Correlational analysis of energy burden and eviction rate

Paichen Li
Duke University, Nicholas School of the Environment

Dr. Lori Bennear, Adviser

April 22, 2019

Masters project submitted in partial fulfillment of the requirements for the Master of Environmental Management degree in the Nicholas School of the Environment of Duke University

Abstract

Evictions occur when a landlord expels renters from residing in property the landlord owns. Recent data suggest that approximately 40% of residential households in California from years 2012 to 2016 are occupied by renters. The prevalence of renting along with increasing awareness of evictions make studying the causes of eviction a topic of interest for public officials, scholars, housing service providers, and the renter population among others. High cost of living is a direct common cause of evictions across the US. This paper examines and presents a study on the connection between energy burden (how much a single household pays for electricity out of its total household income) and eviction rate. Analysis relies on the application of quantitative research methods using census tract level data from 2012 to 2016 over the service territory of Southern California Edison (SCE). This study uses models that account for both time-variant and time-invariant effects of other key cost and household demographic variables on eviction rate. By taking this approach, the author attempts to separate an unbiased effect of energy burden, which could inform predictions about whether high energy burden is generally accompanied by high eviction rates. Preliminary results suggest that there is a borderline significant positive correlation between energy burden and the unobserved time-invariant census tract level heterogeneity that contributes to higher eviction rates.

Introduction

Evictions occur when a landlord expels renters from residing in property the landlord owns, or involuntary moves initiated by the landlord that drive renters from their residence. Evictions may affect social, economic, and mental well-being of their victims. While the most direct consequence for a family would be losing their home, evictions may also cause children to switch schools as the family moves, loss of possessions, and a negative court record that may prevent them from finding residence in a safe neighborhood. Other effects include but are not limited to job loss, depression, and exacerbated poverty conditions for low-income families.

There are many possible causes of eviction, including but not limited to taking on boarders, damaging property, engaging in illegal activities, causing disturbances that violate leasing agreements, etc. According to the Eviction Lab at Princeton University, by far the most evictions are results of renters' failure to pay their rent and utilizes. While failure to pay rent directly signals to the landlord limited financial capacity of the renters, not paying utility bills on time accrues arrearages to utility service providers which could often lead to electricity or gas disconnections. In a residential compound, the latter could spell disconnection consequences not only for the household responsible but also for other households that share the same electricity/gas meter.

A recent survey by the Eviction Lab suggests that most poor renting families spend over half of their income on housing-related costs, with one out of four spending more than 70% of their income on rent and utilities alone. The survey attributes this to surging housing and utility costs in the US while incomes for those previously below or around the poverty line did not improve, making it difficult for low-income renter households to keep up. Lack of affordable housing is critical to a spectrum of social problems, from poverty and homelessness to health care and educational disparities. While the monthly rent is communicated upfront, utility spending is usually not, and may become a major determining factor on whether a family can afford housing-related costs.

This study approaches the eviction crisis through identifying a connection between utility spending and eviction. Utility spending can be significant for low-income families, especially in summer and winter months when energy use rises as a result of higher demands for cooling or heating. Energy burden, defined as the percentage of utility spending out of total household income, is a bigger threat to low-income households as they are families that tend to live in housing with less energy efficient designs. As a result, these families consume more energy and struggle to pay utilities to avoid utility disconnections and evictions. A 2016 study of America's biggest cities by Energy Efficiency for All (EEFA) at the Natural Resources Defense Council (NRDC) and the American Council for an Energy-Efficient Economy (ACEEE) suggests that the median low-income household spends 7.2 percent of its total income on energy, twice as much as the median for all households which is 3.5 percent.

This study identifies the connection between utility spending and eviction rate in the service territory of Southern California Edison (SCE). In doing so, the author establishes and presents an analytical framework for quantitatively estimating the correlations between energy burden and chances of eviction. This study also addresses the limitations in data availability and inconsistencies in data collection and maintenance that can be addressed through developing more systematic and comprehensive data programs and partnerships with research institutions, followed by a discussion of the potential role of energy efficiency programs in reducing evictions and improving welfare for low-income families and communities.

Data

Quantitative analysis requires data on eviction rates, household demographics, utility rates and billing plans, and energy consumption. This study analyzes census tract level data on a yearly time resolution from 2012 to 2016, covering 2,634 census tracts within the service territory of SCE. Graphs delineating time trends of some key variables and their distribution across energy burden deciles can be found in Appendix i. Some summary statistics of the studied variables can be found in Appendix ii.

Eviction records data are collected, cleaned, and maintained by Eviction Lab at Princeton University. Eviction rates have a range from 0 up to 20%. Due to data collection inconsistencies, eviction rate information is only available for approximately 2,000 census tracts within in SCE serviced territory, which makes the eviction rate the variable with the most missing values. Because we might be concerned that census tracts with eviction data are systematically different from those without such data, a t-test (assuming unequal variances) was performed. Table 1 presents the mean of each key variable for observations separated by the available of eviction records, as well as significance level of whether the means across groups are different.

Results suggest that the pattern of missing data is not random. There is statistically significant difference in all variables if at the 90% confidence level. We see that tracts with missing eviction data, that are not a part of the regression analysis in the study, are characterized by significantly higher energy burden, poverty rate, rent burden, and percentage of Hispanic/Latinos, and significantly lower percentage of renter-occupied housing, multi-family housing, as well as the percentages of Asian, African American, and other minorities' populations. The differences in socio-economic variables imply that including only tracts with complete information in the model omits those tracts with more low-income residents, and likely those whose chances of being evicted are most affected by energy burden. This implies that the relationship between energy burden and eviction rates observed in our sample may understate the true relationship between these two variables.

Census tract level aggregate demographics data are available from the United States Census Bureau's American Community Survey (ACS). ACS has been collecting and presenting census data of American households on a yearly basis from 2005 to 2016 and reports key statistics separately for renter-occupied and owner-occupied households on a census tract level. As such, the author uses reported statistics from ACS on median annual household income when calculating rent burden and energy burden. The author also selects other key variables from the ACS dataset as predictor variables, including median yearly rent (in 2012 USD), poverty rate, percentage of renter occupied housing, percentage of multi-family housing, as well as percentage of Hispanic/Latinos, African American, Asian, and other minorities' populations. Median annual rent burden is calculated as the quotient of median yearly rent and median annual household income. All ACS data are publicly available via the American Factfinder data portal.

Table 1. Tests of Differences in Means among Census Tracts with and without Eviction Data

Variables (Unit: %)	Tracts without eviction records (n=2,731)	Tracts with eviction records (n=10,040)	Difference
Energy burden	1.328	1.267	0.061**
Poverty rate	13.167	12.494	0.673***
Percentage of renter-occupied housing	38.966	42.300	-3.334***
Rent burden	35.269	34.765	0.504***
Percentage of multi-family housing	0.136	0.196	-0.06***
Percentage Hispanic/Latinos	44.620	43.847	0.773*
Percentage African Americans	5.130	5.771	-0.641***
Percentage Asians	6.150	14.291	-8.141***
Percentage other minorities	3.019	3.093	-0.074***

*Significance levels: * if p-value < 0.1; ** if p-value < 0.05; *** if p-value < 0.01.*

Utility rates, billing plans, and energy consumption data on the census tract level are not available in public datasets. The author requested data from three major utility companies in California, Pacific Gas & Electric (PG&E), Southern California Edison (SCE), and San Diego Gas & Electric (SDG&E) through Energy Data Request Programs (EDRP). Constrained by the timeline of this project, the author only had sufficient time to analyze SCE data that came the earliest. The SCE dataset samples observations of roughly 40 households from each census tract that have over 100 SCE-serviced households. The sample size is proportionately smaller from census tracts that contain fewer than 100 households under SCE service. With data on utility rates, billing plans, and monthly electricity consumption, the author was able to calculate monthly electricity spending for each household, which can then be used to generate annual electricity spending. The median of annual household electricity spending is divided by the median annual household income to produce an estimate of the median energy burden for each census tract in a particular year.

A number of assumptions are made to allow construction of models/data from existing statistics. First, energy use in households is not restricted to electricity use and often includes the consumption of gas, biofuels, and other energy sources. In this study, electricity usage and spending statistics are used to approximate total energy use and expenditure for residential households. Second, electricity usage and spending statistics are not reported separately by tenure (renter-occupied or owner-occupied). Energy use and spending across the two groups are assumed to be identical in this study, thus allowing the use of aggregate data to represent the renter population.

Third, energy burden is calculated as the quotient of median energy spending per household divided by median household income on a census level. A more reliable identification strategy should be based on a dataset that directly collects energy burden information on the household level and takes the median of household energy burden statistics. In the absence of such a dataset, the calculation used in this study provides the most reasonable estimate.

Also, observations of census tracts in cities with historically high median household income and low percentage of renter population (e.g. Beverly Hills), uncommon household demographics (e.g. Bradbury, where many families operate horse ranches), and within unincorporated communities (e.g. Hinkley) are removed to produce a dataset that better describes average renter households. Sampled households that are listed under certain dwelling types (e.g. Domestic, Non-dwelling) are also removed for the same reason. For more information, please refer to Appendix iii.

Methodology

The main hypothesis of this study is energy burden is positively correlated eviction rate, even holding constant other explanations of eviction rates. This hypothesis is tested through three quantitative analysis models, using census tract level data over a 5-year period from 2012 to 2016 (refer to the data section for details). In each model, the same set of variables are included: energy burden, poverty rate, percentage of renter-occupied households, rent burden, percentage of multi-family housing, percentage of Hispanic/Latino population, percentage of African American population, percentage of Asian population, percentage of other minorities' population. The percentage of white population and the year dummy variable for 2012 are left out for collinearity.

The first model is a pooled ordinary least squared (OLS) regression model. In this model, census tract level observations over five years are pooled across the entire geographic territory examined. The model does not account for unobserved individual census tract level heterogeneities, or in other words, assumes that conditional on the observed variables in the regression, the error term across all census tracts is independently and identically distributed (iid). We can relax the iid assumption somewhat by allowing for observations within the same census tract to be correlated but assuming independence among observations in different census tracts. That is the approach taken here. This model can be expressed as

$$Y_i = \beta_0 + \beta_1 \cdot X_i + \beta_2 \cdot poverty\ rate_i + \beta_3 \cdot rent\ burden_i + \beta_4 \cdot multifamily_i + \beta_5 \cdot renter\ pct_i + \beta_6 \cdot afam_i + \beta_7 \cdot asian_i + \beta_8 \cdot hispanic_i + \beta_9 \cdot others_i + \beta_{10} \cdot year\ dummies + \varepsilon_i$$

where Y_i represents the response variable (eviction rate) for a census tract i, and X_i represents the regressor of interest (median annual energy burden) for the same census tract. Including year dummies provides information as to how eviction rate changes over time absent variations in other observables.

The second model is a panel regression model that assumes census tract level random effects. In this model, I assume that there exists unobserved time-invariant census tract level heterogeneity that explains variations in eviction rate, but that it is not correlated with the observed census tract variables in the model, such as energy burden and poverty rate.

The third model is a panel regression model that assumes census tract level fixed effects. In this model, I also assume that there exists unobserved time-invariant census tract level heterogeneity that explains variations in eviction rate. In contrast to the previous model, this model assumes that this heterogeneity **is** correlated with the observed census tract variables in the model. Hence, this model can be expressed as

$$Y_{it} = \beta_0 + \beta_1 \cdot X_{it} + \beta_2 \cdot poverty\ rate_{it} + \beta_3 \cdot rent\ burden_{it} + \beta_4 \cdot multifamily_{it} + \beta_5 \cdot renter\ pct_{it} + \beta_6 \cdot afam_{it} + \beta_7 \cdot asian_{it} + \beta_8 \cdot hispanic_{it} + \beta_9 \cdot others_{it} + \beta_{10} \cdot year\ dummies + \boldsymbol{\alpha_i} + v_{it}$$

where Y_{it} represents the response variable (eviction rate) for a census tract i at time t, and X_{it} represents the regressor of interest (median annual energy burden) for the same census tract and time. α_i represents the time-invariant (hence no subtext t) census tract level heterogeneity that is not correlated with census tract level observables.

In implementing the fixed effects model, coefficients of multiple regressors are not estimated due to high collinearity among them within observations of the same census tract over time, and a lack of variation in the regressors over time. As a result, a variation of the fixed-effects model is applied, separating the regression into two stages to identify how energy burden explains eviction rate.

The new third model runs the fixed-effects panel regression with only year dummies in the first stage. This allows us to eliminate time-variant effects that cause eviction rate to vary across years. The first stage can be expressed as

$$Y_{it} = \beta_0 + \beta_1 \cdot year\ dummies + \boldsymbol{\gamma_i} + \boldsymbol{u_{it}}$$

We then estimate the fixed effects term α_i, which captures the time-invariant census tract level heterogeneity that explains variations in eviction rate. γ_i here is different from α_i as it includes effects from relatively time-invariant observables as well. Upon collapsing the observations across five years into an average during the 5-year period, I ran a second stage regression with γ_i as the response variable as the original set of regressors except for the year dummies, which can be expressed as

$$\boldsymbol{\gamma_i} = \beta_0 + \beta_1 \cdot X_i + \beta_2 \cdot poverty\ rate_i + \beta_3 \cdot rent\ burden_i + \beta_4 \cdot multifamily_i + \beta_5 \cdot renter\ pct_i + \beta_6 \cdot afam_i + \beta_7 \cdot asian_i + \beta_8 \cdot hispanic_i + \beta_9 \cdot others_i + u_i$$

The second stage produces results that provide us with information on how significant energy burden predicts time-invariant census-tract level fixed effects that explain variations in eviction rate.

Each model includes a specification that clusters standard errors at the census tract level. In doing so, I assume that there is a level of similarity between/correlation across observations of the same census tract over 5 years, and account for serial correlation of the error term within a census tract. This helps remove bias in standard error estimates.

Results and findings

A pair-wise correlation test with significance testing is done before the implementation of the models. Results shown in Table 2 shows that almost all the socio-economic covariates in the explanatory variables are highly significantly correlated to one another, except for energy burden and percentage of renter-occupied housing. This finding explains the collinearity issue in panel regression when including all these variables and census tract level fixed effect.

Table 2. Pair-wise correlation of coefficients and significance levels

Variables	Eviction rate	Energy burden	Poverty rate	Pct. renter housing	Rent burden	Pct. multi-family
Eviction rate	1					
Energy burden	0.0596***	1				
Poverty rate	0.2260***	0.1337***	1			
Pct. renter housing	0.0715***	0.0104	0.5972***	1		
Rent burden	0.1339***	0.1016***	0.4362***	0.1904***	1	
Pct. multi-family	-0.013	-0.1654***	0.1868***	0.4949***	0.0440***	1

*Significance levels: * if p-value < 0.1; ** if p-value < 0.05; *** if p-value < 0.01.*

Results from running a pooled OLS regression on the entire dataset are shown in Table 3. When ignoring unobserved heterogeneity at the census tract level, we find that the coefficient estimate on energy burden is positive and statistically significant at 0.05. Holding all else constant, a 1% increase in energy burden is associated with a 0.03% increase in eviction rate. There appears to be a steady decline

Table 3. Pooled OLS regression estimates on eviction rates

Variables	Coefficient
Energy burden	**0.0304****
Poverty rate	0.1509***
Percentage of renter-occupied housing	-0.0059***
Rent burden	0.0012
Percentage of multi-family housing	-0.0542
Percentage of Hispanic/Latinos	0.0096***
Percentage of African Americans	0.0252***
Percentage of Asians	-0.0028***
Percentage of other minority races	0.0345***
Year 2013	-0.0637***
Year 2014	-0.2624***
Year 2015	-0.2505***
Year 2016	-0.3266***
Constant	0.7264***

*Significance levels: * if p-value < 0.1; ** if p-value < 0.05; *** if p-value < 0.01.*

In eviction rate across the years when controlling for socio-economic indicators. The coefficient estimates of poverty rate and rent burden are also positive, which is consistent with the understanding that eviction rates are higher where there is more poverty and where rent takes up a larger proportion of the renter's income. Estimates on the effects of racial compositions are also realistic.

Results from a random-effects panel regression model are shown in Table 4. This model accounts for unobserved heterogeneity on the census tract level and assumes that they are not correlated with the observables included. Since it is reasonable to believe that there are things we do not observe that affect how likely households are evicted in a census tract, this model is a step further than the previous one at disentangling the true impact of energy burden on eviction rate.

Table 4. Panel regression (census tract level random effects) estimates on eviction rates

Variables	Coefficient
Energy burden	0.0012
Poverty rate	0.0174***
Percentage of renter-occupied housing	-0.0061***
Rent burden	0.0008
Percentage of multi-family housing	-0.0612
Percentage of Hispanic/Latinos	0.0088***
Percentage of African Americans	0.0248***
Percentage of Asians	-0.0036***
Percentage of other minority races	0.0340***
Year 2013	-0.0716***
Year 2014	-0.2688***
Year 2015	-0.2460***
Year 2016	-0.3270***
Constant	0.8186***

*Significance levels: * if p-value < 0.1; ** if p-value < 0.05; *** if p-value < 0.01.*

As can be seen from Table 4, coefficient estimate of energy burden is still positive but no longer significant. This implies that some of its effect is absorbed by the census tract level random effects. Poverty rate still has a significant positive effect on eviction rate, holding all else constant, while rent burden has a positive yet insignificant effect. Year trends and estimates on the effects of racial compositions do not change much, with signs that are consistent with common perceptions.

Findings from the random-effects model leads to the construction of the final model, results of which are shown in table 5a and 5b. In the random-effects model, some effect of energy burden on eviction rate is found to be absorbed by census tract level fixed effect. As such, there is good reason to believe that these unobservables are correlated with the observables in the model and that a fixed-effects panel regression would produce more accurate estimates.

Including all variables along with census tract fixed effects creates issues of collinearity that led to the omission of multiple key explanatory variables from the actual estimation process. Therefore, the new model is implemented in two stages (more details in the methodology section).

Table 5a. Stage 1: panel regression (census tract level fixed-effects) estimates

Variables	Coefficient
Year 2013	-0.0719***
Year 2014	-0.2753***
Year 2015	-0.2404***
Year 2016	-0.3300***
Constant	1.3713***

*Significance levels: * if p-value < 0.1; ** if p-value < 0.05; *** if p-value < 0.01.*

Table 5b. Stage 2: regressing variables against census tract level fixed-effects estimates

Variables	Coefficient
Energy burden	0.0381*
Poverty rate	0.0174***
Percentage of renter-occupied housing	-0.0058***
Rent burden	-0.0002
Percentage of multi-family housing	-0.0838
Percentage of Hispanic/Latinos	0.0086***
Percentage of African Americans	0.0251***
Percentage of Asians	-0.0036***
Percentage of other minority races	0.0334***
Constant	-0.5557***

*Significance levels: * if p-value < 0.1; ** if p-value < 0.05; *** if p-value < 0.01.*

Results from the first stage suggest a year trend that is similar to what was found in the pooled OLS and random-effects models. In the second stage, the author regresses the five-year average of explanatory variables against the time-invariant census tract heterogeneity that explains different eviction rates across census tracts. The coefficient estimate on energy burden is positive and statistically significant at 0.1 (p-value=0.06), and is greater than that in the previous models. This suggests that energy burden is a good predictor for eviction rate. Poverty rate is still estimated to have a significant positive effect, and coefficients on racial compositions changed very little from the previous models. The reliability of results of this model is contingent on the relative time-invariance of energy burden and the other variables. T-test results of year-to-year comparisons suggest that there is significant variation in energy burden from year 2016 to the other years, which may be the result of policies that increased the provision of assistance programs in California such as California Alternate Rates for Energy (CARE) and Family Electric Rate Assistance Program (FERA).

Limitations and recommendations

There are some limitations in the study, in terms of how the hypothesis is constructed, how the datasets are produced, and how the models are designed and implemented. This section will examine these limitations and recommend ways to overcome them.

The first limitation lies in the construction of the hypothesis. Initially, the author made attempts to identify the causation between utility disconnections and evictions/eviction rate. Initially, the author intended to collect data on changes in the provision of state-wide assistance programs in California (e.g. CARE and FERA). Were these data available, they would be used to construct an instrumental variable for the endogenous explanatory variable of energy burden. Change in assistance program provision would be exogenous to individual households, in theory explain variations in household energy burden, and only affects eviction rate through its influence on energy burden. Applying the IV method would allow us to identify the causal effect of energy burden on eviction rate for renter households whose energy burden is reduced via enrollment in assistance programs. Constrained by the timeline of this project, a data request to researchers at Boston University did not return significant data for the IV method to be helpful. Future research aimed at inferring causality could benefit from taking this approach upon acquiring the data.

The second limitation is in data access. Statistics of a number of key variables in the models are not collected directly from surveys but constructed using data of other variables and possibly from other sources. Construction of median household energy burden and rent burden using census tract level aggregate data may bias the coefficient estimates. This problem could be resolved through the use of primary data collection methods, such as surveys at the household level to determine household energy burden and rent burden, which may be costly to implement. Data can also be inconsistent across different data sources. For example, household income data from the Eviction Lab are generally 40% above those from the ACS. Census data is recommended for analysis as the Census Bureau collects and reports key household demographics statistics separately by tenure, allowing researchers to model the renter population more accurately. It should be noted that the pattern of missing data on eviction records from the Eviction Lab is not random, which reduces external validity and narrows the possible scope of inference. Increasing data collection coverage can enable a more comprehensive and equitable study, since data tend to be missing from places occupied by households with lower income.

The third limitation is the constraint of the panel regression model. OLS can produce unbiased estimates of linear correlation when a number of assumptions are met. In this study, the problematic multi-collinearity from high correlation of predictors in a number of models challenges the assumption that individual predictors should be independent of one another. This may bias individual coefficient estimates without undermining the overall strength of the model at explaining variations in eviction rate. An alternative to be considered in future research is contingent on the availability of a large household-level dataset that includes information on a set of key variables that influence a family's chances of being evicted, such as monthly household income, monthly energy spending, and monthly gross rent. Should there also be a variable that indicates whether each household has been evicted, future research may apply a logistic regression model that fits the probability of being evicted to the other indicators in the dataset and identify how each factor may affect the chances of eviction.

Conclusion

This study provides some evidence for the argument that higher energy burden is associated with higher eviction rate, holding all else constant. While there are many assumptions on which the analytical model is based, and a number of limitations to the study that preclude more accurate estimates of the examined correlations, there are ways to prepare data that satisfy key assumptions, and a series of techniques and research methods that can be applied in response to the limitations.

Based on results of this study, future regulatory efforts can target increasing the coverage of data collection as well as its transparency. A good initiative is to have utilities consistently collect and maintain census tract/zip code level data on household energy use, energy spending, and utility disconnections from failure to pay bills. Energy efficiency programs has the potential to significantly reduce energy burden of low-income families, drop the number of arrearages, and lower the chances of eviction. Identifying and recognizing positive connections between energy burden and eviction rate, especially for low-income families, will raise attention to and support for the work of policy-makers and NGOs that delivers energy efficiency improvement to low-income multi-family housing.

References

American Council for an Energy-Efficient Economy (2016). "Lifting the High Energy Burden in America's Largest Cities: How Energy Efficiency Can Improve Low Income and Underserved Communities." Retrieved from https://aceee.org/sites/default/files/publications/researchreports/u1602.pdf

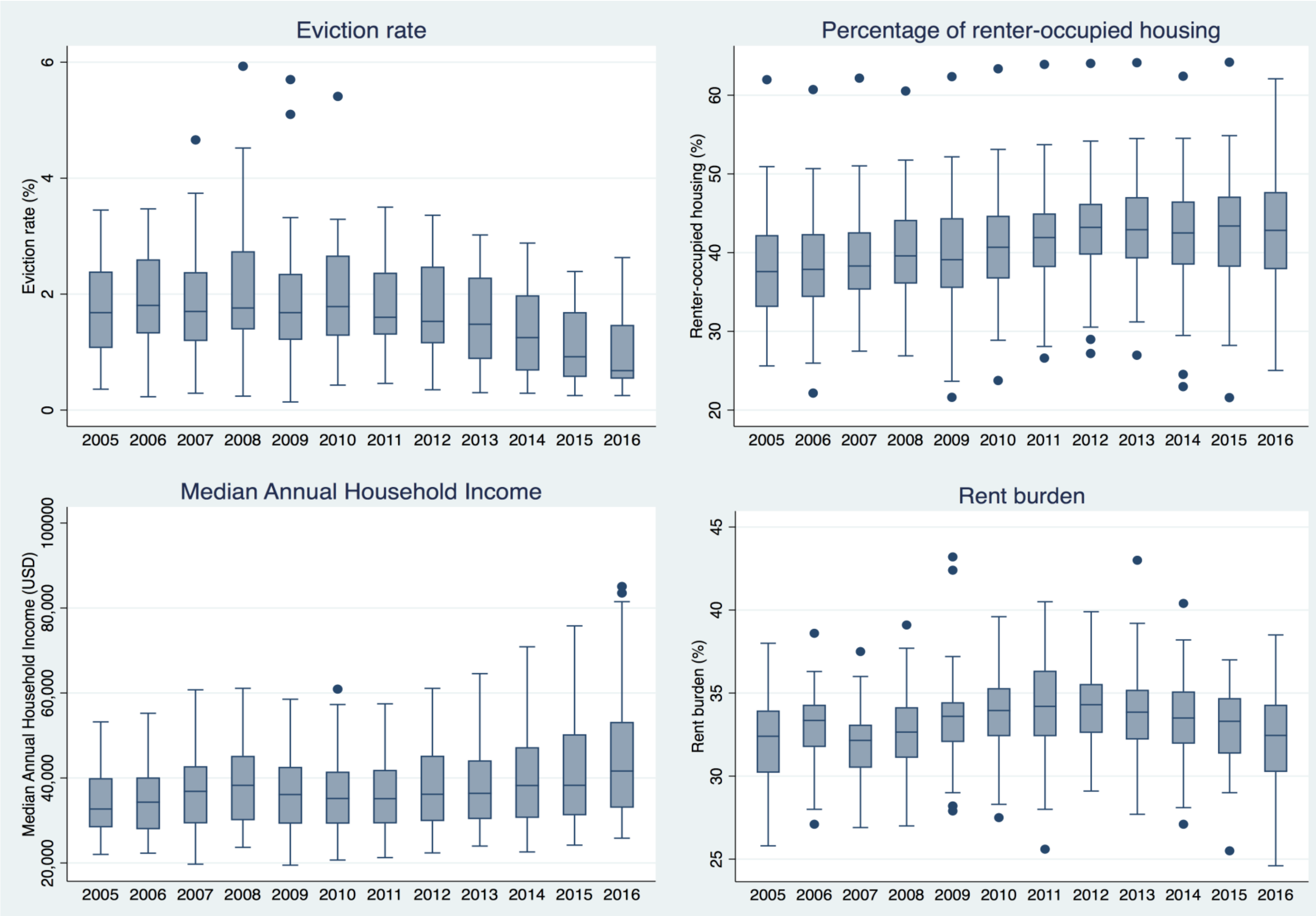
Eviction Lab, Princeton University (2018). California census tract level evictions data. Retrieved from https://data-downloads.evictionlab.org/

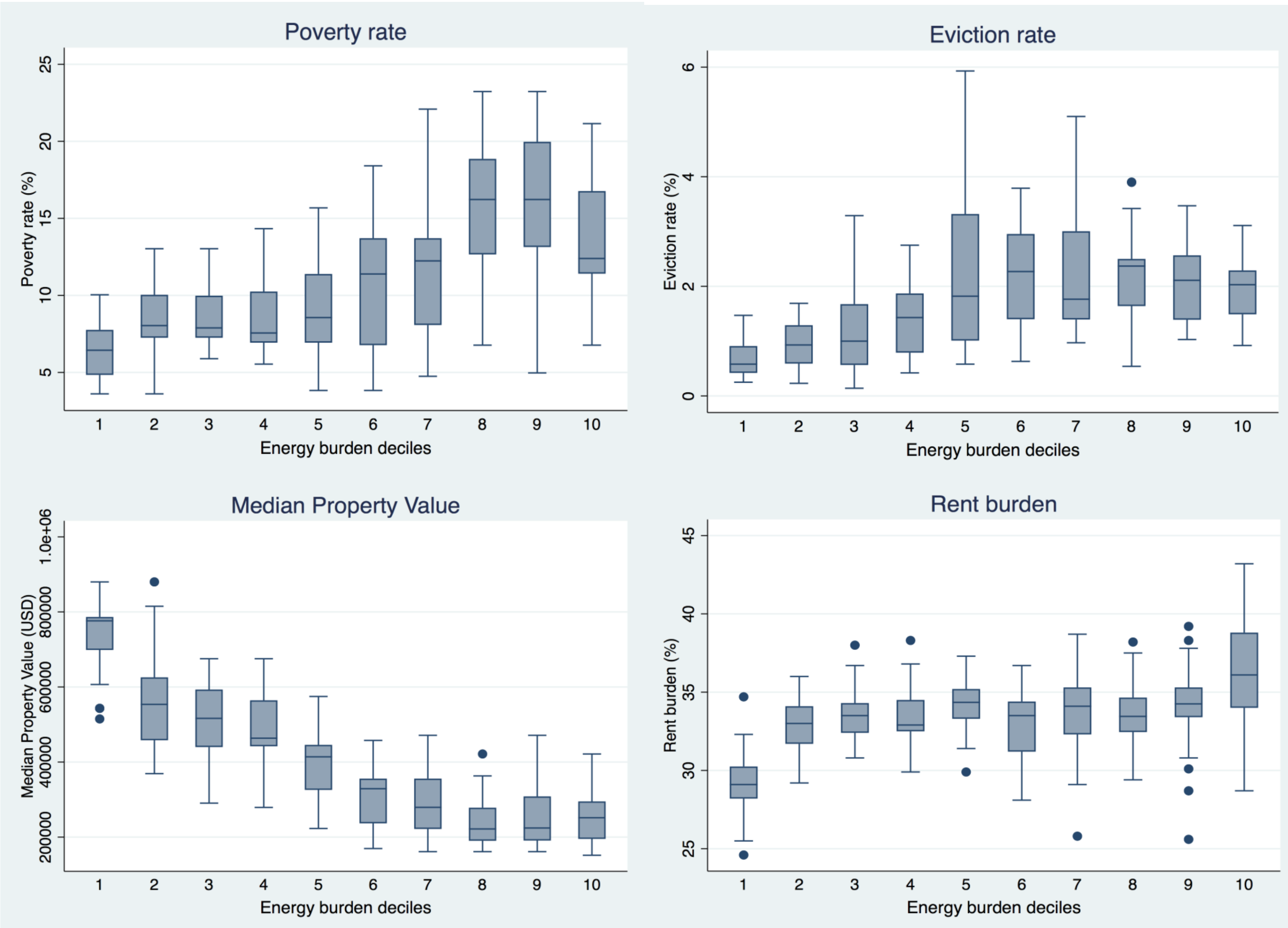
Eviction Lab, Princeton University (n.d.). "What is the relationship between the affordable housing crisis and the eviction epidemic?" Retrieved from https://evictionlab.org/why-eviction-matters/#affordable-housing-crisis

Southern California Edison. Household dwelling type, electricity use and billing data. Requested in January 2019.

United States Census Bureau. Datasets on household income, gross rent, and race composition. Retrieved at https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml

Appendix i. Graphs of distribution of key variables over time and energy burden deciles





Poverty rate
Poverty rate (%)
25
20
15
10
5
1 2 3 4 5 6 7 8 9 10
Energy burden deciles
Eviction rate
Eviction rate (%)
6
4
2
0
1 2 3 4 5 6 7 8 9 10
Energy burden deciles
Median Property Value
Median Property Value (USD)
1.0e+06
800000
600000
400000
200000
1 2 3 4 5 6 7 8 9 10
Energy burden deciles
Rent burden
Rent burden (%)
45
40
35
30
25
1 2 3 4 5 6 7 8 9 10
Energy burden deciles

Appendix ii. Summary statistics of explanatory variables (unit: %)

-> YEAR = 2012

stats	energy~n	povert~e	pctren~d	rentbu~n	multi_~t	pcthis~c	pctafam	pctasian	pct_ot~s
N	2477	2479	2479	2471	2478	2479	2479	2479	2479
mean	1.404037	12.51025	41.29787	34.81781	.1631669	43.82617	5.597568	12.6059	3.071343
sd	1.818625	10.31454	22.01258	7.073709	.2923665	26.78554	9.732869	15.33497	2.60283

-> YEAR = 2013

stats	energy~n	povert~e	pctren~d	rentbu~n	multi_~t	pcthis~c	pctafam	pctasian	pct_ot~s
N	2514	2517	2517	2507	2516	2517	2517	2517	2517
mean	1.294781	12.60289	41.52839	34.84791	.1722021	43.93942	5.631708	12.57154	3.078264
sd	1.283645	10.40765	22.12348	7.067868	.2942177	26.77823	9.734414	15.29454	2.664788

-> YEAR = 2014

stats	energy~n	povert~e	pctren~d	rentbu~n	multi_~t	pcthis~c	pctafam	pctasian	pct_ot~s
N	2559	2563	2563	2553	2562	2563	2563	2563	2563
mean	1.275686	12.65313	41.62535	34.89091	.1829281	44.04975	5.650433	12.54688	3.072844
sd	1.225717	10.41677	22.14777	7.078076	.2965542	26.81218	9.75005	15.29579	2.657549

-> YEAR = 2015

stats	energy~n	povert~e	pctren~d	rentbu~n	multi_~t	pcthis~c	pctafam	pctasian	pct_ot~s
N	2591	2595	2595	2585	2594	2595	2595	2595	2595
mean	1.258941	12.67242	41.67079	34.89366	.1916566	44.10808	5.642012	12.54392	3.081156
sd	1.065437	10.40855	22.16806	7.06751	.2958428	26.78434	9.703217	15.26011	2.667601

-> YEAR = 2016

stats	energy~n	povert~e	pctren~d	rentbu~n	multi_~t	pcthis~c	pctafam	pctasian	pct_ot~s
N	2630	2634	2634	2624	2634	2634	2634	2634	2634
mean	1.17288	12.74275	41.79973	34.91098	.2026594	44.12596	5.646591	12.49854	3.083516
sd	.9730617	10.47095	22.26232	7.089191	.2965803	26.78426	9.68544	15.23477	2.66886

Variable order from left to right: energy burden, poverty rate, percentage renter-occupied households, rent burden, percentage of multi-family housing, percentage Hispanic/Latino, percentage African American, percentage Asian, percentage of other minorities' population.

Appendix iii. Data preparation—removing observations that are not representative of the average renter household

By city

High household income and low percentage of renter households

- Beverly Hills
- Newport Coast
- Indian Wells
- Hidden Hills
- Rolling Hills
- Coto de Caza

Uncommon residential household demographics

- Bradbury

Unincorporated community with uncommon residential demographics

- Hinkley
- Tipton
- Nipton

(Observations from these cities amount to 1.36% of total observations)

By dwelling type

- Dwelling, Unknown
- Mobile home park master meter
- Residential/Commercial combination
- Residential hotel
- Residential/DM/Multiple
- Domestic, Non-dwelling
- Domestic, Electric vehicle charging
- Unknown

(Observations of these dwelling types amount to 2.34% of total observations)

Appendix iv. Graphs and maps of descriptive statistics—evictions and eviction rate

Figure A1. US map of number of eviction cases 2016

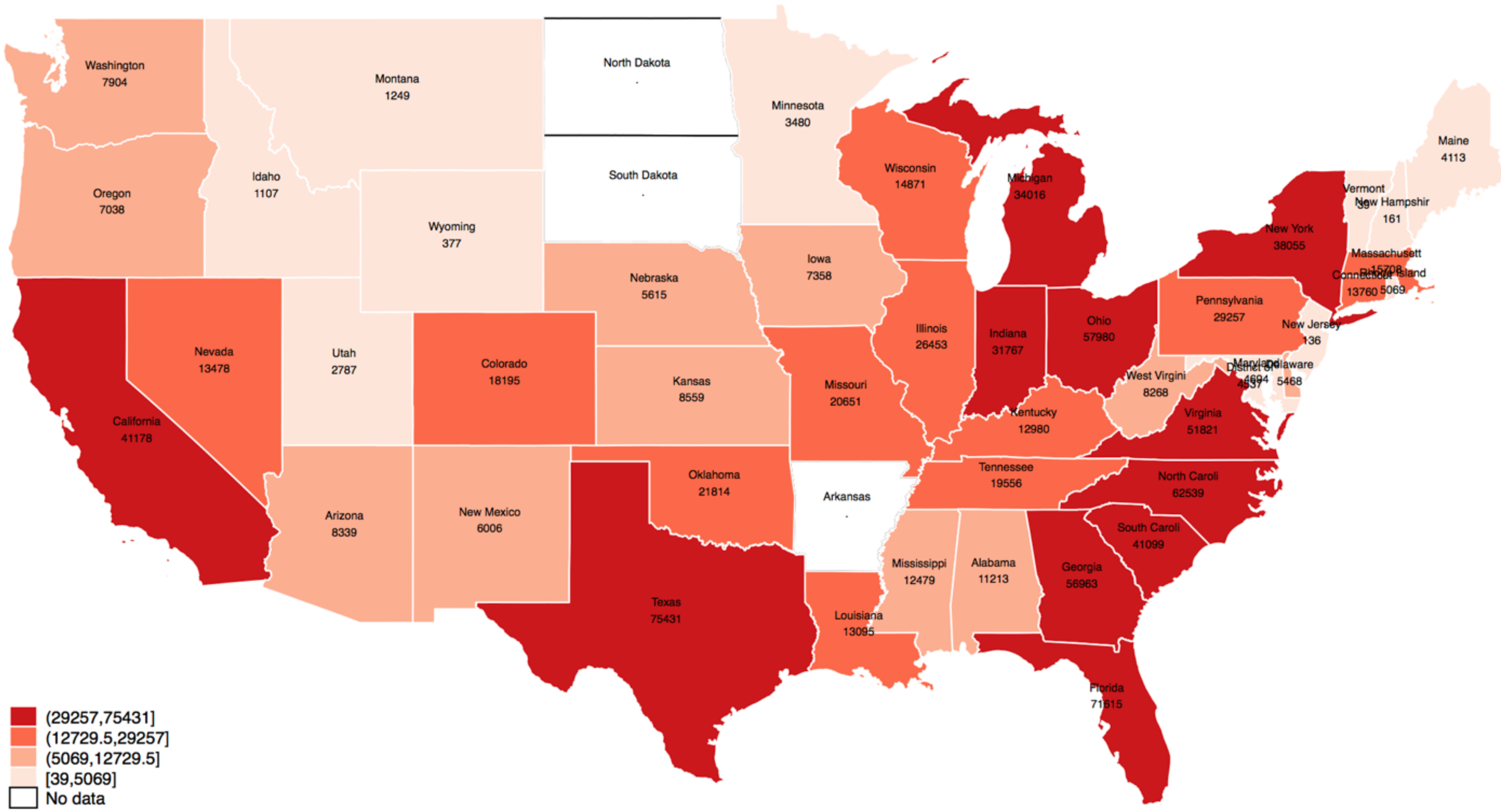


Figure A2. US map of number of eviction rate 2016