. (2013)).

B2. Data Processing

Firstly, I convert all data types into the correct format for estimation and change system size from kilowatts to watts. To improve the estimation quality, I also generate more predictors from the given data. Most importantly, I create per-watt-values for installed price, rebate or grant and sale tax. Furthermore, I calculate the tax rate applied by dividing the total sales tax paid by 55% of the installed price as it is assumed that sales taxes are levied only on the hardware costs of the installation. I also remove the four-digit zip code extension given for some observations to obtain five-digit zip codes for every installation. Additionally, I add the installation year, month, and week as variables to the data set, while making sure that one year only has 52 distinct weeks.

Additionally, I group systems by the state they are installed in, selecting all states with more than 20,000 installations and grouping the rest into 'other' states (7.8% of the estimation sample). Likewise, I cluster observations by customer segment into residential, small non-residential, and large non-residential and I group module technology into polysilicon, monocrystalline silicon, and other technologies to obtain a significant number of observations per category. For some installations, up to three different modules are listed in the original data set. However, for the second and third module, less than 4% and 1% of all observations, respectively, show data on module technology. Therefore, the second and third module values on technology, efficiency, etc. are only considered if information for the preceding first or second is not available. Thereby, I merge the information on module technology, efficiency, and use of microinverter for all modules belonging to one installation.

Subsequently, I add dummies to indicate a premium module with efficiency no less than 20% (about 8% of the full sample and in line with Barbose and Darghouth (2019)), the presence of MLPE, and an indicator for whether the observation meets the criteria to be included in the price sample or not. Finally, I factor all non-numeric variables, mapping them to integer values to be able to use them for subsequent estimations.

It needs to be kept in mind that several observations exhibit missing data for one or more of the relevant variables. I do not impute missing data or outliers which is more common in machine learning, as there is substantial uncertainty about most of the missing values. However, simply discarding observations with missing predictor values can likewise lead to biased estimates, especially if installations with missing values differ systematically from the completely observed cases or if there are only very few complete observations in the sample, significantly reducing the data finally used for estimation (Gelman & Hill, 2006). This could be relevant in this estimation as missing data are often related to the phase down of incentive programs which have primarily been used for data collection (Barbose & Darghouth, 2019).

The above represents a brief summary of the most relevant data processing steps. For more details, please refer to the corresponding code provided in the R Markdown file.

B3. Data Transformation

After left joining all additional data sources to the estimation sample constructed from the TTS data set, the whole sample needs to be transformed in order to enable the estimation of elasticities. First, I convert tax rate and module efficiency values to percentages. Subsequently, to specify a log-log model, I log-transform all continuous variables. For those which exhibit zero values, I add one to the data in order to still map zero values to zero after taking the log, as log(1) equals 0 (Benoit, 2011).

Standardisation and normalisation of variables is reasonable in case the model requires comparable scales. However, this is not the case for linear algorithms, wherefore I refrain from any such transformations in order to maintain high interpretability of the estimation results. The resulting pooled cross-sectional data set is well-suited for studying dynamics of change and transition behaviour like the adaption of a new technology (Baltagi, 2008; Dielman, 1983). Unfortunately, as for some states and years, there are substantially more data points available, the panel is notably unbalanced.