Received: 12 May 2022 | Accepted: 9 March 2023

DOI: 10.1111/coep.12610

ORIGINAL ARTICLE

CONTEMPORARY ECONOMIC POLICY

Methods in open policy analysis: An application to California's building energy codes

Matthew J. Holian

San Jose State University, San Jose, California, USA

Correspondence
Matthew J. Holian, San Jose State University, One Washington Square, San Jose, CA 95192, USA.
Email: matthew.holian@sjsu.edu

Abstract

Have building energy codes lowered energy consumption, and have their benefits outweighed costs? Using 2000 Census data, I estimate household energy expenditures by decade of home construction, controlling for household and home characteristics. I find homes built in the 1980s used $35 less in electricity and $46 less in natural gas, per year, compared to 1970s era homes. For Sacramento, energy codes pass a cost-benefit test when low-end policy costs are used, but fail with base-case costs. This study also clarifies how a cost-benefit analysis (CBA) for a representative household fits into a comprehensive CBA.

KEYWORDS
energy, environment, housing, regulation, urban

JEL CLASSIFICATION
R1, Q4, D61

1 | INTRODUCTION

California was the first state to adopt energy efficiency standards for new homes in 1978. These regulations, now known as the California Energy Code (Title 24, Parts 6 and 11), set forth building energy efficiency standards that are updated every few years. Over time, other states adopted state energy codes. Have building energy codes succeeded in lowering energy consumption, and have their benefits outweighed their costs? While recent scholarship has tended to cast doubt on their cost-effectiveness, insights from earlier research designs has not been fully extracted. Also missing from our current state of knowledge is a clear understanding of how "representative household" effects—the focus of much of the literature—fit into a comprehensive social cost-benefit analysis (CBA).

In this paper, I revisit Costa and Kahn's (2011) model of electricity expenditures, and reanalyze it by including natural gas expenditures as a dependent variable. I then use the estimates in a social CBA for a representative household. The use of replication methods in this study lowers the cost of doing policy analysis, minimizes researcher bias (by limiting both functional form and policy cost search) and is an example of open policy analysis (Hoces de la Guardia et al., 2021).

Using 2000 Census microdata for California, I find homes built in the 1980s spent $35 less in electricity and $46 less in natural gas, per year, compared to observationally identical homes built in the 1970s. In an extension of these main results to Sacramento, I find natural gas reductions are larger than electricity reductions, but, not quite as large as a

Abbreviations: CBA, cost-benefit analysis; CEC, California Energy Commission; IECC, International Energy Conservation Code; NPV, net present value; PUMA, Public-Use Microdata Area.

 © 2023 Western Economic Association International.

conjecture from Novan et al. (2022). These authors did not have data on natural gas use, but thought it was plausible that natural gas reductions from energy codes would be about five times larger than electricity reductions; I find natural gas reductions to be about four times larger than electricity reductions. A wide range of policy cost estimates exists in the literature. Based on my policy impact estimates for Sacramento, energy codes pass a cost-benefit test (present value of benefits exceeds compliance costs) when the best-case (low-end) policy cost estimates are used, and fail the test with base-case (mid-point) policy costs.

Five recent studies have laid important groundwork for a comprehensive CBA of energy codes. Costa and Kahn (2011) use a large sample of microdata from the 2000 U.S. Census, but they restricted their attention to electricity only and did not carry out a policy analysis. Jacobsen and Kotchen (2013) and follow-on work by Kotchen (2017), use smaller sample sizes but more detailed energy billing data, from Gainesville, Florida, to study the effect of an energy code change. They focus on a representative household, that uses both electricity and natural gas, and present social payback periods for a representative household. Levinson (2016) uses multiple household surveys and empirical approaches, including a selection on observables approach. I adopt his notation in the next section. He also presents evidence on policy (compliance) cost for Sacramento. Finally, and most recently, Novan et al. (2022) use extremely detailed, hourly electricity billing data from Sacramento to estimate the impact of the initial adoption of California's energy codes. They also carry out archival research on policy cost, allowing us to better understand the cost of the initial versus subsequent changes to energy codes, and they carry out a social CBA for a representative household.

This paper advances CBA methods for this policy setting, and also the "replicate and extend" approach laid out in Holian (2021). It employs open research practices including transparent reporting, open data, and clearly documented code, which unifies the econometric and cost-benefit analyses, and is written in R, an open source programming language. The steps behind the policy analysis, as well as differences with previous analyses, are also documented.

The remainder of this paper is organized as follows. The next section presents background on the policy context. Section 3 describes the empirical approach, which relies on replicating, reanalyzing and extending Costa and Kahn (2011). Section 4 presents original empirical results and compares them to earlier studies. Section 5 uses these results, along with plug-in values from the literature, to carry out a CBA of the first decade of California's building energy codes for the Sacramento region. Section 6 presents robustness checks and examines the impact of home type. Section 7 revisits the case study of Gainesville in Jacobsen and Kotchen (2013) and Kotchen (2017), by using the public-use microdata to investigate the number of households with two energy mix types. I find the large majority of households in Florida do not use natural gas at all. Section 8 presents an equation for comprehensive NPV, Section 9 extends the analysis by dividing the estimation subsample by California's climate zones, and finds evidence consistent with natural gas reductions being larger in colder climates. A brief conclusion summaries the contributions of the study, and the Appendix presents additional robustness checks using the American Community Survey.

2 | BACKGROUND

The Warren-Alquist Act established the California Energy Commission (CEC) in 1974. One of the CEC's main requirements was to adopt and maintain energy conservation standards for new residential buildings. The CEC first adopted building energy standards in 1977 (CEC, 1981b). In 1978, state legislation (SB 331, Robbins) mandated that all building standards, governing fire safety, accessibility for disabled, energy codes, etc. be unified in a single code within the California Code of Regulations, and designated as Title 24, the California Building Standards Code.[1] Although energy codes are sometimes referred to as "Title 24" in the literature, energy codes as a specific type of building standard comprise only Article 6, Part 6 of Title 24 of the California Administrative Code.

The initial codes included both prescriptive and performance standards (CEC, 1981b). For example, performance standards set forth energy budgets new homes had to meet, and builders are free to use practices including insulating walls and ductwork, and infiltration control (i.e., caulking, weatherstripping, etc.). It was also possible to comply by following prescriptive standards from approved lists of construction practices and appliances, for example, which specified the quality of window glazing. The initial energy codes developed by the CEC were more stringent in colder climates (Novan et al., 2022) suggesting energy savings from heating was the primary concern. Although the focus within the economics literature has been on single-family homes, the initial codes "...apply to all new hotels, motels, apartment houses, lodging houses, dwellings and other residential buildings which are heated or mechanically cooled." (Horn, 1980, p. 1).

14657287, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/coep.12610 by TPS: Western Economic Association International - WEAI - TPS, Wiley Online Library on [22/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14657287, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/coep.12610 by TPS: Western Economic Association International - WEAI - TPS, Wiley Online Library on [22/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Today nearly all states have adopted energy codes. Currently, nine states do not have a statewide energy code (including North Dakota and Alaska), two (California and Washington) have "custom" state codes that are unique and stricter than model code, and the remainder adopt versions of International Energy Conservation Code (IECC) model code.[2] Meanwhile California and other states have continued to strengthen their standards. For example, the 2019 Building Energy Efficiency Standards required that new homes have solar panels. Some cities impose stricter regulations, such as Berkeley, California, which no longer permits the use of natural gas in new construction.

Many factors can influence how a home is built apart from regulations. Some homes were built to higher standards before the codes were adopted or enforced, although Costa and Kahn (2011) note that real electricity prices were falling in the 1960s and 1970s, which may have led home buyers and builders to pay less attention to energy efficiency in construction. Working in the other direction, oil crises in the 1970s may have focused consumer attention on conservation in home building. What I estimate below are home vintage effects. The regression-controlled difference between an average 1970s home and 1980s home provides a measure of what could be gained in terms of lower energy expenditures by bringing a noncompliant home into compliance, but like all measures, it is not without possible sources of bias. I discuss these in the next section.

At the time energy codes were adopted, some CEC reports argued the regulation would cause household energy expenditures to fall substantially, and be well-worth their cost. For example, CEC (1981c, p. I-1) claimed, "Under the new standards, new homeowners will be paying only half the money for heating, cooling, and water heating that they would have paid for homes built to current standards." Another report (CEC, 1981a, p. 2) claimed, "For the household residing in a \$75,000 single family detached home, additional construction costs attributable to the standards (average cost \$1500) increase home income requirements by less than 2–1/2 percent...[while] increased home energy efficiency results in positive annual cash flow in the fourth year of homeownership." Despite possible sources of bias, the home vintage effects I present in Section 4 should be able to detect energy expenditure reductions if they are anywhere near this level of magnitude. I now turn to the data and model.

3 | DATA AND METHODS

In this section, I describe the electricity expenditure model from Costa and Kahn (2011), who estimated it on a sample of California homes using household-level microdata from the 2000 Decennial U.S. Census, 5% sample. I first replicate and reanalyze this model by substituting natural gas expenditure for the dependent variable, and then extend it by estimating the energy models on a sample of homes in Sacramento county. All data and code from this study is archived to facilitate future research.[3]

The empirical approach estimates decade of construction effects in an energy expenditure model, while controlling for home and household characteristics. The empirical equation is:

$$lnE_i = X_i\beta + \sum_j \theta_j ConstructPeriod_{ji} + \varepsilon_i,$$

where the dependent variable lnE_i is the natural log of annual energy expenditures (electricity or natural gas) reported by household i. The matrix X_i contains control variables such as number of rooms, household income, and household size. The model also includes fixed effects at the Public Use Microdata Area (PUMA) level. PUMAs are small and mostly self-contained within California's five electric utility districts, so these fixed effects will account for differences in contemporaneous electricity prices as well as other time-invariant and geographic-specific factors like climate. The estimation subsample includes homes built between 1960 and 2000.

Variable descriptions and summary statistics are presented in Tables A1 and A2 in the Appendix. We see in Table A2 that average electricity expenditures for the California sample are \$1128, almost double average natural gas expenditures, which were \$607, and the average home in our sample had 6.27 rooms. The θ_j's are the key coefficients in the model.[4] These are the coefficients on the construction period dummies. If California's building energy codes were highly effective, we would expect θ_{1980} to be lower than θ_{1970} (the reference category).

California's first building energy codes went in effect in 1978 and due to the delay in permitting and enforcement, the initial policy was in full effect for homes built in 1980 and 1981. In 1982, the initial codes were strengthened (e.g., in Sacramento, they now required builders following the prescriptive standards to use double paned windows; see Novan et al., 2018, p. 48 and tab. A.11). It is likely that the revised codes were in full effect for essentially all homes built between

1983 and 1989. Thus, the estimate of the θ_{1980} coefficient reflects a mix of the impact of the initial and revised codes, but more heavily weighted toward the impact of the revised 1982 codes, as more homes were built in this longer period.

Homes built in the 1970s and 1980s, and households that live in them, will use differing amounts of energy for reasons apart from the design of the home that is due to the energy codes that existed when the home was built. As two examples, on average, older homes are smaller, and, their residents have lower incomes. Thus homes built in the 1980s may use more energy on average than homes built in the 1970s even if codes were effective.

The use of control variables in the equation above, such as number of rooms and average household income, aims to account for these and other energy-using characteristics of homes and households, bringing the comparison of period of construction effects closer to *ceteris paribus*. Like any control variable strategy, there remains the possibility that some unaccounted for variable might affect both a household's selection into a 1970s or 1980s home and their energy use. One example of this is when a household knows it will be using a lot of energy, and therefore selects a more energy efficient home. In this case, a home may be energy efficient and still use a lot of energy. Not all omitted variables serve to lower the estimate. Innovations in building practices that lead to greater efficiency and would have been adopted even without regulation will make the energy codes look more effective than they otherwise would.

Another possibility is that the energy codes affected some of the control variables used in the empirical model. Bruegge et al. (2019) find evidence that by raising construction costs, California's energy codes led to reductions in home size. Number of rooms thus to some extent may be a "bad control" in my model. In the Census sample I use to estimate the model, average home size increases from 6.09 in 1960 to 6.46 in the 1990–1994 period, and continues to rise throughout the 1990s, but interestingly, the difference in average home size between 1980s and 1970s homes is not statistically significant. This simple difference in means is consistent with regulation causing homes to be smaller than they otherwise would be given the secular rise in home size during this period. However, in a regression of rooms on the period of construction dummies with control variables (reported in the archived code), 1980s homes are slightly (and marginally significantly) larger than 1970s homes. In addition Bruegge et al. (2019) find energy codes reduced home size only by less than 2%. Taken together this suggests keeping rooms as a control for secular trends in home size.

Yet another consideration concerns the interpretation of the results. A large fraction of homes built in the 1970s were at least partially compliant with the 1978 energy codes. Novan et al. (2022) report figures suggesting 31% of homes were fully non-compliant, with 42% fully compliant and the remainder partially compliant. To the extent we interpret the difference between the 1980 and 1970 vintage effects as the effect of regulation in making a noncompliant home complaint, it will underestimate the effect because voluntary compliance lowers energy use in 1970s era homes from what it would be without it.

The empirical strategy here uses less detailed data than Levinson (2016), Jacobsen and Kotchen (2013) and Novan et al. (2022). For example, this model does not contain a control variable for duration of stay, as did Levinson. People who have lived in a home longer will have had more time to retrofit for energy efficiency.[5] The data here nevertheless have some virtues, principally a large sample size, making it representative with regard to geography, energy mix and housing type. Even if some bias remains in the estimates, if the bias affects gas and electricity similarly, the estimates are still able to shed valuable light on, for example, whether the *relative* natural gas savings are greater than electricity savings.

One way to evaluate the magnitude of the bias is to compare the results of estimating the equation above with results presented in the literature, which used different, more sophisticated data sources and modeling approaches. If the results here are similar, they can be held to have higher validity. I provide such comparisons at the end of the next section, and I also present robustness checks in Section 6 and the Appendix.

4 | EMPIRICAL RESULTS

Table 1 presents results of estimating the previous equation, for two dependent variables and two estimation sub-samples. The table shows the main coefficient of interest, which is the coefficient on the 1980 construction period variable, YB1980. To help the reader understand the empirical model, the full regression results are shown in Table A3 in the Appendix. Column (1) of Table 1 contains results for electricity expenditures, estimated on a population of single-family homes in California, and is a replication of Costa and Kahn (2011).[6] The coefficient of −0.031 can be interpreted as homes built in the 1980s use 3.1% less electricity than homes built in the 1970's, holding observable characteristics of the home and household constant. In column (2) I present new results concerning natural gas; homes built in California in the 1980s use 7.6% less natural gas than homes built in the 1970's, ceteris paribus. This is consistent with California's energy codes being responsible for a larger (percentage change) reduction in natural gas use than electricity use.

14657287, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/coep.12610 by TPS: Western Economic Association International - WEAI - TPS, Wiley Online Library on [22/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Regression results, census 2000.

	Dependent variable: Log of the cost of			
	Electricity (1, California)	**Gas (2, California)**	**Electricity (3, Sacramento)**	**Gas (4, Sacramento)**
YB1980	−0.031***	−0.076***	−0.027***	−0.221***
	(0.008)	(0.007)	(0.007)	(0.036)
Observations	144,296	135,473	6582	5801
Controls?	Yes	Yes	Yes	Yes
Adjusted R^2	0.163	0.117	0.160	0.142
Residual Std. Error	3.014	3.108	3.128	3.263

Note: The data used are from the 2000 decennial Census and were obtained from Ruggles et al. (2019). The sample includes only single-family, owned homes constructed in California between 1960 and 2000, where the head of household is aged 30–65, and the household has nonnegative income. All models include PUMA-level fixed effects, and standard errors are clustered at the electricity district level. The omitted construction decade dummy is the 1970s.

Abbreviation: PUMA, Public-Use Microdata Area.

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

In columns (3) and (4) I estimate the same models as in (1) and (2) but for Sacramento county. I will discuss the statewide results first and then the results for Sacramento.

How do my estimates for California compare with the selection on observables strategy in Levinson (2016)? A comparison is facilitated by the fact that both he and I use a log dependent variable, however it is complicated by the fact that we have different construction periods; one of the construction periods in his Table 3 combines homes built between 1978 and 1982 and another between 1983 and 1992. The difference in the point estimates of these coefficients is about 0.025 for electricity and 0.04 for gas, though neither difference is statistically significant. We thus both find larger reductions for natural gas than electricity, and the magnitudes are very similar, but my estimates are statistically significant.[7]

To calculate the dollar amount saved by an average household, I multiply the coefficient estimates by the averages of the dependent variable. Homes built in 1980s spend about \$35 less on electricity than homes built in the 1970s, and this figure is \$46 for gas.[8]

I next describe the results of extensions reported in Table 1 columns (3) and (4), compare my electricity estimates to Novan et al. (2022), and calculate the ratio of electricity and natural gas savings to test a conjecture from Novan et al. (2022). The statewide estimation subsample used in columns (1) and (2) contains about 6000 households in Sacramento county.

In columns (3) and (4) the coefficients on the 1980s dummy are −0.027 and −0.221, for electricity and gas, respectively. The percentage change in natural gas expenditure is over eight times the percentage change in electricity expenditure. Average expenditures on electricity and natural gas in Sacramento were \$1105 and \$558, respectively. Thus, converting the percentage changes to expenditure changes, I find homes built in the 1980s in Sacramento spend \$29.8 less on electricity, and \$123.3 less on gas. The reduction in natural gas expenditure is 4.1 times the reduction in electricity expenditure.

In their study of Sacramento, Novan et al. (2022, p. 497) conclude, "…the cumulative gas savings would need to at least exceed …roughly 5 times more than the electricity savings - for the total benefits of Title 24 in Sacramento to exceed the compliance costs." The authors conjecture that they do, and conclude Title 24 likely passes a cost-benefit test in Sacramento. My results in Table 1 and the calculations I present in the previous paragraph suggest natural gas reduction benefits are greater than electricity reduction benefits by a factor of about four, which is close to but less than five. As argued above, even if my estimates are biased, if the relative bias is the same across gas and electricity, then my findings could be considered to cast some doubt on the conclusion in Novan et al. (2022) that the initial adoption of energy codes passes a cost-benefit test, although it suggests it is close.[9]

How do the magnitude of my estimates from Table 1 compare with Novan et al. (2022)? They find a reduction in total electricity use of 1.6%–2.6% whereas I find a reduction of 3.1%. It is important to note that our estimates are not directly comparable, as they estimate the effect of the initial adoption of the codes, whereas what I estimate is a mix of both the effect of the initial adoption and the 1982 changes. Still, the closeness of these estimates would seem to lend validity to my findings.

14657287, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/coep.12610 by TPS: Western Economic Association International - WEAI - TPS, Wiley Online Library on [22/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

5 | COST-BENEFIT ANALYSIS

This section presents a CBA based on the results from the previous section. I focus on Sacramento, because compliance costs estimates for this area are available in the literature, however one could easily adapt this analysis as a statewide analysis for California as a whole, or for other regions, using relevant compliance costs estimates for these geographies.

The net present value (NPV) of building energy codes for a representative household in Sacramento can be written as follows:

$$NPV_R = -8,176 + \sum_{t}^{T} \frac{30\epsilon + 123\gamma}{(1+r)^t}$$

where the subscript R on NPV reminds that it is for a representative household. On the right-hand side of this equation, \$8176 is an estimate of the upfront cost of the policy. Levinson (2016) cited Horn et al. (1980) who estimated the increase in initial cost of compliance for a home was \$8080 in year 1980 dollars. However, in an earlier version of their published paper, Novan et al. (2018) write, we "...estimate that the upfront cost of implementing Title 24 was substantially lower than the cost highlighted in Levinson's analysis..." (p. 45), because the Horn et al. (1980) study was of a set of policies that were stricter than those in the initial energy codes. Novan et al. (2018, 2022) used \$782 as an estimate of the initial policy cost. Because the effect I estimated above reflects both the initial adoption and subsequent strengthening, the average upfront cost for a home in my sample is likely in the range of \$782 and \$8080. Thus I take the average of these two numbers, which is \$4431, and then convert from year 1980 dollars to year 2000 dollars, to arrive at the \$8176 figure.

The values \$30 and \$123 are the annual electricity and gas expenditure savings, respectively, which are treated here as the annual resource cost savings,[10] expressed in year 2000 dollars. These figures were produced based on the estimates from Table 1 and were reported in the text toward the end of the previous section. Both of these annual benefit estimates are weighted, by ϵ for electricity and γ for natural gas. By specifying different values of these multipliers, we can compactly incorporate various assumptions about the social value of the energy savings.[11] Finally, the time horizon is determined by the values of t and T, and the discount rate by r.

Assuming no external benefits ($\epsilon = \gamma = 1$), a 3% real discount rate ($r = 0.03$) and a 40-year time horizon ($t = 1$ and $T = 40$), I find the sum of discounted present value of benefits is \$3537, and NPV = \$-4639.

Next, I incorporate external benefits of reduced energy use in the calculation. Previous studies have made different assumptions regarding their size. Novan et al. (2022) assume the resource cost of electricity generation is \$0.08/kWh, while external costs are \$0.02/kWh (due to the social cost of carbon, assumed to be \$50/ton).[12] This translates to a value of ϵ equal to 25%. Jacobsen and Kotchen (2013) incorporate both electricity and natural gas use in their analysis. The external benefits of reduced natural gas consumption was between 4.2% and 47% of the energy cost savings in their study, with a midpoint of 25%.[13] I recalculate NPV with $\epsilon = 1.25$ and $\gamma = 1.25$ and under a 3% discount rate find the discounted sum of annual benefits are \$4421, and NPV = −\$3755. California's energy codes do not pass a cost-benefit test under my baseline assumptions. However, the benefits (with or without external benefits) are high enough for it to pass if the low-end policy cost estimates (of \$1443, which is the \$782 from Novan et al., 2022, but in 2000 dollars) are used.

Table 2 shows NPV under low, baseline and high cost assumptions, with and without external benefits.

I have produced a CBA of California's initial energy codes considering impacts and costs in the Sacramento region that relied on several open science principles. The impact estimates came from a reanalyzed model from the literature, which is possible because the model was reproducible, and relied on public-use raw data. Future studies can use this

TABLE 2 Net present value (in \$) under different assumptions.

Compliance cost	Without external benefits	With external benefits
Low	2094	2978
Baseline	−4639	−3755
High	−11,373	−10,489

Note: This table shows net present value under different assumptions: low, baseline and high compliance costs of \$1443, \$8176 and \$14,910, respectively, and with external benefit multiplier of 1.0 or 1.25. Calculations are shown in archived research data.

14657287, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/coep.12610 by TPS: Western Economic Association International - WEAI - TPS, Wiley Online Library on [22/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14657287, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/coep.12610 by TPS: Western Economic Association International - WEAI - TPS, Wiley Online Library on [22/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 Robustness checks.

	Model and subsample			
	2000, log	2000, level	2000, log	2000, level
	Detached and duplexes		Detached only	
	(1)	(2)	(3)	(4)
$\beta_{CA}^{Electric}$	−0.031***	−27.895***	−0.019**	−19.233*
β_{CA}^{NGas}	−0.076***	−30.711***	−0.062***	−26.163***
$\beta_{Sac}^{Electric}$	−0.027***	−27.432**	−0.035***	−30.491***
β_{Sac}^{NGas}	−0.221***	−79.091***	−0.207***	−71.460***

Note: Each number is the estimate of the coefficient on YB1980 in one of 16 models. The data rules are identical to those described under Table 1. The full regression results for all models in Table 3 is available in the log file in the online supplement, https://doi.org/10.3886/E187383V1.

$^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

method to derive impact estimates for other settings; the impact estimates from Table 1 could be plugged in the previous analysis if an estimate of average policy cost for California as a whole can be found. The externality multipliers and average policy costs for Sacramento are from documented sources in the literature, and the archived code for both the empirical and NPV analyses runs from the same file with one click.

The results are intended to provide information for policy makers, who consider the success of previous policies when considering new policies. My conclusions concerning the policy are not definitive, the range of estimates for the upfront cost of the policy in the literature is very wide, thus future research should aim to provide more certainty on the true value of policy costs.

The next two sections use the richness of the decennial Census data, in particular with respect to home-type, geography and energy mix, to carry out robustness checks, and to explore Gainesville, Florida, the setting of a key case study in the literature.

6 | HOME-TYPE EXTENSION AND LOG-ROBUSTNESS

The Census microdata contains 10 categories of home type, including both single- and multi-family buildings, and also mobile homes, trailers, and 50+ family building home types. Some of these other home types were regulated under Title 24 (Horn, 1980) but analyzing most of these is beyond the scope of this paper. Instead this section focuses on two fairly similar home types that may or may not be appropriate to combine. The estimation subsample in Costa and Kahn (2011) included only single-family homes, but it combined two categories of housing types, detached and attached single-family homes. Attached houses include duplexes, and they are regulated somewhat differently under California's building energy codes.

A separate issue is that there is no consensus in the literature regarding whether dependent variables should be in logs or levels. To explore the effect of combining two home-type categories in the estimation subsample, and also the modeling choice of log versus level models, I carried out robustness checks reported in Table 3. Each number is the estimate of the coefficient on YB1980, the main variable of interest, where the dependent variable and estimation subsample are indicated by subscripts on β; for example, $\beta_{CA}^{Electric}$ indicates the estimated coefficient is from a model where electricity costs are the dependent variable, and the estimation subsample is the entire state of California. Column (1) contains the same estimates as in Table 1, reproduced here to facilitate comparison. This is the log-specification and home-type mix that Costa and Kahn used. In Column (2) I keep the home-type mix but change to a level-specification. In Columns (3) and (4) I remove the duplexes and re-estimate the log and level models, respectively.

What we see in Table 3 is that the results are somewhat sensitive to the log and level modeling choice, with the log models showing a larger impact. To facilitate comparison, recall at the end of Section 4 I converted estimates in the log models to dollar-values, for electricity and natural gas, where savings for the California sample were \$34.96 and \$46.13, respectively. In Table 3 the comparable figures are \$27.895 and \$30.711, respectively. We also see an effect of home-type, where duplexes contribute to a larger and more significant effect in the California sample.

7 | ENERGY MIX: GAINESVILLE

In Jacobsen and Kotchen's (2013) study of Gainesville, they find evidence that a change in Florida's energy codes lead to savings in both electricity and natural gas use. In follow-up work, Kotchen (2017) finds that the electricity savings disappears, but the natural gas savings persists. Thus, for their representative household, there is at least some benefit from the energy codes in both of the studies.

The representative household in Jacobsen and Kotchen (2013) and Kotchen (2017) used both electricity and natural gas. Out of the 95,658 owned, single-family Florida homes in the 2000 Census 5% sample, I find that 80.37% have no value for COSTGAS and are coded as "no charge or no gas used." Compare this to California, where only 6.17% of households in the sample did not use gas. In both states, nearly all households use electricity, but clearly there are important differences in energy mix across states.

If the changes to the energy codes that Jacobsen and Kotchen (2013) studied led to savings in natural gas use but not electricity use, then 80% of Florida households will not see any benefit from the code changes, although possibly may still have incurred costs. As we have seen, many studies in this literature focus on representative households. In the next section, I illustrate how a CBA for a representative household fits into a comprehensive CBA.

8 | TOWARDS COMPREHENSIVE NPV

To illustrate how a CBA for a representative household fits into a comprehensive CBA, I start by generalizing the equation from Section 5. The equation below is the NPV for a representative household, which accounts for home-type, geography and energy mix:

$$NPV_{gtm} = -APC_{gtm} + \sum_{t}^{T} \frac{ABE_{gtm}\epsilon_g + ABG_{gtm}\gamma}{(1+r)^t}.$$

where NPV_{gtm} is the net present value for a representative household. The subscripts g, t and m refer to geography, type and mix, respectively. California has many microclimates, but 16 official climate zones.[14] In addition, it considers three housing types: detached and attached single-family homes, and multi-family homes. Finally, mix refers to energy mix; as highlighted in the previous section, some homes are electric-only, while others use electricity and natural gas.[15]

The equation above contains several variables. First, APC_{gtm} is the Average Policy Cost in region g, for housing type t which uses energy mix m. Next, ABE_{gtm} is Annual Benefits from Electricity reduction. The parameter ϵ_g is the externality multiplier which varies by geography.[16] ABG_{gtm} is the Annual Benefits from gas reduction, and γ is the natural gas multiplier, which does not vary by geography (because a pound of gas burned in Florida produces the same emissions as a pound of gas burned in California.)

Next I show what comprehensive NPV of a state energy code policy would look like for California:

$$\begin{aligned} NPV = {} & NPV_{1,1,EG} \times N_{1,1,EG} + NPV_{1,1,EO} \times N_{1,1,EO} + NPV_{1,2,EG} \times N_{1,2,EG} + \\ & + \cdots + NPV_{16,3,EO} \times N_{16,3,EO} \end{aligned}$$

There are 16 climate zones, three home types, and two energy mix types: electric and gas (EG) and electric-only (EO). This gives us 96 weighted representative household NPVs, where the weights are the total number of homes in each category ($N_{g,t,m}$).

This comprehensive NPV equation makes apparent that the CBA presented in, for example, Novan et al. (2022) or in Section 5 of the present study, is just one of the 96 weighted NPVs needed for a more complete policy evaluation. The 2000 US Decennial Census can be used to estimate the policy impacts for many representative household NPVs, as well as to provide estimates of the number of homes in each category; most these tasks are left to future research, but the next section re-estimates the models from Sections 3 and 4 using estimation subsamples of households in California's climate zones.

14657287, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/coep.12610 by TPS: Western Economic Association International - WEAI - TPS, Wiley Online Library on [22/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

9 | CLIMATE ZONES

In this section, I exploit the scale of the Census data to examine geographic heterogeneity and to test the hypothesis that zones with colder climates had larger natural gas savings. The stringency of energy codes varies by climate, with stricter energy codes in more extreme climates. Bruegge et al. (2019) present a list of energy budgets for each climate zone (more extreme climate zones are given higher energy budgets) and they also present evidence that energy code strictness is highly correlated with climate measures.

Figure 1 presents a map of California showing its 16 climate zones with thick lines, that also shows its 233 Public-Use Microdata Areas (PUMAs) with thin lines. Many PUMAs straddle two or more climate zones; of the 233 PUMAs in the 2000 Census, 93 are in one climate zone and 140 are in more than one. I restrict analysis to households in the 93 PUMAs that are fully contained within a climate zone. The 2000 Census data does not identify the PUMA for households in areas with low population. Thus for both of these reasons we have Census data for households in the 11 climate zones listed in Table 4.

After assigning households to climate zones, I re-estimate the models presented in Sections 3 and 4, using climate zones as separate subsamples. Table 4 presents the coefficients on the 1980 decade of construction indicator (YB1980) for each of these 11 climate zone subsamples in both the electricity and natural gas expenditure models. Section 4 presented results for California as a whole, and for Sacramento county; Sacramento county is fully contained within Climate Zone 12, thus Sacramento households are combined with other California households who live in areas with a similar climate in the results presented in Table 4. The full results from these regressions are presented in the archived research data and code. Sample sizes range from 715 households in zone 7 (which was matched to 3 PUMAs) to 10,464 households in zone 12 (which was matched to 15 PUMAs). All zones had sample sizes of at least 1500 except for zone 7.

This table also presents average climate conditions within the climate zones. To calculate these climate measures, I use data from the 2007 City and County Data Book, which provides climate data for cities with population of 25,000 or more. There are 242 California cities in this file and I matched 233 of them to a climate zone using the information presented in CEC (1995).[17] I was able to match between 6 and 35 cities to each climate zone. The archived research data presents summary statistics for each climate zone based on these data; in Table 4 I present averages only for six climate measures. The notes under Table 4 describe the measures.

Next I present evidence consistent with the hypothesis that natural gas reductions were larger in colder climate zones because regulations were stricter in these areas. Figure 2 plots the coefficients on YB1980 in natural gas regressions from each climate zone subsample on the y-axis, and the Heating Degree Days (HDD) for the climate zone on the x-axis. An HDD is the annual sum of degrees under 65 degrees Fahrenheit, so colder climates have a higher value of HDD. We see in Figure 2 that in general, colder climates were associated with coefficients on the 1980 decade of construction indicator that are larger in magnitude, consistent with the hypothesis.

It is tempting, because we have the data on Cooling Degree Days (CDD), to also test the hypothesis that electricity expenditure reductions were larger in hotter areas. There is less support for this hypothesis, however. The results in Table 1 show larger natural gas reductions, and only seven of the coefficients in Table 4 are significant from the electricity regressions, whereas all the coefficients from the natural gas regressions by climate zone are statistically significant. In addition, the conjectures and findings of previous research (Kotchen, 2017; Novan et al., 2022) also suggest natural gas reductions are larger. The archived research data does carry out a similar analysis as shown in Figure 2 but for electricity. There, I don't find the expected relationship, and I find some large electricity expenditure reductions in more temperate climates. This could perhaps be explained by the fact that more temperate climate zones are also richer on average, and households in them may have more energy efficient appliances. A more sophisticated analysis is left for future research,[18] but I have shown that the 2000 Census Microdata is large enough to enable us to estimate the Costa and Kahn energy expenditure model by climate zone for most of California's climate zones, and, I do find natural gas reductions are larger in colder climates, which is the hypothesis we would have more confidence in a priori.

10 | CONCLUSION

Drawing on Open Science principles (Christensen & Miguel, 2018), this study has presented original impact estimates of energy codes on energy consumption. My estimates are in-line with previous studies of California's energy codes (Levinson, 2016; Novan et al., 2022). I then used my original estimates to carry out a CBA of California's early energy codes, and found that energy codes pass a cost-benefit test under low compliance cost assumptions. This study used

14657287, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/coep.12610 by TPS: Western Economic Association International - WEAI - TPS, Wiley Online Library on [22/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

FIGURE 1 California climate zones have thick boundaries and are numbered, Public-Use Microdata Areas have thin boundaries and are not numbered.

insights from analysis of public-use microdata to enhance our understanding of two recent case studies from Sacramento and Gainesville. It then provided a formal definition of a representative household NPV and clarified how it fits into a comprehensive CBA. This clarification should help policy makers and others to better interpret findings from the literature. Future research can expand upon this study in a number of ways, including by adapting the methodology to other states and regions, home types, and energy mixes.

14657287, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/coep.12610 by TPS: Western Economic Association International - WEAI - TPS, Wiley Online Library on [22/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 4 Climate zone data and regression coefficient estimates.

Climate zone	AvgDailyJan	AvgDailyJuly	HDD	CDD	BetaE	BetaG
2	49	68	2677	478	−0.085*	−0.147*
3	49	66	2746	383	−0.105*	−0.162*
4	49	69	2633	636	−0.042*	−0.077*
6	56	68	1675	653	−0.032	−0.050*
7	56	71	1534	918	−0.136*	−0.141*
8	57	73	1232	1224	−0.013	−0.069*
9	56	74	1485	1389	−0.052*	−0.136*
10	54	77	1700	1568	−0.056*	−0.089*
12	46	75	2652	1250	−0.015	−0.133*
13	46	81	2455	1916	−0.058*	−0.105*
14	46	82	2774	2000	0.0002	−0.153*

Note: AvgDailyJan (AvgDailyJuly) is the average of average daily January (July) temperatures for cities matched to the climate zone. HDD (CDD) is the average heating (cooling) degree days for cities matched to the climate zone. BetaE (BetaG) is the estimated coefficient on YB1980 in a regression where electricity (natural gas) is the dependent variable. **p* < 0.01. The full regression results for these regressions are available in the archived research data.

Abbreviations: CDD, Cooling Degree Days; HDD, Heating Degree Days.

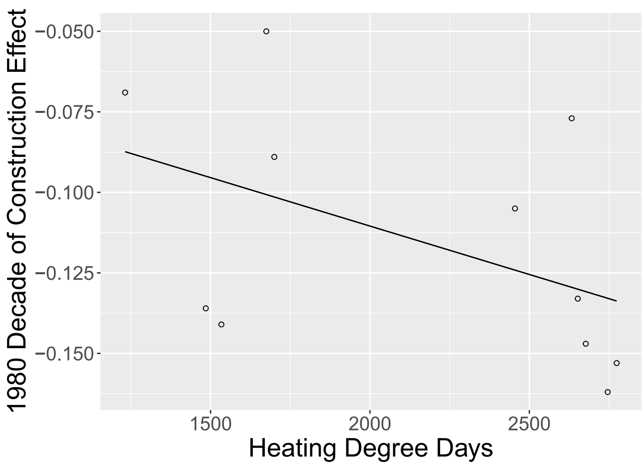

FIGURE 2 Natural gas reductions are larger in colder climates; figure shows HDD and the coefficient on 1980 decade of construction dummy in natural gas regressions. HDD, Heating Degree Days.

ACKNOWLEDGMENTS

I thank two anonymous reviewers for very helpful comments, and Grant Jacobsen, Matthew Kotchen, Kevin Novan and Aaron Smith for helpful discussion. All errors are my own.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are openly available in: Holian, Matthew. COEP Replication Package for "Methods in Open Policy Analysis: An Application to California's Building Energy Codes". Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2023-03-25. https://doi.org/10.3886/E187383V1.

ORCID

Matthew J. Holian https://orcid.org/0000-0002-7319-4294

14657287, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/coep.12610 by TPS: Western Economic Association International - WEAI - TPS, Wiley Online Library on [22/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

14657287, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/coep.12610 by TPS: Western Economic Association International - WEAI - TPS, Wiley Online Library on [22/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

ENDNOTES

[1] "History of the California Building Code - Title 24, Part 2" https://www.dgs.ca.gov/BSC/About/History-of-the-California-Building-Code--Title-24-Part-2. Accessed January 22, 2023.

[2] Department of Energy, "Status of State Energy Code Adoption" https://www.energycodes.gov/status Accessed January 22, 2023. Infographic: Residential energy code: state energy index relative to current model code (2021 IECC).

[3] Holian (2023).

[4] As shown in Table A3, the construction period variables include YB1960, an indicator for homes built between 1960 and 1969 (YB stands for "year built"), and analogous variables for the homes built in the 1970s and 1980s. For homes built in the 1990s, the Census Bureau reported more refined information: YB1990 is equal to one for homes built between 1990 and 1994, YB1995 is one for homes built between 1995 and 1998, and YB1999 is one for homes built between 1999 and 2000.

[5] In the Appendix present and discuss robustness checks using American Community Survey data and additional control variables, including one measuring how long someone has lived in the house.

[6] There is only one difference between the model I estimate and the one presented in their column (1). They used the most recent period as a reference group, whereas I omit the 1970s period to facilitate interpretation of the results. This only affects the interpretation; the coefficients on the other variables are identical, as can be seen by reference to Costa and Kahn (2010, tab. 5), an earlier version of the published paper where the authors presented the full set of regression results.

[7] Levinson's sample size ranged from only 11,644 to 14,045 which could explain why his standard errors were high relative to the coefficient estimates. Another difference between our studies is that his dependent variable is log quantity while mine is log expenditures; I could divide expenditures by state average price and do my regressions in quantity too, but it wouldn't change the coefficient estimates.

[8] These calculations are $-0.031 \times 1128 = -34.968$ and $-0.076 \times 607 = -46.132$, respectively. Averages for COSTELEC and COSTGAS are reported in Table A2.

[9] There are important differences in the CBA from Novan et al. (2022) compared to the one I present in the next section, and it is worth highlighting one concerning inflation adjustments. In their NPV calculation, Novan et al. (2022) assumed constant nominal benefits while in Section 4, by contrast, I assume constant real benefits.

[10] This may not hold if energy providers are either earning monopoly rents or subsidizing customers.

[11] For example, with $\epsilon = \gamma = 1$ the equation reduces to a private NPV calculation. By making $\epsilon > 1$ and $\gamma > 1$ we can incorporate external benefits of energy savings, such as reduced carbon emissions from electricity generation and burning natural gas.

[12] They assumed resource cost of generation was 6 cents and added 2 cents for line losses. I assume the utility would charge consumers for line losses.

[13] In their Table A1, the low and high end external benefit of avoided natural gas emissions are \$0.92 and \$10.37, respectively, while avoided energy costs were \$21.9. Carbon dioxide emissions are responsible for over 90% of the external benefits in their analysis, though they also monetize impacts from sulfur dioxide, nitrous oxide, and particulate emissions.

[14] In Horn et al. (1980) climate zones were based on counties. Boundaries of California's climate zones were subsequently changed in 1982 after incorporating more refined meteorological information and no longer followed county boundaries; see Bruegge et al. (2019).

[15] Of course, some homes use wood and other less-common energy sources. To make this exposition more tractable, here I focus on just two energy mix types.

[16] Electricity generation in some areas is low-carbon (e.g., in areas with large hydroelectric facilities) while it is high-carbon in others (e.g., areas with many coal-fueled electricity generating facilities). See Holian and Pyeon (2017, tab. 29) for state-specific emissions factors.

[17] I did not use information from a few cities that straddle two climate zones and a few cities were not listed in CEC (1995). The City and County Data Book can be accessed at: http://web.archive.org/web/20120119122603/http://www.census.gov/statab/ccdb/ccdbcityplace.html (accessed January 22, 2023). See C-6 Government and Climate.

[18] Another promising angle for future research is, following Bruegge et al. (2019), to exploit the fact that some households that are near climate zone borders face a similar climate but a different regulatory environment compared to households on the other side of the border. Looking closely at Figure 1 we see there are very small PUMAs on both sides of the border between climate zones 11 and 12; there are also other areas around the state where this is true. Sometimes climate zones really do demarcate very different climates, but they sometimes also coincide with purely administrative boundaries which do not depend on climate.

REFERENCES

Bruegge, C., Deryugina, T. & Myers, E. (2019) The distributional effects of building energy codes. *Journal of the Association of Environmental and Resource Economists*, 6(S1), S95–S127. Available from: https://doi.org/10.1086/701189

California Energy Commission. (1981a) Financing the energy efficient home: analysis of the impact of the new residential building standards on housing costs. Proceedings of a conference held at the Sacramento Community Convention Center. Available from: https://hdl.handle.net/2027/uc1.31822003925252 [Accessed 22nd January 2023].

California Energy Commission. (1981b) Proposed residential building standards: final environmental impact report. Available from: https://hdl.handle.net/2027/uc1.31822040988347 [Accessed 22nd January 2023].

California Energy Commission. (1981c) New residential building standards: briefing package. Available from: https://hdl.handle.net/2027/uc1.31822040988354 [Accessed 22nd January 2023].

California Energy Commission. (1995) California climate zone descriptions for new buildings. Available from: https://web.archive.org/web/20190324082915/https://www.energy.ca.gov/1995publications/P400-95-041.pdf [Accessed 22nd January 2023].

Christensen, G. & Miguel, E. (2018) Transparency, reproducibility, and the credibility of economics research. *Journal of Economic Literature*, 56(3), 920–980. Available from: https://doi.org/10.1257/jel.20171350

Costa, D.L. & Kahn, M.E. (2011) Electricity consumption and durable housing: understanding cohort effects. *The American Economic Review*, 101(3), 88–92. Available from: https://doi.org/10.1257/aer.101.3.88

Costa, D.L. & Kahn, M.E. (2010) *Why has California's residential electricity consumption been so flat since the 1980s? A microeconometric approach. No. w15978*. National Bureau of Economic Research.

Hoces de la Guardia, F., Grant, S. & Miguel, E. (2021) A framework for open policy analysis. *Science and Public Policy*, 48(2), 154–163. Available from: https://doi.org/10.1093/scipol/scaa067

Holian, M.J. (2020) The impact of building energy codes on household electricity expenditures. *Economics Letters*, 186, 108841. Available from: https://doi.org/10.1016/j.econlet.2019.108841

Holian, M.J. (2021) *Data and the American dream: contemporary social controversies and the American Community Survey*. London: Palgrave Macmillan.

Holian, M. (2023) *COEP replication package for "Methods in open policy analysis: an application to California's building energy codes"*. Ann Arbor: Inter-University Consortium for Political and Social Research [distributor]. Available from: https://doi.org/10.3886/E187383V1

Holian, M.J. & Pyeon, J.-H. (2017) *Analyzing the potential of hybrid and electric off-road equipment in reducing carbon emissions from construction industries*. San Jose: Mineta Transportation Institute.

Horn, M. (1980) *High rise buildings*. Project Report No. 18. California Energy Commission. Available from: https://hdl.handle.net/2027/uc1.31970021339509 [Accessed 22nd January 2023].

Horn, M., Dean, T. & Borden, M. (1980) *Cost effectiveness. 1980 Residential Standards Development Project*. Sacramento: California Energy Commission. Available from: https://hdl.handle.net/2027/uc1.31970021339525 [Accessed 22nd January 2023].

Jacobsen, G.D. & Kotchen, M.J. (2013) Are building codes effective at saving energy? Evidence from residential billing data in Florida. *The Review of Economics and Statistics*, 95(1), 34–49. Available from: https://doi.org/10.1162/rest_a_00243

Kotchen, M.J. (2017) Longer-run evidence on whether building energy codes reduce residential energy consumption. *Journal of the Association of Environmental and Resource Economists*, 4(1), 135–153. Available from: https://doi.org/10.1086/689703

Levinson, A. (2016) How much energy do building energy codes save? Evidence from California houses. *The American Economic Review*, 106(10), 2867–2894. Available from: https://doi.org/10.1257/aer.20150102

Novan, K., Smith, A. & Zhou, T. (2018) *Residential building codes do save energy: evidence from hourly smart-meter data*. UC Davis.

Novan, K., Smith, A. & Zhou, T. (2022) Residential building codes do save energy: evidence from hourly smart-meter data. *The Review of Economics and Statistics*, 104(3), 483–500. Available from: https://doi.org/10.1162/rest_a_00967

Ruggles, S., Flood, S., Goeken, R., Grover, J., Meyer, E., Pacas, J., et al (2019) *IPUMS USA: version 9.0. dataset*. Minneapolis: IPUMS.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Holian, M.J. (2023) Methods in open policy analysis: an application to California's building energy codes. *Contemporary Economic Policy*, 1–16. Available from: https://doi.org/10.1111/coep.12610

14657287, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/coep.12610 by TPS: Western Economic Association International - WEAI - TPS, Wiley Online Library on [22/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

APPENDIX: TABLES AND FURTHER ROBUSTNESS CHECKS

TABLE A1 Variable descriptions.

Variable	Description
logCOSTELEC	Natural log of annual household electricity expenditures
logCOSTGAS	Natural log of annual household natural gas expenditures
YB1960	Indicator for homes built between 1960 and 1969
YB1970	Indicator for homes built between 1970 and 1979
YB1980	Indicator for homes built between 1980 and 1989
YB1990_94	Indicator for homes built between 1990 and 1994
YB1995_98	Indicator for homes built between 1995 and 1998
YB1999_00	Indicator for homes built between 1999 and 2000
AGE	Age of head of household
ROOMS	Number of rooms in home
logHHINCOME	Natural log of annual household income
HHSIZE	Number of persons in household
WHITE	White race indicator for head of household
ELEHEAT	Electric heat indicator for home
SEI	Socioeconomic Index for householder (composite variable)

Note: All variables are from the 2000 US Census, 5% sample, Ruggles et al. (2019).

TABLE A2 2000 5% census CA summary statistics.

Statistic	*N*	Mean	St. Dev.	Min	Max
logCOSTELEC	144,301	6.80	0.74	1.39	8.65
logCOSTGAS	135,477	6.16	0.74	1.39	8.37
COSTELEC	144,301	1128.41	810.81	4.00	5700
COSTGAS	135,477	607.60	505.76	4.00	4300
YB1960	144,914	0.22	0.41	0	1
YB1970	144,914	0.27	0.44	0	1
YB1980	144,914	0.26	0.44	0	1
YB1990_94	144,914	0.13	0.33	0	1
YB1995_98	144,914	0.09	0.29	0	1
YB1999_00	144,914	0.03	0.17	0	1
AGE	144,914	47.06	9.25	30	65
ROOMS	144,914	6.27	1.71	1	9
logHHINCOME	144,909	11.21	0.81	1.39	14.03
HHINCOME	144,909	98,209.21	83,832.37	4.00	1,237,000
HHSIZE	144,914	3.28	1.69	1	24
WHITE	144,914	0.72	0.45	0	1
ELEHEAT	144,914	0.16	0.36	0	1
SEI	144,914	0.49	0.26	0.00	0.96

Note: All variables are from the 2000 US Census, 5% sample, Ruggles et al. (2019). Summary statistics for the Sacramento subsample are available in the log file in the online supplement, https://www.openicpsr.org/openicpsr/project/187383.

14657287, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/coep.12610 by TPS: Western Economic Association International - WEAI - TPS, Wiley Online Library on [22/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE A3 Regression results, census 2000.

	Dependent variable			
	logCOSTELEC (1)	**logCOSTGAS (2)**	**logCOSTELEC (3)**	**logCOSTGAS (4)**
AGE	0.005***	0.004***	0.006***	0.002
	(0.0003)	(0.0005)	(0.001)	(0.001)
ROOMS	0.090***	0.072***	0.101***	0.078***
	(0.003)	(0.002)	(0.009)	(0.010)
logHHINCOME	0.116***	0.088***	0.111***	0.067***
	(0.004)	(0.004)	(0.014)	(0.017)
HHSIZE	0.056***	0.048***	0.058***	0.059***
	(0.004)	(0.003)	(0.009)	(0.012)
WHITE	0.086***	0.028***	0.091***	−0.024
	(0.008)	(0.006)	(0.011)	(0.028)
ELEHEAT	0.221***	−0.184***	0.307***	−0.417***
	(0.011)	(0.021)	(0.022)	(0.126)
SEI	0.066***	0.039***	0.071	0.048
	(0.010)	(0.011)	(0.062)	(0.076)
YB1960	−0.019***	0.036***	−0.054***	0.015***
	(0.007)	(0.007)	(0.006)	(0.004)
YB1980	−0.031***	−0.076***	−0.027***	−0.221***
	(0.008)	(0.007)	(0.007)	(0.036)
YB1990_94	−0.078***	−0.040***	−0.175***	−0.077***
	(0.010)	(0.008)	(0.012)	(0.011)
YB1995_98	−0.129***	−0.073***	−0.310***	−0.183***
	(0.012)	(0.007)	(0.011)	(0.012)
YB1999_00	−0.157***	−0.022*	−0.326***	−0.160***
	(0.019)	(0.012)	(0.015)	(0.013)
Observations	144,296	135,473	6582	5801
R^2	0.164	0.119	0.162	0.145
Adjusted R^2	0.163	0.117	0.160	0.142
Residual Std. Error	3.014	3.108	3.128	3.263

Note: This table shows the full set of regression coefficients (except PUMA fixed-effects) whereas Table 1 only showed estimated coefficients on YB1980. The data rules are the same as those described under Table 1.

Abbreviation: PUMA, Public-Use Microdata Area.

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

As an additional set of robustness checks, I use the 2012 to 2017 American Community Survey (ACS) data to estimate models that are very similar to those in Table 3. The numbers in Table A4 are coefficient estimates on the 1980s construction period dummy (YB1980). The models used to produce the estimates in Table A4 differ from the ones in the body of this paper that used the 2000 5% sample in that the control strategy is slightly different. Holian (2020) estimated a version of the model presented in Table A4, row one, column (1), which differed only in that a different reference group was chosen for the decade of construction dummies. The analysis reported in Table A4 reanalyzes the model from Holian (2020) by adding natural gas as a dependent variable, by presenting results for levels and logs, and by extending the analysis to detached homes.

14657287, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/coep.12610 by TPS: Western Economic Association International - WEAI - TPS, Wiley Online Library on [22/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE A4 Robustness checks, ACS Data 2012–2017.

	Model and subsample			
	2012–17, log	2012–17, level	2012–17, log	2012–17, level
	Detached and duplexes		Detached only	
	(1)	(2)	(3)	(4)
$\beta_{CA}^{Electric}$	−0.011	−8.393	−0.003	0.134
β_{CA}^{NGas}	−0.023***	−5.461	−0.012*	−0.494
$\beta_{Sac}^{Electric}$	0.021**	17.931	0.021**	14.617
β_{Sac}^{NGas}	−0.029	−2.266	−0.030	−7.811

Note: The data used are from the 2012–2017 ACS and were obtained from Ruggles et al. (2019). The sample includes only single-family, owned homes constructed in California between 1960 and 2000, where the head of household is aged 30–65, and the household has nonnegative income. All models include PUMA-level fixed effects, and standard errors are clustered at the electricity district level. The omitted construction decade dummy is the 1970s.

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

The results of the analysis of the ACS data are surprising, in that they show *positive* though mostly statistically insignificant effects for Sacramento. For California as a whole, the evidence for a difference between 1980 and 1970 homes in terms of electricity is weak, as none of the coefficient estimates are statistically significant. The effect of gas is significant in column (1) for California, but insignificant or marginally significant in the other three columns. One possible explanation for the insignificant findings in Table A4 could be that many homes built in the 1970s and 1980s were retrofitted with new windows, weatherstripping, etc. between 2000 and 2017, making them similar in terms of their level of energy efficiency.

14657287, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/coep.12610 by TPS: Western Economic Association International - WEAI - TPS, Wiley Online Library on [22/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License