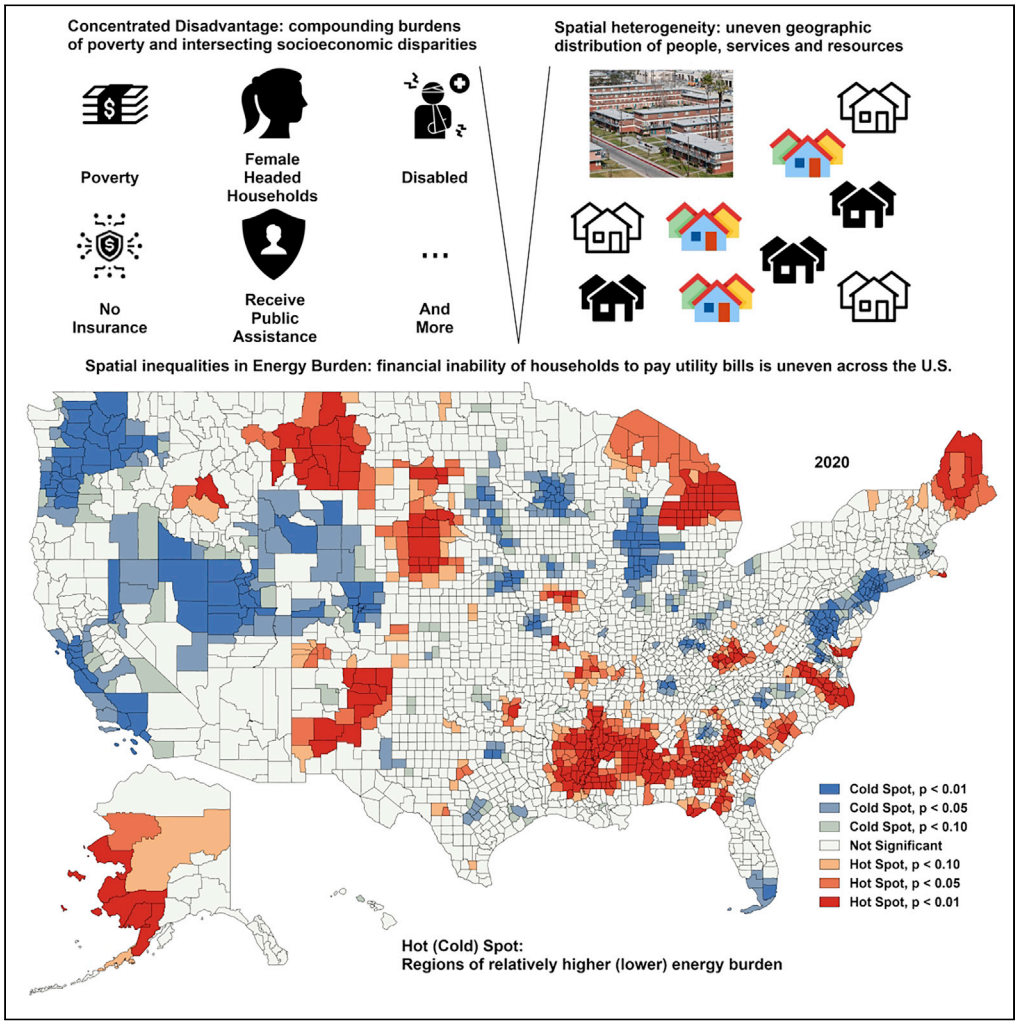
iScience

CellPress
OPEN ACCESS

Article

Localized energy burden, concentrated disadvantage, and the feminization of energy poverty



Chien-fei Chen, Jimmy Feng, Nikki Luke, Cheng-Pin Kuo, Joshua S. Fu

cchen26@utk.edu

Highlights

Energy burden arises from racial, socioeconomic, and spatial stressors

This study shows the local effects of concentrated disadvantage on energy burden

The counties with households headed by Black women have a severe energy burden

Policies to address energy and healthcare issues for low-income areas are needed

Chen et al., iScience *25*, 104139
April 15, 2022
https://doi.org/10.1016/j.isci.2022.104139

iScience

CellPress
OPEN ACCESS

Article

Localized energy burden, concentrated disadvantage, and the feminization of energy poverty

Chien-fei Chen,[1,5,*] Jimmy Feng,[2] Nikki Luke,[2] Cheng-Pin Kuo,[3] and Joshua S. Fu[3,4]

SUMMARY

Energy burden directly influences households' health and safety. Amid a growing literature on energy, poverty and gender remains relatively understudied. We evaluate socioeconomic, geographic, and health factors as multidimensions of concentrated disadvantage that magnify energy burden in the United States over time. We show that the energy burden is more pronounced in disadvantaged counties with larger elderly, impoverished, disabled people, and racialized populations where people do not have health insurance. Neighborhoods with households headed by women of color (especially Black women) are more likely to face a high energy burden, which worsened during the COVID-19 pandemic. Although energy costs are often regarded as an individual responsibility, these findings illustrate the feminization of energy poverty and indicate the need for an intersectional and interdisciplinary framework in devising energy policy directed to households with the most severe energy burden.

INTRODUCTION

Energy poverty magnifies the material, health, educational, and social stressors of poverty and creates barriers for participation in society (Bouzarovski, 2018). Research on energy burden, which is a measure of energy poverty that describes the percentage of household income spent on utility expenditure, such as bills for electricity, gas, and water, shows compounding negative effects on the mental and physical health of vulnerable populations (Hernández, 2013; Mayer and Smith, 2019) that exacerbate pervasive social inequalities, especially during the COVID-19 pandemic (Castán Broto and Kirshner, 2020; Chen et al., 2020, 2021; Graff and Carley, 2020; Memmott et al., 2021). Several factors contribute to the high energy burden; however, even after controlling for household size, age, heating source, and local weather, high electricity consumption remains a key component (U.S. Department of Energy Office of Energy Efficiency and Renewable Energy, 2018). In addition, in the United States, low-income, Black/African American, Hispanic/Latino, multifamily, and renting households, on an average, consume less electricity than their counterparts, yet have higher energy use when normalized by housing quality and efficiency (Reames, 2016).

Although these disparities are geographically specific, neither national or state averages nor cross-sectional survey analyses (i.e., non-longitudinal studies) can demonstrate the particular and long-term effects of energy burden in vulnerable households. There are significant regional and local variations (i.e., county-level) in the energy burden that low-income households (LIHs; defined as earning less than 80% of the Area Median Income) face; for example, 38% of LIHs in southeast states suffer an energy burden of 6% or higher, compared to 29% of LIHs in other regions of the United States. Further, LIHs across the United States. spend up to three times more of their income on utility bills than higher-income households (Drehobl et al., 2020). More seriously, energy burden follows a pattern sociologists have described within the framework of *concentrated disadvantage*, whereby low-income areas, and especially communities of color, are characterized by intersecting and compounding socioeconomic disparities, such as high levels of households headed by a single parent, rentership, and disability (Sampson et al., 2008). With the recent growth in studies of energy insecurity, this article aims to situate energy as it relates to social stressors to understand changes in energy burden in the United States over time while also accounting for the effects of the COVID-19 pandemic. Echoing recent studies that point to the feminization of energy poverty outside of the United States (Casabonne et al., 2019; Nguyen and Su, 2021; Petrova and Simcock, 2019), our results indicate the need for an intersectional account of the magnified effects of gender, racial, and income

[1]Center for Ultra-wide-area Resilient Electrical Energy Transmission Networks (CURENT), University of Tennessee, TN, USA

[2]Department of Geography, University of Tennessee, Knoxville, TN, USA

[3]Department of Civil & Environmental Engineering, University of Tennessee, Knoxville, TN, USA

[4]Computational Science and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA

[5]Lead contact

*Correspondence: cchen26@utk.edu

https://doi.org/10.1016/j.isci.2022.104139

disparities in energy burden (Crenshaw, 1989). This analysis recognizes that neither gender nor race is reducible to class and calls for greater attention, as several key studies have already evidenced, to the specific social and historical patterns that reproduce gender and racial inequalities in energy (Luke, 2021; Petrova and Simcock, 2019; Reames, 2016).

The extant literature provides important contributions to understand the combination of spatial and social dimensions of energy poverty (Gouveia et al., 2019; Mashhoodi et al., 2019; Robinson et al., 2019; Walker et al., 2012), although analyses of the intersections of geography, race, and gender have been more limited. This literature also contains limitations that arise in part from scant data on energy use among vulnerable populations in the United States. First, nationally representative data such as the Residential Energy Consumption Survey (RECS) does not provide state or county-level geographical identifiers or detailed demographic information. Second, the widely used American Community Survey (ACS) did not capture the most updated energy-specific expenditures (e.g., electricity, natural gas) in 2019 or during the pandemic. Third, recent efforts to use survey data to fill these gaps rely on limited sample observations over time (Memmott et al., 2021). Finally, limited research has been devoted to local geographies as a social context that may influence energy burden and insecurity. To address these shortcomings, this study examines the relationship between *concentrated disadvantage* in socioeconomic, race/ethnicity, gender, and energy burden across a range of spatial and temporal scales from 2014–2018 and during the COVID-19 pandemic in 2020.

Conceptualizing energy burden and concentrated disadvantage

This study analyzed the five-year average energy burden of the U.S. low-income households between 2014–2018 and 2020 and calculated the annual energy burden from the monthly average at the county and state levels. The data are from Low-income Energy Affordability Data (LEAD) tool by the U.S. Department of Energy and the Energy Information Administration (EIA) (U.S. Department of Energy Office of Energy Efficiency and Renewable Energy, 2021).

We bring these data into conversation with sociological literature on *concentrated disadvantage*, which is guided by the understanding that spatial relationships that are determined through interrelated and interdependent political, economic, and social relations define the everyday lived experience of communities (Massey and Denton, 1993; Sampson et al., 2008). The concentrated disadvantage framework is used to explain the intensification of inequalities in ways that systematically produce less favorable outcomes for individuals or groups within specific locations. Recent scholarship has combined concentrated disadvantage with issues of environmental justice to analyze the negative health effects associated with lead and ambient environmental pollution (Liévanos, 2019; Mennis et al., 2016; Winter and Sampson, 2017). We contend that this framework can also be applied to the concentrated and uneven spatial distribution of high energy burden.

Energy burden is in part linked to historic patterns of racial discrimination in the lending and other racist housing policies, as well as racial disparities in poverty and accumulated wealth (Lewis et al., 2020). This study accounts for electricity price, location, and demographic indicators to substantiate that energy burden is not only a result of higher rates of an individual or household poverty among racialized and ethnic groups but also because of the high rates of residential racial segregation experienced by people of color, particularly in predominately Black communities. The concentrated disadvantage neither been used to investigate energy burden despite growing attention to the racial, class, and geographic disparities in energy burden in the United States (Bednar and Reames, 2020; Drehobl et al., 2020) nor has been used to investigate energy burden and account for the interactions between gender, racial, and economic factors. These disparities, alongside inequality in access to healthcare, have also shaped differences in COVID-19 exposure and death. In this study, we are less focused on the direct effects of COVID-19 on energy burden, which other analyses have expertly examined (Chen et al., 2021, p. 19; Graff and Carley, 2020; Memmott et al., 2021), and are instead interested in trends in energy burden over time and interconnected factors influencing energy burden. More importantly, this study expands prior uses of the concentrated disadvantage framework that focus on limited variables, namely the share of people in poverty, unemployed, under the age of 18, who receive public assistance, and who live in female-headed households, as well as the racial/ethnic composition of neighborhoods (Life Course Metrics Project, 2013). Here, we explore ten socioeconomic and health variables (see Table 3), as well as the geographic differences between rural and urban settings.

Table 1. Comparison of energy burden by county and state between 2020 and 2014–2018

County - state	Energy burden 2020 (%)	Energy burden 2014-8 (%)	Energy burden difference (%)	Expenditure 2014-8 ($)	Expenditure 2020 ($)	Expenditure difference ($)
Quitman GA	30.45	30	0.45	403.42	409.5	6.09
Hyde NC	28.15	29	−0.85	437.92	425.04	−12.87
Buffalo SD	26.64	26	0.64	451.42	462.58	11.16
Lake MI	26.29	25	1.29	407.08	428.11	21.02
Carter MT	25	24	1	472.75	492.51	19.76
Harding NM	22.77	21	1.77	341.67	370.47	28.8
Catron NM	22.72	21	1.72	359.58	389.12	29.53
Greene AL	22.49	23	−0.51	254	248.42	−5.58
Danville City VA	21.68	22	−0.32	344.75	339.72	−5.03
McCone MT	21.64	21	0.64	452.17	465.88	13.72
Bullock AL	21.56	22	−0.44	227.58	223.06	−4.52
Franklin FL	21.39	20	1.39	308.83	330.26	21.43
Clare MI	21.18	20	1.18	344.5	364.8	20.3
Treasure MT	20.95	20	0.95	463.25	485.29	22.04
Powder River MT	20.89	20	0.89	431.17	450.31	19.15
Alcona MI	20.85	20	0.85	342.67	357.29	14.62
Edwards TX	20.68	20	0.68	341.75	353.31	11.56
Stewart GA	20.41	20	0.41	236	240.84	4.84
Holmes MS	20.35	21	−0.65	233.08	225.91	−7.18
Jackson SD	20.23	20	0.23	322.08	325.77	3.68
Crawford MI	20.19	19	1.19	338.33	359.51	21.17

This Table lists those counties with an energy burden of at least 20% across the United States in 2020.

RESULTS AND DISCUSSION

Energy burden shows a strong local-level pattern

Average energy burden in 2020 and the five-year average of 2014–2018, across state and county levels show that, in 2020, approximately half of LIHs in 24 states had an energy burden greater than 10%, with the highest percentage of LIHs in Mississippi (13.48%), followed by 13.02% in Maine and 12.97% in Alabama (Table 1). The energy burden patterns are similar between 2014–2018 and 2020, yet there is a noticeable monthly difference (Figure 1). Overall, the energy burden is most severe during the winter and summer months. However, our comparison finds that people in 2020 had a relatively lower energy burden during January, February, March, October, and November compared to 2014–2018. Most stark is the nearly 2% decrease in energy burden in January 2020 compared to January 2014–2018. Between May and July of 2020, the energy burden rose relative to the 2014–2018 period, which may be because of stay-at-home orders during the COVID-19 pandemic (Figure 1, also see Figure S3) (Chen et al., 2021). At the state level, residents in Alaska noticeably contributed a significantly higher share of their income to energy in the first four months of the year in both 2014–2018 and 2020, as shown in Figure 2 (See Figure S1 for 2014–2018).

Our analysis at the county level indicates spatial heterogeneity or local variation in energy burden across the country is more significant than the differences observed at the state level. For example, in 2020, the energy burden ranges from 6.04% to 13.48% at the state level compared to 2.93%–30.45% at the county level. At the county level, Quitman County, Georgia, has the highest monthly energy burden of 30.45%, virtually the same as the 2014–2018 energy burden (30.00%). In this part of Georgia, the energy burden has not improved over time. This more granular analysis points to potential aggregation effects at the state level that may obscure understanding of energy burden and analysis at more local levels. Therefore, finer-scale analysis provides a more nuanced and accurate snapshot.

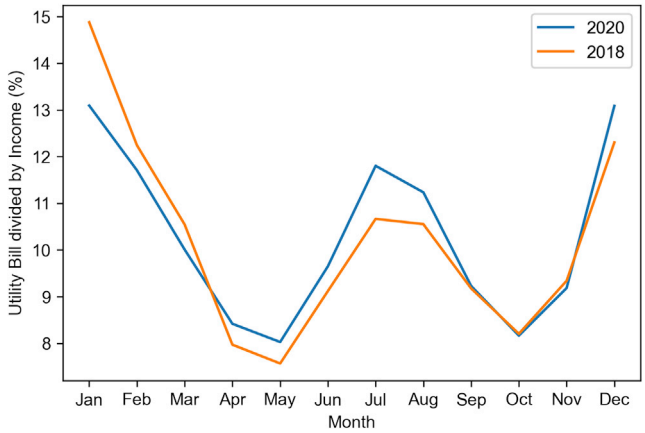


Figure 1. Comparison of monthly energy burdens (%/month) for the five-year average of 2014–2018 and 2020

At the regional level, energy burden rose across the Mountain West, including Nevada, Utah, New Mexico, Arizona, Idaho, Montana, and Wyoming; parts of the Midwest, including Michigan and Wisconsin; and some outlier states, including West Virginia, Oklahoma, Delaware, and Florida, as depicted in Figure 4. For many of these states, there have been more record-setting temperatures and increasingly volatile weather events (e.g., droughts and wildfires) that could contribute to both higher utility bills (resulting from heating and cooling costs) and a diminished ability to pay unexpected bills. Conversely, we observe a few cold spots or clusters of decreased energy burden in parts of Vermont, North Carolina, Illinois, Kentucky, Mississippi, and Alabama, which corroborates the visual analysis of the map of energy burden between 2014–2018 and 2020 in Figure 3.

While examining specific states, we found that Delaware has the largest increase in monthly energy burden (1.61% or an additional $37.53 per month). In addition, there are six states that have at least a 0.5% relative increase in their energy burden between 2014–2018 and 2020, which also increased annual energy costs in 2020, including Arizona (1.03%; $203.84), New Mexico (0.81%; $158.51), Michigan (0.57%; $131.92), Hawaii (0.55%; $179.35), Nevada (0.55%; $146.08), and Florida (0.51%; $109.00). However, analysis at the county level shows notable local incidents where a higher energy burden was witnessed: twelve counties' energy burden increased over 1%, ranging from a $225.96 to $354.36 increase in yearly utility costs.

Local effects of concentrated disadvantage and energy burden

We observed the compounding effects of social factors on energy burden by testing for an expanded set of variables that have been used to describe conditions of concentrated disadvantage. Our first step in this analysis focused on differences in energy expenditure (i.e., the number of energy bills) and energy burden in rural and urban areas. There is a notable difference in the amount of money allocated to utility bills between urban (n = 1,321) and rural (n = 1,821) counties across the country (Table 2). Rural LIHs spent 25.2%

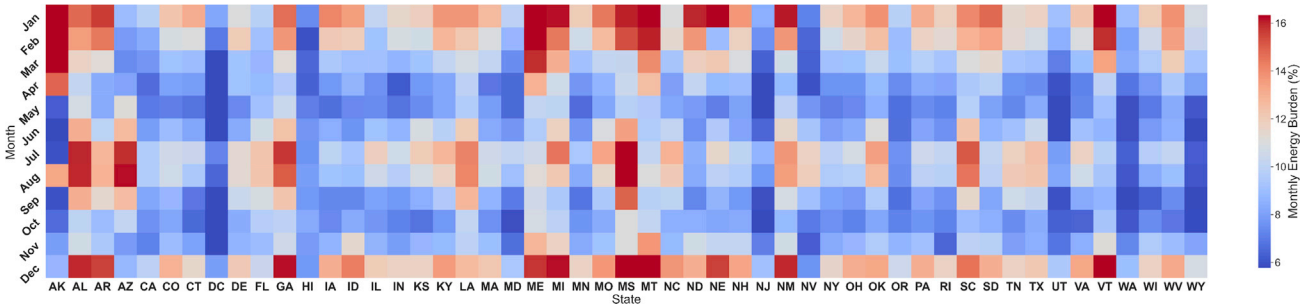


Figure 2. Average energy burdens by state from January to December 2020

Darker red colors indicate a higher energy burden, whereas darker blue colors indicate a lower energy burden.

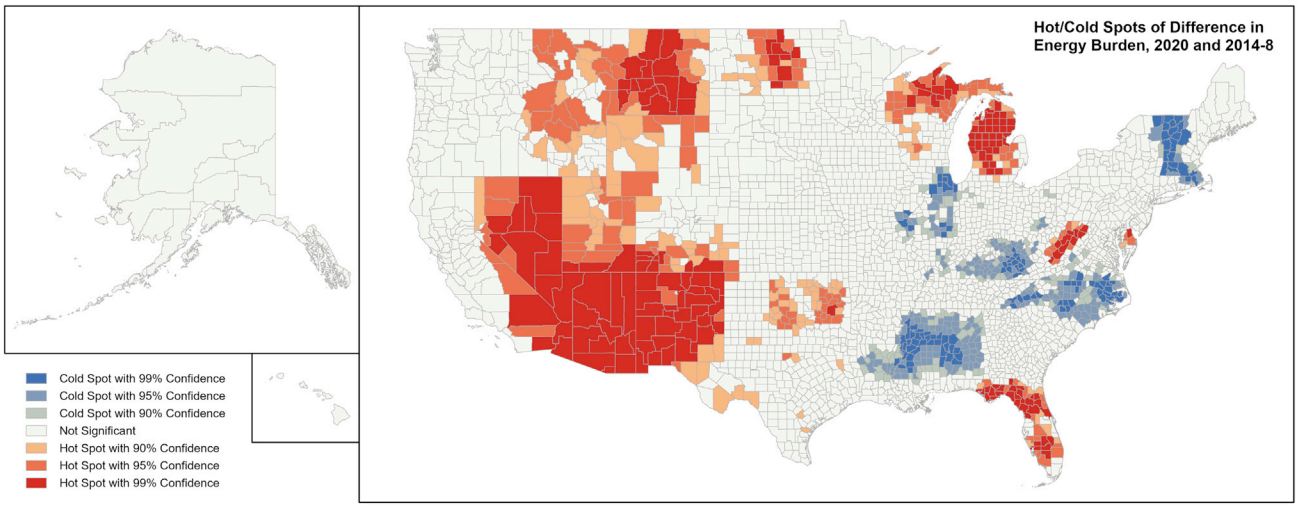


Figure 3. Hot/Cold spots of difference in energy burden, 2014–2018 vs. 2020

A statistically hot/cold spot is a feature that has both a high/low energy burden and has neighboring feature locations that have a similar high/low value. Cold and hot spots are classified by their significance levels. Blue indicates statistically significant spatial clusters of counties with the decreased difference in energy burden, whereas red spots indicate an increased difference in energy burden between 2014–2018 and 2020.

more on monthly utilities (additional $38.37/month, $460.44/year), compared to their urban counterparts between 2014 and 2018. This disparity grew marginally larger in 2020, as rural LIHs spent 25.4% more on utilities (an additional $39.59/month, $475.08/year) than urban LIHs. Analysis of energy burden shows a similar pattern: rural areas have an average monthly burden of 10% or more, which is higher than urban areas (Table 2).

Our second analysis on the concentrated disadvantage shows the potential impacts of socioeconomic and health factors on energy burden. Our estimated models include ten independent variables: the percentage of (1) non-white people: *PerMinorit*, (2) people receiving public assistance via social welfare and social insurance programs: *PerSSI*, (3) people who are disabled people: *PerDisable*, (4) people over age 65: *PerAge65*, (5) people under age 18: *PerAge17*, (6) people with household incomes at or below 200% of the Federal Poverty Level (FPL): *PerPoverty*, (7) people without insurance: *PerNoInsur*, (8) female-headed households: *PerFemHead*, (9) people with COVID-19 in 2020: *PercCase20*, (10) people who died because of COVID-19 in 2020: *PercDeat20*, and (11) dummy variables for each state, excluding North Dakota, to estimate our dependent variable, energy burden. North Dakota was excluded to avoid the dummy trap, as it is the state with the energy burden closest (10.28%) to the average national energy burden (10.30%). As opposed to traditional OLS models, we also explicitly included spatially-lagged variables that relate the values at a given location to those in surrounding and nearby locations, which can reflect how geography may mediate energy burden and capture the observed spatial dependence. In effect, we can also capture site and situational effects of concentrated disadvantage and energy burden.

Results of our spatial regression model with a spatially-autocorrelated error term and spatially-lagged independent and dependent variables find all of the explanatory variables, excluding *PerSSI*, *PerFemHead*, *PercAge17*, *PercCase20*, and *PercDeat20*, to be statistically significant at $p < 0.05$. All else were equal; a 1% increase in the population at or below 200% of FPL correlates to a 0.149% increase in energy burden. A 1% increase in the following factors also leads to corresponding changes in energy burden: without health insurance (0.055% increase), who are racial minorities (0.013%), who are disabled people (0.114%), and who are at least the age of 65 (0.241%). The age of 65 or over appears to be the strongest predictor.

More importantly, there are also indirect effects that can be quantified by the change in the spatial lag of each independent variable. Most spatially lagged terms, excluding *PerPoverty*, are statistically significant at $p < 0.05$. A 1% increase in the average level of neighboring counties' minority population and youth population would increase energy burden by 0.016% and 0.041%, respectively, in the typical county. Conversely, a 1% increase in the average population on some type of public assistance in neighboring counties corresponds to a reduction of 0.046%, whereas a 1% increase in female-headed households in

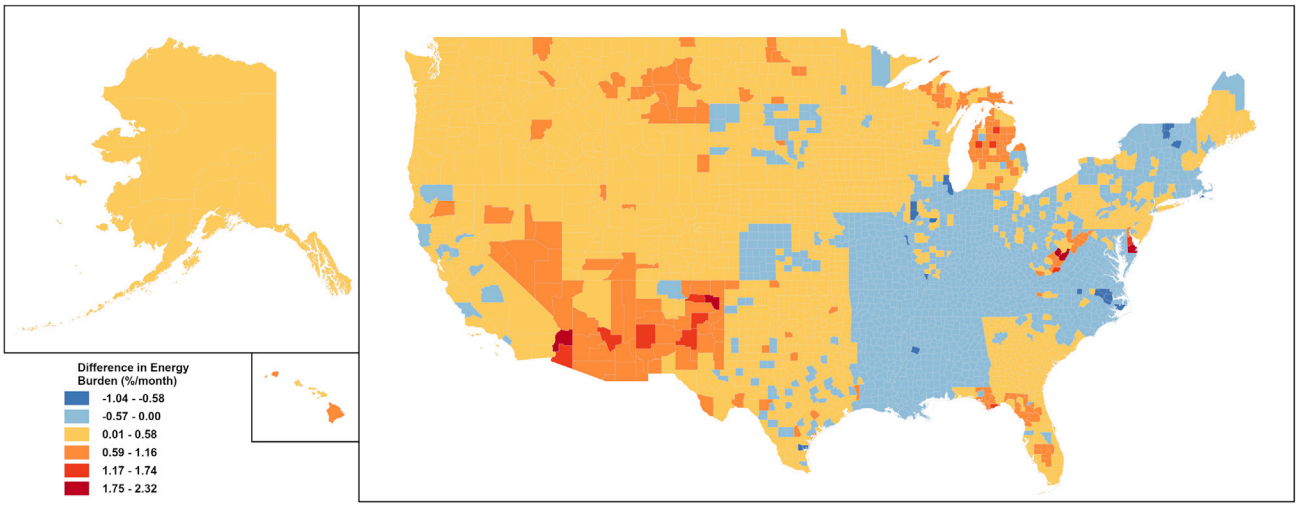


Figure 4. Percentage of monthly energy burden differences at the county level between 2014–2018 and 2020

A positive value, indicated by hues of yellow, orange, and red, represents counties that have experienced more considerable energy burdens in 2020 compared to 2014–2018. Meanwhile, a negative value, indicated by hues of blue, represents counties that have experienced a reduction in their energy burden in the same time frame.

neighboring counties corresponds to a reduction of 0.110% in energy burden after controlling all other factors. These findings indicate that public assistance to support household income may reduce the energy burden. Yet, more information is needed to understand whether this is because households already receiving public assistance are more likely to be aware of or participating in seasonal, federal programs for utility bill assistance (i.e., LIHEAP) and/or rate-payer funded initiatives offered in some states to cap energy costs to a percent of income. Meanwhile, a 1% increase in energy burden across an average set of neighboring counties correlates to a 0.544% increase in energy burden in the typical county. Although the variable of families with a female head shows a negative relationship with energy burden after accounting for other factors, we further examine this predictor in various contexts in the next section.

Racial/ethnic minorities, public assistance, and the feminization of energy poverty

To gain a deeper understanding of the influence of concentrated disadvantage on energy burden, we further analyzed three interconnected variables, including female-headed households, public assistance, and minority status, to investigate why these variables had a negative relationship to energy burden in the earlier models. First, we hope to understand where clusters of counties with a high minority population and high energy burden exist. Second, we examine differences in energy burden across racial/ethnic groups by analyzing the relationships between women of color-led households, public assistance, and COVID-19 death rate to consider the impacts of the COVID-19 pandemic in investigating energy burden in 2020.

To understand the relationship between specific minority populations and energy burden, we carry out bivariate local indicators of spatial autocorrelation test. The Moran's I statistic describes the pattern of spatial autocorrelation – clustered, dispersed, or random; a positive value indicates that counties are likely to be nearby other counties with a similar value (clustered). A negative value indicates that counties are likely to be nearby other counties with a dissimilar value (dispersed). Most Moran's I statistical values in our models are relatively small but significant, indicating weak patterns of spatial autocorrelation. More importantly, we find different patterns by racial/ethnic group at the county level. Notably, the Moran's I statistic for Black populations and energy burden is 0.19, indicating that counties are more likely to be around the counties that share a similar local relationship. For example, counties with a relatively large Black population and energy burden are likely to be surrounded by other counties with the same characteristics. The corresponding map (Figure 5) below illustrates some clusters that have both a relatively large Black population and high energy burden, most notably in the southern United States. In examining the Black populations and the number of COVID-19 deaths using the OLS multiple regression model to predict energy burden in 2020, we found that a county-level increase of 1% of the Black population correlates to a 0.045%

Table 2. Mean and median energy burdens of LIHs in rural and urban areas in the U.S., 2014–2018 & 2020

Living areas	Energy burden 2020 (%)	Energy burden 2018 (%)	Energy burden difference (%)	Expenditure 2020 ($)	Expenditure 2018 ($)	Expenditure difference ($)
a. Mean energy burdens						
Rural	11.53	11.42	0.11	214.28	211.80	2.41
Urban	8.60	8.54	0.05	174.69	173.44	1.25
b. Median energy burdens						
Rural	11.15	11.00	0.15	202.97	201.73	1.24
Urban	8.38	8.00	0.38	175.46	173.08	2.38

Counties are identified as urban if they, in any way, spatially intersect "urbanized areas (UA) and urban clusters (UC)" as classified by the 2010 U.S. Census. UAs comprise contiguous census blocks that contain at least 50,000 people, whereas a UC contains between 2,500 and 50,000 people. Based on Shapiro–Wilk tests, data for both urban and rural counties were can be inferred to have an nonnormal distribution ($p < .001$). In addition, ensuing results from a Kruskal-Wallis H-test to compare the median of each population strongly support a statistically significant difference between energy burden in rural and urban counties ($p < .001$).

increase in energy burden, in comparison with the nonblack population. Counties with a significant Hispanic population show slightly different geographical patterns in 2020. For instance, many counties in New Mexico have both a relatively large number of Hispanic residents, ranging from 5.51% to 54.85%, and a high monthly energy burden, ranging from 5.29% to 22.77% (Figure 6). Specifically, Lincoln with 40.7% Hispanic has 19.6% energy burden, Guadalupe with 37.1% Hispanic has 19.5% energy burden, and Socorro with 37.1% Hispanic has 16.1% energy burden are a few counties in New Mexico where a sizable cluster of both a large Hispanic population and high energy burden exists.

Second, we conducted three steps of analysis by performing a series of OLS regression models to examine how households led by women of color are affected by a high energy burden. These models also included the overall minority population and public assistance receipt, after accounting for the factors of 50 states and Washington D.C. (indicating different socioeconomic status, electricity price, and weather) and COVID-19 death rate in each county. While examining only the predictor of female-headed households, we found that counties with a higher percentage of families with female heads ($B = .10$; $p < .001$) saw their energy burden increase in 2020 after controlling for COVID-19 death and states, which is the opposite result from the earlier model (see Table 3). The earlier model indicates that the effect of female-headed households becomes less important (negative) when including other socioeconomic factors, such as race/ethnicity and poverty. Results of this model reveal a direct positive influence of female-headed households on energy burden.

While comparing differences across racial groups in female-headed households, we found households headed by Black women faced the highest energy burden in 2020 ($B = 0.212$; $p < .001$), higher than those headed by white, Hispanic, Native Hawaiian or Pacific Islander, and American Indian or Alaskan Native women, though there is no significant relationship between Asian female-headed households and energy burden, after accounting for COVID-19 death and states. After considering specific female minority groups, public assistance, and COVID-19 deaths based on five OLS regression models, we found that counties with a higher percentage of Black women household heads ($B = 0.785$; $p < .001$) have greater energy burden than the counties with a large overall Black population ($B = -0.770$; $p < .001$), receipt of public assistance ($B = 0.394$; $p < .001$), or COVID-19 mortality ($B = 0.212$; $p < 0.001$). In contrast, counties with a higher percentage of white overall female-headed household are less likely to have higher energy burden ($B = -0.588$; $p < .001$) than overall white population ($B = 0.635$; $p < .001$), recipients of public assistance ($B = 0.400$; $p < .001$), female-headed households ($B = 0.158$, $p < .001$), and counties with high rates of COVID-19 deaths ($B = 0.085$; $p < .001$). Examining the Hispanic population model, we found Hispanic female household heads are not significant to their energy burden, whereas counties with a higher percentage of overall female-headed households ($B = -0.390$; $p < .001$) are less likely to have a higher energy burden. However, the COVID-19 mortality rate ($B = 0.092$; $p < .001$) and public assistance recipients ($B = 0.401$; $p < .001$) are positively related to energy burden. This result is similar to the findings for Native Hawaiian and Pacific Islander, American Indian and Alaskan Native, and Asian women-led households. For all racial/ethnic groups except for Black women-headed households, COVID-19 mortality and public assistance are more important influencers of energy burden.

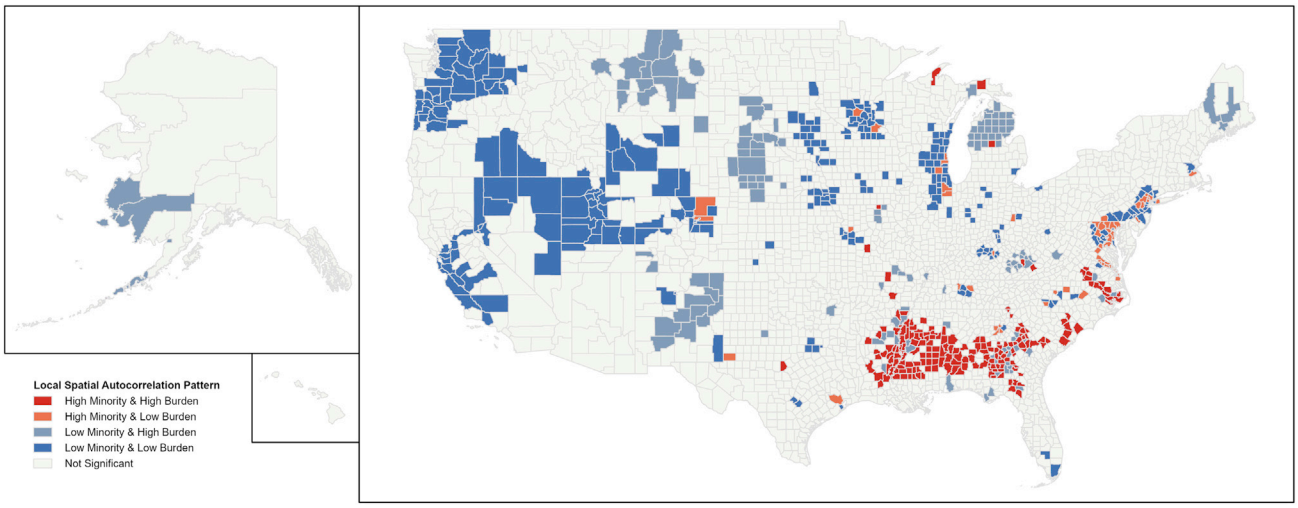


Figure 5. Bivariate hot spot map: Black population and energy burden in 2020
Deep red indicates counties with a large minority population and high energy burden; lighter red indicates counties with a large minority population and low energy burden; light blue indicates counties with a small minority population and high energy burden; deep blue indicates the counties that have a small minority population and low energy burden.

DISCUSSION

This study analyzes a novel dataset compiled from publicly available sources to examine the geographic, socioeconomic, and health factors that magnify energy burden through the lens of concentrated disadvantage. We have identified a global positive spatially autocorrelated pattern of energy burden in the U.S., such that counties are likely to be nearby other counties with similar values, either high or low. We find that energy burden is more pronounced in counties with larger elderly, impoverished, disabled, and racialized populations, as well as in counties where people are less likely to have health insurance. While examining these intersecting socioeconomic factors that contribute to concentrated disadvantage, we first find counties where a larger share of the population receives public assistance show lower levels of energy burden, suggesting that financial assistance can help reduce inequitable energy burden in the most impacted areas. However, existing federal programs (i.e., the Weatherization Assistance Program and the Low-Income Home Energy Assistance Program) that assist with energy bill payment and aim to improve underlying conditions that contribute to energy vulnerability, including housing quality, the functionality of

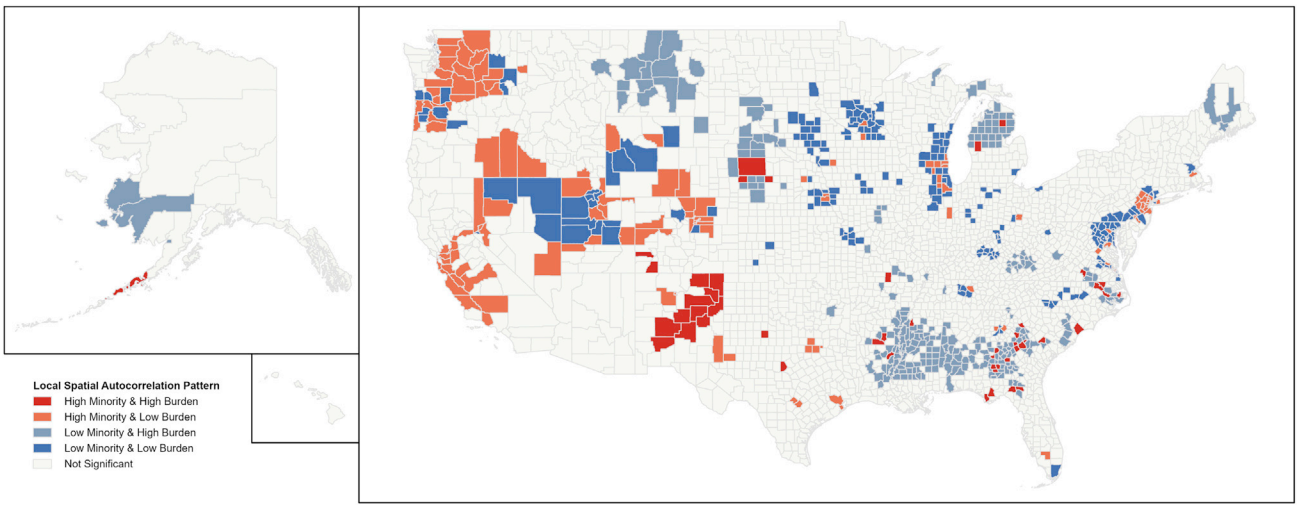


Figure 6. Bivariate hot spot map: Hispanic population and energy burden in 2020
The color scheme is the same as is described in Figure 5.

Table 3. Results of a spatial regression model with a spatially-autocorrelated error term and spatially-lagged concentrated disadvantage variables and energy burden in 2020

Independent variables	Unstandardized coefficient (b)	S.E.	z-Statistic	p value
Public assistance	−0.009	0.006	−1.468	0.142
No insurance	0.055	0.012	4.465	0.000
Minority status	0.013	0.004	2.961	0.003
Female household heads	−0.019	0.012	−1.620	0.105
Disable	0.114	0.014	8.222	0.000
Below poverty	0.149	0.012	12.662	0.000
Age 65 or older	0.241	0.013	18.660	0.000
Age 17 or younger	0.014	0.016	0.874	0.382
COVID-19 cases	−0.022	0.020	−1.079	0.280
COVID-19 death	0.621	0.556	1.118	0.264
w-public assistance	−0.046	0.011	−4.361	0.000
w-minority group	0.016	0.005	3.192	0.001
w-age 17 or younger	0.041	0.020	2.043	0.041
w-poverty	0.012	0.019	0.635	0.526
w-female household head	−0.110	0.011	−9.977	0.000
w-P2020 (energy burden)	0.544	0.023	23.947	0.000
Lambda	−0.099			
F-statistic	(67, 3074) = , 72.848***			
N	3142			
Adjust R^2	0.593			

*$p < 0.05$, **$p < 0.01$, ***$P < 0.001$; coefficients of categorical variables, including 50 states and Washington D.C, were considered as controlled variables, and therefore, were not listed here. Each of the 'w-' variables relate the values at a given county to those in surrounding counties. In contrast to the coefficients of the other variables, which reflect global averages, variables such as w-P2020 reflect the average energy burden of every county's neighboring counties. A 1% increase in the energy burden of surrounding counties correlates to a 0.544% increase in energy burden for the average county, and this interpretation can be held for all other 'w-' variables.

heating/cooling equipment, and utility debt management, are underfunded and limited in scope (Bednar and Reames, 2020; Franklin et al., 2017).

Later, we found that when focusing only on the intersections of COVID-19 mortality and public assistance receipt in households headed by women of different racial/ethnic backgrounds, public assistance is linked to a higher energy burden. This indicates the need for additional research to understand whether and how women householders access energy assistance programs and the need for targeted efforts to link women, householders to weatherization and programs that address factors contributing to the likelihood of becoming energy poor. This analysis also revealed certain limits to the concentrated disadvantage thesis that occluded the relationships between gender and energy insecurity. Testing for this relationship specifically, we observed that women of color, especially Black women heads of households, are more likely to face a high energy burden, which worsened during the pandemic. Women lost employment at higher rates than men and Black women were especially affected. The pandemic worsened employment outcomes in a labor market that is already highly unequal because of employment segregation, racial discrimination, lack of affordable childcare, and insufficient paid family leave, meaning many women may be forced to take unpaid leave to care for family (Wilson, 2021). Following research outside of the United States (Morris et al., 2019), this finding indicates the need for an intersectional framework that addresses gender and racial disparities in energy policy and programs. Building energy justice into policy design would also require engaging with the most affected people in decision-making processes that govern changes in energy infrastructure (Baker et al., 2019; Carley and Konisky, 2020).

Although our study attempts to capture changes across time, this perspective also offers contemporary insight into the context of the COVID-19 pandemic. We find that counties with a higher COVID-19 mortality

rate are observed also to have a higher energy burden when examining only minority status, public assistance, and female-headed households. Energy burden is a geographic phenomenon and nonstationary across the U.S. and is more pronounced in rural than urban areas, as other studies have noted (Gouveia et al., 2019; Walker et al., 2012). Based on the results of our spatial regression modeling, we observe that both local and situational effects are present but local-specific (i.e., own-county) are more sizable than situation-specific (i.e., neighboring counties) effects, such that neighboring counties may have an impact on a county's energy burden. Thus, it is not enough to examine a county solely within itself, as the characteristics of nearby counties form a larger region that also has an effect on the energy burden it faces. This relationship also affirms other scholars' findings of the need for attention to how social processes are spatialized (Mashhoodi et al., 2019; Robinson et al., 2019) as well as more refined geographic data on utility providers, which often stretch across multiple counties, as regions needing greater consideration (Brown et al., 2020; Graff et al., 2021).

Utilities are an important site of intervention to change policies related to disconnections, customer debt, and late fees and can also help to ameliorate the factors that contribute to energy burden for LIHs through rate design, on-bill financing, and community solar programs among other interventions. At a more micro-geographic scale, additional analysis at census tract or block levels to account for neighborhood effects can also improve the understanding on where and how different neighborhoods within counties are most and least burdened. Investigating the ongoing interactions between relationships at a local, state, and national level policy will be necessary to devise policies and programs capable of ameliorating the geographically-specific concentrations of severe and localized energy burden.

Limitations of the study

There are several limitations to the present study that could inspire future research. First, this paper only focuses on the effects of household energy expenditure in electricity, fuel, and natural gas; however, future research can investigate other household burdens, such as expenses for rent, medicine, food, transportation, and since LIHs may be forced to make tradeoffs between energy expenditure and other necessities. Second, the datasets analyzed here did not have the variables related to psychological stress, which is an important research area relating to LIHs' energy poverty; therefore, qualitative research incorporating focus groups, interviews, or other methods may help researchers reach LIHs and underserved communities (e.g., elderly, people with disabilities) without internet access to conduct psychological or attitudinal related research. Third, this study only focuses on moderate and lower-income households at the county level. Future research should study other geographic scales to capture neighborhood, urban, and utility-scale processes. Finally, the datasets analyzed here did not have policy-related variables, and future research should investigate the impacts of local and national policies on energy poverty issues.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHOD DETAILS
  - Variables
  - Methodological justification
  - Data analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2022.104139.

ACKNOWLEDGMENTS

General: The authors thank Hannah Nelson for comments and Gregory Bonilla for data collection. C-F.C. thanks the support of 2019 Fulbright Global Scholar Award through US-UK Fulbright Commission. J.F. was supported by the Office of Research Innovation and Economic Development Summer Graduate Research

Assistantship Fund at the University of Tennessee. C-P.K. is supported by the Institute for a Secure & Sustainable Environment at the University of Tennessee.

AUTHOR CONTRIBUTIONS

C-F.C. conceived and designed this project. J.F. compiled most of the data used in the analysis, created the models, and led the result interpretation. C-P.K. estimated energy burden. C-F.C. led the overall manuscript writing and conducted OLS regression models. N.L. led literature review and conclusion writing and revised the entire manuscript. C-P.K. and J.S.F. contributed to writing literature review. J.S.F. edited the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in science. One or more of the authors of this paper received support from a program designed to increase minority representation in science. While citing references scientifically relevant for this work, we also actively worked to promote gender balance in our reference list.

Received: November 5, 2021
Revised: February 24, 2022
Accepted: March 18, 2022
Published: April 15, 2022

REFERENCES

Anselin, L. (1988). Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. Geogr. Anal. *20*, 1–17. https://doi.org/10.1111/j.1538-4632.1988.tb00159.x.

Anselin, L. (2003). Spatial econometrics. In A Companion to Theoretical Econometrics, B.H. Baltagi, ed. (Blackwell Publishing Ltd), pp. 310–330. https://doi.org/10.1002/9780470996249.ch15.

Baker, S., DeVar, S., and Prakash, S. (2019). The-Energy-Justice-Workbook-2019-web.pdf (Initiative for Energy Justice).

Bednar, D.J., and Reames, T.G. (2020). Recognition of and response to energy poverty in the United States. Nat. Energy *5*, 432–439. https://doi.org/10.1038/s41560-020-0582-0.

Bouzarovski, S. (2018). Energy Poverty: (Dis) Assembling Europe's Infrastructural Divide (Palgrave Macmillan).

Brown, M.A., Soni, A., Lapsa, M.V., Southworth, K., and Cox, M. (2020). High energy burden and low-income energy affordability: conclusions from a literature review. Prog. Energy *2*, 042003. https://doi.org/10.1088/2516-1083/abb954.

Carley, S., and Konisky, D.M. (2020). The justice and equity implications of the clean energy transition. Nat. Energy *5*, 569–577. https://doi.org/10.1038/s41560-020-0641-6.

Casabonne, U., Faria, M.M., Graiff Garcia, R., and Sacks, A. (2019). Energy Vulnerability in Female-Headed Households: Findings from the Listening to Citizens of Uzbekistan Survey (International Bank for Reconstruction and Development/World Bank).

Castán Broto, V., and Kirshner, J. (2020). Energy access is needed to maintain health during pandemics. Nat. Energy *5*, 419–421. https://doi.org/10.1038/s41560-020-0625-6.

CDC Social Vulnerability Index (2020). Agency for toxic substances and disease registry. https://www.atsdr.cdc.gov/placeandhealth/svi/index.html.

Chen, C., Zarazua de Rubens, G., Xu, X., and Li, J. (2020). Coronavirus comes home? Energy use, home energy management, and the social-psychological factors of COVID-19. Energy Res. Social Sci. *68*, 101688. https://doi.org/10.1016/j.erss.2020.101688.

Chen, C., Nelson, H., Xu, X., Bonilla, G., and Jones, N. (2021). Beyond technology adoption: examining home energy management systems, energy burdens and climate change perceptions during COVID-19 pandemic. Renew. Sustain. Energy Rev. *145*, 111066. https://doi.org/10.1016/j.rser.2021.111066.

COVID-19 United States Cases by County (2021) (Johns Hopkins University). https://coronavirus.jhu.edu/us-map.

Crenshaw, K. (1989). Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics (University of Chicago Legal Forum), pp. 139–167.

Drehobl, A., Ross, L., and Ayala, R. (2020). How High are Household Energy Burdens? An Assessment of National and Metropolitan Energy Burden across the United States (American Council for an Energy-Efficient Economy).

FACTS, U. (2021). Detailed Methodology and Sources: COVID-19 Data.

Franklin, M., Kurtz, C., Alksnis, M., Steichen, L., and Younger, C. (2017). Lights out in the Cold: Reforming Utility Shut-Off Policies as if Human Rights Matter (NAACP Environmental and Climate Justice Program).

Gouveia, J.P., Palma, P., and Simoes, S.G. (2019). Energy poverty vulnerability index: a multidimensional tool to identify hotspots for local action. Energy Rep. *5*, 187–201. https://doi.org/10.1016/j.egyr.2018.12.004.

Graff, M., and Carley, S. (2020). COVID-19 assistance needs to target energy insecurity. Nat. Energy *5*, 352–354. https://doi.org/10.1038/s41560-020-0620-y.

Graff, M., Carley, S., Konisky, D.M., and Memmott, T. (2021). Which households are energy insecure? An empirical analysis of race, housing conditions, and energy burdens in the United States. Energy Res. Social Sci. *79*, 102144. https://doi.org/10.1016/j.erss.2021.102144.

Hernández, D. (2013). Energy insecurity: a framework for understanding energy, the built environment, and health among vulnerable populations in the context of climate change. Am. J. Publ. Health *103*, e32–e34. https://doi.org/10.2105/AJPH.2012.301179.

Lewis, J., Hernández, D., and Geronimus, A.T. (2020). Energy efficiency as energy justice: addressing racial inequities through investments in people and places. Energy Effic. *13*, 419–432. https://doi.org/10.1007/s12053-019-09820-z.

Life Course Metrics Project (2013). Life Course Indicator: Concentrated Disadvantage.

Liévanos, R.S. (2019). Racialized structural vulnerability: neighborhood racial composition, concentrated disadvantage, and fine particulate matter in California. Int. J. Environ. Res. Publ. Health *16*, 3196. https://doi.org/10.3390/ijerph16173196.

Luke, N. (2021). Powering racial capitalism: electricity, rate-making, and the uneven energy geographies of Atlanta. Environ. Plann. E: Nat. Space. 251484862110167. https://doi.org/10.1177/25148486211016736.

Ma, O., Laymon, K., Day, M., Oliveira, R., Weers, J., and Virmont, A. (2019). Low-Income Energy Affordability Data (LEAD) Tool Methodology (National Renewable Energy Laboratory (NREL), Golden, CO. 2021). https://www.nrel.gov/docs/fy19osti/74249.pdf, Accessed 01 July 2021.

Manski, C.F. (1993). Identification of endogenous social effects: the reflection problem. Rev. Econ. Stud. *60*, 531–542. https://doi.org/10.2307/2298123.

Mashhoodi, B., Stead, D., and van Timmeren, A. (2019). Spatial homogeneity and heterogeneity of energy poverty: a neglected dimension. Ann. GIS *25*, 19–31. https://doi.org/10.1080/19475683.2018.1557253.

Massey, D.S., and Denton, N.A. (1993). American Apartheid: Segregation and the Making of the Underclass, Later Printing edition (Harvard University Press).

Mayer, A., and Smith, E.K. (2019). Exploring the link between energy security and subjective well-being: a study of 22 nations. Energy Sustain. Soc. *9*, 34. https://doi.org/10.1186/s13705-019-0216-1.

Memmott, T., Carley, S., Graff, M., and Konisky, D.M. (2021). Sociodemographic disparities in energy insecurity among low-income households before and during the COVID-19 pandemic. Nat. Energy *6*, 186–193. https://doi.org/10.1038/s41560-020-00763-9.

Mennis, J., Stahler, G., and Mason, M. (2016). Risky substance use environments and addiction: a new frontier for environmental justice research. IJERPH *13*, 607. https://doi.org/10.3390/ijerph13060607.

Morris, E., Greene, J., and Healey, V.M. (2019). Blueprint Guide for Creating Gender-Sensitive Energy Policies (No. NREL/TP-7A40-73927, 1544544) (Clean Energy Solutions Center). https://doi.org/10.2172/1544544.

Nguyen, C.P., and Su, T.D. (2021). Does energy poverty matter for gender inequality? Global evidence. Energy Sustain. Dev. *64*, 35–45. https://doi.org/10.1016/j.esd.2021.07.003.

Petrova, S., and Simcock, N. (2019). Gender and energy: domestic inequities reconsidered. Soc. Cult. Geogr. 1–19. https://doi.org/10.1080/14649365.2019.1645200.

Prevention, C.for D.C. (2020). Calculating percent positivity. https://www.cdc.gov/coronavirus/2019-ncov/lab/resources/calculating-percent-positivity.html.

Project, T.C.T. (2021). About the data. https://covidtracking.com/about-data.

Reames, T.G. (2016). Targeting energy justice: exploring spatial, racial/ethnic and socioeconomic disparities in urban residential heating energy efficiency. Energy Pol. *97*, 549–558. https://doi.org/10.1016/j.enpol.2016.07.048.

Robinson, C., Lindley, S., and Bouzarovski, S. (2019). The spatially varying components of vulnerability to energy poverty. Ann. Am. Assoc. Geogr. *109*, 1188–1207. https://doi.org/10.1080/24694452.2018.1562872.

Sampson, R.J., Sharkey, P., and Raudenbush, S.W. (2008). Durable effects of concentrated disadvantage on verbal ability among African-American children. Proc. Natl. Acad. Sci. *105*, 845–852. https://doi.org/10.1073/pnas.0710189104.

Surgo Ventures (2021). The U.S. COVID community vulnerability index (CCVI). https://precisionforcovid.org/ccvi.

Tobler, W.R. (1970). A computer movie simulating urban growth in the detroit region. Econ. Geogr. *46*, 234–240. https://doi.org/10.2307/143141.

United States Census Bureau. American Community Survey 2015-2019 5-Year Estimates [WWW Document]. URL https://data.census.gov/, 2021.

U.S. Census Bureau (2019). American Community Survey and Puerto Rico Community Survey 2019 Subject Definitions (United States Census Bureau), Accessed 30 January 2022.

U.S. Department of Energy Office of Energy Efficiency and Renewable Energy (2018). Low-Income Household Energy Burden Varies Among States — Efficiency Can Help in All of Them. https://www.energy.gov/sites/prod/files/2019/01/f58/WIP-Energy-Burden_final.pdf, Accessed 01 July 2021.

U.S. Department of Energy Office of Energy Efficiency and Renewable Energy (2021). LEAD tool. https://www.energy.gov/eere/slsc/maps/lead-tool.

U.S. Energy Information Administration (2020). Monthly Energy Review. https://www.eia.gov/totalenergy/data/monthly/. Accessed 01 July 2021.

US Coronavirus Cases & Deaths by State (2021). USAFACTS. https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/.

van Ginkel, J.R., Linting, M., Rippe, R.C.A., and van der Voort, A. (2020). Rebutting existing misconceptions about multiple imputation as a method for handling missing data. J. Personal. Assess. *102*, 297–308. https://doi.org/10.1080/00223891.2018.1530680.

Walker, R., McKenzie, P., Liddell, C., and Morris, C. (2012). Area-based targeting of fuel poverty in Northern Ireland: an evidenced-based approach. Appl. Geogr. *34*, 639–649. https://doi.org/10.1016/j.apgeog.2012.04.002.

Wang, J., and Johnson, D.E. (2018). An examination of discrepancies in multiple imputation procedures between SAS ® and SPSS ® an examination of discrepancies in multiple imputation procedures between SAS and SPSS. Am. Statistician *0*, 1–9. https://doi.org/10.1080/00031305.2018.1437078.

Wilson, V. (2021). Black Women Face a Persistent Pay Gap, Including in Essential Occupations during the Pandemic (Economic Policy Institute). https://www.epi.org/blog/black-women-face-a-persistent-pay-gap-including-in-essential-occupations-during-the-pandemic/?utm_source=Economic+Policy+Institute&utm_campaign=1abf423812-EMAIL_CAMPAIGN_2021_04_30_07_37_COPY_01&utm_medium=email&utm_term=0_e7c5826c50-1abf423812-59964657&mc_cid=1abf423812&mc_eid=9536d0b206.

Winter, A.S., and Sampson, R.J. (2017). From lead exposure in early childhood to adolescent health: a chicago birth cohort. Am. J. Publ. Health *107*, 1496–1501. https://doi.org/10.2105/AJPH.2017.303903.

CellPress
OPEN ACCESS

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Social and demographic data in the United States	U.S. Census Bureau	https://data.census.gov/
Low-income energy affordability data (LEAD) tool	Department of Energy	https://www.energy.gov/eere/slsc/low-income-energy-affordability-data-lead-tool
COVID-19 cases and deaths in the United States	Johns Hopkins University & USAFacts	https://coronavirus.jhu.edu/us-map https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/
COVID-19 community vulnerability Index	Centers of Disease and Control & Agency for Toxic Substances and Disease Registry	https://www.atsdr.cdc.gov/placeandhealth/svi/index.html
Software and algorithms		
ArcGIS Pro 2.7	Esri	https://www.esri.com/en-us/arcgis/products/arcgis-pro/
GeoDa 1.18.0	Luc Anselin, University of Chicago	https://geodacenter.github.io/
Python 3.8.5	Python Software Foundation	https://www.python.org/downloads/release/python-385/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Chien-fei Chen (cchen26@utk.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Data: This study analyzes existing, publicly available data. The key resources table documents five nationally representative data sources, including the COVID-19 Community Vulnerability Index (*CCVI)* data (“CDC Social Vulnerability Index,” 2020), the U.S. Department of Energy’s Low-Income Energy Affordability Data (LEAD) tool (U.S. Department of Energy Office of Energy Efficiency and Renewable Energy, 2021), energy expenditure from the U.S. Energy Information Administration (EIA) (U.S. Energy Information Administration, 2020), John Hopkins’ COVID-19 data(“COVID-19 United States Cases by County,” 2021), and the U.S. Census Bureau American Community Survey (ACS) (U.S. Census Bureau, 2021).
- Code: Original code was written in Python and is available from the lead contact upon request.
- All additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

METHOD DETAILS

Variables

Household income and energy burden (EB)

The household annual income data before taxes were generated from the American Community Survey (ACS) (U.S. Census Bureau, 2019). We used this income information and energy expenditure to establish the 2020 county-level energy burden for LIHs in the U.S. Due to the lack of official energy consumption data at the zip code and county level in 2020, historical energy expenditure (electricity, fuel, and natural gas) and burden estimations among LIHs were collected from the LEAD Tool by the National Renewable

Energy Laboratory (NREL) and the U.S. Department of Energy (DOE) (Ma et al., 2019), as well as the energy expenditure data from the EIA (U.S. Energy Information Administration, 2020). Stakeholders have used the LEAD tool to fill knowledge gaps, conduct strategies to improve energy program planning, and promote public awareness of LIH issues (Ma et al., 2019). Technically, the 2020 county-level energy burden data were estimated from the 2014-2018 county-level information of LIHs' annual income ($\text{Income}_{2014-2018\ ij}$) and 2020 state-level energy expenditure data by using the following equation ($\sum_1^k \text{Exp}_{2020\ ijk}$). Because 2020 county-level income data had not been reported until the end of the study period, the 2014-2018 income data were assumed to represent the 2020 income of each county.

$$\text{EBs}_{2020\ ij} = \frac{\sum_1^k \text{Exp}_{2020\ ijk}}{\text{Income}_{2014-2018\ ij}}$$
$$= \left(\sum_1^k \text{Exp}_{2014-2018\ ik} \times \frac{\text{Exp}_{2014-2018\ ijk}}{\overline{\text{Exp}_{2014-2018\ ik}}} \times \frac{\text{Con}_{2020\ ijk}}{\text{Con}_{2014-2018\ ijk}} \right) \times \frac{\text{EBs}_{2014-2018\ i}}{\sum_1^k \text{Exp}_{2014-2018\ ik}}$$

where $\text{EB}_{2020\ ij}$ was 2020 EB for the *i*th county in the *j*th month for LIHs, $\text{Exp}_{2020\ ijk}$ was the sum of energy expenditure for the *k*th source (electricity, fuel, and natural gas) of the *i*th county in the *j*th month in 2020, and $\text{Income}_{2014-2018\ ij}$ was the average income of *i*th county in the *j*th month during 2014-2018 for LIHs. $\text{Exp}_{2020\ ijk}$ was further calculated by using energy expenditure for the *k*th source of the *i*th county during 2014-2018 ($\text{Exp}_{2014-2018\ ik}$), energy expenditure fluctuation ratio in *j*th month for *k*th source compared with the monthly average ($\text{Exp}_{2014-2018\ ijk}/\overline{\text{Exp}_{2014-2018\ ik}}$), and 2020 state-level residential energy consumption ratio for *k*th source of *i*th county in *j*th month compared with 2014-2018 data ($\text{Con}_{2020\ ijk}/\text{Con}_{2014-2018\ ijk}$). The applied monthly energy consumption indices for electricity, fuel, and natural gas in each state were sales of electricity to the residential sector, prime supplier sales volume (propane), and natural gas consumption by the residential sector. $\text{Income}_{2014-2018\ ij}$ was calculated by using the EB for *i*th county for LIHs ($\text{EB}_{2014-2018\ i}$) and the sum of energy expenditure for the kth source of the ith county during 2014-2018 ($\sum_1^k \text{Exp}_{2014-2018\ ik}$). The missing values of our EBs data were replaced by the estimations from the multiple imputation techniques with the regression method (van Ginkel et al., 2020; Wang and Johnson, 2018), which employed county-level COVID-19 Community Vulnerability Index (*CCVI)* data. The U.S. Center for Disease Control and Prevention (CDC) built the CCVI, has complete datasets for all the U.S. counties, and reflects the socioeconomic and health vulnerability status of each county.

Concentrated disadvantage and Community Vulnerability Index (CCVI)

The concentrated disadvantage was measured by the proportion of households located in census tracts with a high level of socioeconomic disadvantage, calculated using ten census variables. The CCVI was used to measure concentrated disadvantage at the county-level to identify the factors influencing EBs, such as socioeconomic and health vulnerabilities that indicate communities that may be less resilient to the impacts of natural disasters and other extreme events, such as the COVID-19 pandemic (Surgo Ventures, 2021). The CCVI is a validated metric that uses census tract and county-level data include six themes: (1) socioeconomic status; (2) household composition and disability; (3) minority status and language; (4) housing type, transportation, household composition, and disability; (5) healthcare system factor; and (6) epidemiological factor. Each county is ranked from least vulnerable to most vulnerable in each of these categories. We extended five original variables in the concentrated disadvantage framework and used the subthemes of the CCVI to identify the impacts of specific variables. The variables from the CCVI include the percentage of minority groups, people aged 65 or older, people aged 18 or younger, people older than age 5 with a disability, and people living below 200% of the FPL. The present study also includes female-headed households, public assistance recipients, and households without health insurance from the ACS.

Households with female heads, public assistance recipients, and no health insurance

Since the CCVI does not capture the share of female-headed households or public assistance recipients in the past 12 months, we generated these variables from the ACS. Families with female heads are measured by the share of women householders with no spouse or partner present. Public assistance income, or welfare, provides cash payments to low-income families and includes general assistance and Temporary Assistance to Needy Families (TANF), which replaced Aid to Families with Dependent Children (AFDC) in 1997. Public assistance income also includes Supplemental Security Income (SSI) and non-cash benefits such as food stamps (U.S. Census Bureau, 2019). Additionally, the variable of unemployment in the CCVI was replaced by households without health insurance in the ACS to estimate residents' overall level of safe and secure employment.

COVID-19 confirmed cases and mortality

The information on COVID-19 confirmed cases and deaths were observed from January 21 to December 31, 2020, using the data from John Hopkins' University and USA FACTS ("US Coronavirus Cases & Deaths by State," 2021). These data sources use three main methods to collect this data: first, by drawing aggregate county-level data from the Covid Tracking Project, John Hopkins' utilizes data from 52 U.S. states and territories under CDC guidelines for test positivity (Prevention, 2020; Project, 2021). Second, USA FACTS indicates where presumed cases are included as COVID positive cases and adjusted per capita to represent the cumulative total. Lastly, USA FACTS estimates the gaps in daily cumulative cases and deaths by direct referencing or scraping from state and local agencies (FACTS, 2021). Both sources were used to validate confirmed cases and mortality to ensure accuracy.

Methodological justification

EB is geographic and varies across the U.S. Changes in the relative energy cost in a county may impact the relative energy cost decisions in neighboring counties, and these changes may happen at local (e.g., county-level) and global (e.g., country-level) scales. A Moran's I test of spatial autocorrelation on county-level energy burden in 2020 confirms this idea of spatial dependence. A Moran's Index value of 0.49, *z*-score of 45.37, and a *p*-value < 0.001, indicates a pattern of positive spatial autocorrelation: those similar situations of EB tend to be clustered together. We additionally used the Hot Spot Analysis tool in ArcGIS Pro 2.7 to identify statistically significant hot and cold spots of the areas experiencing either greater or lower EB. This tool calculates the Getis-Ord Gi* statistic, z-score, and p-value for each county to identify spatial clusters of high and low values of EB. Our local hot spot analysis finds that the United States is marked by clusters of counties with high energy burdens and similarly with low energy burdens. Thus, EB in the United States is characterized by spatial dependency and spatial heterogeneity. Spatial dependence refers to the tendency for things closer together to be more similar than things further apart (Tobler, 1970), while spatial heterogeneity refers to local variation across an area due to changes in the underlying properties of the landscape. Classical statistical analysis requires the assumption of independent and identical data but the nonrandom distribution of EB across the United States violates this requirement. A conventional ordinary least squares (OLS) model would likely result in coefficients that are biased and inefficient, and not able to capture this spatial process. We adopted a spatial regression model with spatially-lagged exogenous and endogenous regressors and errors based on concentrated disadvantage theory to predict EB, with additional tests for spatial dependence, and the results corroborate the notion of spatial structure in EB across the country. We discuss our analysis process to understand the relationship between concentrated disadvantage and geography with EB. In the next section, we explain our four models, including (1) a traditional OLS without control variables for the states and compare them to another model with control variables. Based on model diagnostics and tests for spatial dependence (2), we decided to include controls for states and introduce local effects into the equation with a spatially-lagged dependent variable. This paper compared a (3) spatial lag model with a spatial error model and ultimately concluded with a (4) spatial regression model with a spatially-autocorrelated error term and spatially-lagged independent and dependent variables that in addition to global effects, also included neighborhood effects of concentrated disadvantage variables.

Summary of regression models

Independent variables	(1a) OLS No state	(1b) OLS state	(2) Spatial lag	(3) Spatial error	(4) MSARSAR
Public assistance	−0.012	−0.012	−0.013*	−0.008	−0.009
No insurance	0.075***	0.067***	0.068***	0.077***	0.055***
Minority status	0.004	0.015***	0.015***	0.014***	0.013***
Female household heads	−0.017	−0.034***	−0.034***	−0.038***	−0.019
Disable	0.091***	0.135	0.115***	0.119***	0.114***
Below poverty	0.212***	0.186	0.170***	0.161***	0.149***
Age 65 or older	0.272***	0.291	0.274***	0.287***	0.241***
Age 17 or younger	−0.005	0.015	0.012	0.007	0.014

(*Continued on next page*)

Continued

Independent variables	(1a) OLS No state	(1b) OLS state	(2) Spatial lag	(3) Spatial error	(4) MSARSAR
COVID-19 cases	−0.020	−0.012	−0.015	0.002	−0.022
COVID-19 death	2.812***	0.679	0.699	0.741	0.621
w-public assistance	–	–	–	–	−0.046***
w-minority group	–	–	–	–	0.016***
w-age 17 or younger	–	–	–	–	0.041**
w-poverty	–	–	–	–	0.012
w-female household head	–	–	–	–	−0.110***
w-P2020 (energy burden)	–	–	0.181***	–	0.544***
Lambda	–	–	–	0.355	−0.099
F-*statistic*	(11, 3131) = 248.2084***	(61, 3081) = 62.6925***	(62, 3080) = 67.6821***	(61, 3081) = 58.4314***	(67, 3075) = 71.7442***
Adjusted R-Squared	0.4404	0.5410	0.5643	0.5377	0.5930
N	3142				

[1] The number in parentheses refers to the model described in the next section on Data Analysis; for example (3) Spatial Error refers to 3. Spatial Error Model. [2] *** Significant at the 0.1% level. ** Significant at the 1% level. * Significant at the 5% level.

Data analysis

OLS regression model and Lagrange Multiplier (LM) tests for spatial dependence

We first compared a model with and without the dummy variables. A traditional OLS model with these variables, excluding the dummies, is significant, $F(11,3131) = 248.208$, $p < .001$, with an adjusted R-squared of 0.440, Log likelihood of −7197.26, and Akaike info criterion (AIC) of 14,416.53. The OLS model that included the state dummy variables was also statistically significant, $F(61,3081) = 62.6925$, $p < .001$, with an adjusted R-squared value of 0.541, Log likelihood of −6860.82, and AIC of 13,843.642. Based on these model diagnostics, we confirmed that including dummy variables for the states improved model fit and estimation. Introducing states as control variables helped capture the spatial regime or structural differences in energy burdens across the U.S. Each state is likely to have its own distinct contribution of energy burdens owing to relative wages, incomes, and utility costs. Excluding the control variables for the states, with the exception of the percentage of public assistance (which is close to being significant, $p = 0.062$), all variables were significant at $p < 0.05$ (Table S2). The Moran's I test statistic on the residuals from our OLS model is positive, large (0.230), and statistically significant ($p < .001$). This result indicated strong spatial autocorrelation in the errors. Additional spatial dependence tests in linear models including the LM lag and error were, respectively, for a missing spatially lagged dependent variable and error dependence (Anselin, 1988). The Robust LM counterparts were the tests for each in the possible presence of the other (e.g., a Robust LM test for error dependence with the possibility that a spatially lagged dependent variable is missing). Meanwhile, the LM SARMA is a portmanteau test that assessed both LM Error and LM Lag together; and the test statistic was significant when either was also highly significant. Both LM tests were statistically significant at very high significance levels ($p < .001$), indicating the presence of spatial dependence in our model. However, the test statistic for the Robust error test (165.354) was larger than that of the test statistic for the Robust lag test (10.651), indicating a spatial error model would be better suited to capture the spatial dependence. Nonetheless, we used both an SLM and SEM and compare their results in the following sections.

Spatial lag model

In our models, geographic space was introduced by including variables that relate the values at a given location to those in surrounding and nearby locations. With LM Lag tests indicating that there was spatial dependence in our dependent variable of energy burdens, we fit a spatial lag model that includes this dependence in the result. Spatial lag describes how a variable in location i is influenced by variables in neighboring locations j. In other words, a spatially lagged dependent variable is one that averages the values of a location i's neighbors j and they can be used to account for autocorrelation in the model (Anselin, 2003). For example, we believe that energy burdens in a county are similar to that of its neighboring counties. While there are many ways to define the spatial relationship between each pair of

locations *i* and *j*, we adopted a queen criterion (i.e., all shared vertices and edges between a county *i* and county *j* are deemed neighbors). A spatial weights matrix *W* was constructed to capture this spatial relationship between all locations *i* and its neighbors *j* and row-standardized. Our spatial lag model included all the variables from the OLS model with control variables for states and the addition of the spatial lag of EB (i.e., the average energy burdens of neighboring counties around county *i*) to estimate EB. The spatial lag model was also significant, $F(62,3080) = 67.6821$, $p < .001$), with a Log likelihood of −6778.44 and AIC of 13,680.88, and has better predictive ability compared to the OLS model, accounting for 56.43% of the variation in average monthly energy burdens across the country in 2020. All independent variables' coefficients and their respective *p* values were similar to those in the OLS model with the additional inclusion of *PerSSI* and *PerFemHead* as statistically significant variables at $p < 0.05$ (Table S3). The estimated coefficient parameter (Rho) for the spatial lag of energy burdens (*W_P_2020*) was statistically significant and positive (0.181), and this result indicates that there are likely processes of spatial interaction that occur in the relative cost of electricity between counties. Including the spatial lag of our dependent variable has improved the model fit, relative to the traditional OLS model, and is reflected in both the Adjusted *R*-squared and Log likelihood value. However, the significance of the Breusch-Pagan test statistic ($p < 0.001$) indicates that there is heteroskedasticity in the model even when the spatial lag term is included, leading us to believe that there are still spatial effects unaccounted for in the model for further analysis.

Spatial error model

Whereas spatial dependence was incorporated into the dependent variable in a spatial lag model, it is instead incorporated in the errors/residuals in a spatial error model (SEM) (Anselin, 2003). This assumes that error terms are correlated across space; the error for one location affects its neighboring locations. An OLS model may underestimate the standard errors if there is a spatial process. We also use a SEM, $F(61,3081) = 58.431$, $p < .001$, on the same set of independent variables used in the original OLS model. Coefficient estimates and *p* values were somewhat similar to those from the spatial lag model, but with a worse adjusted *R*-squared value of 0.538 and AIC of 13,865.61. Lambda is the spatial error parameter and rather large (0.355), positive and significant, and these results told us that the unexplained variation in EB was spatially correlated (Table S4). However, the Breusch-Pagan (1026.120) test statistic was still significant at $p < .001$, meaning while we explicitly included spatial structure in the error terms, there remained spatial effects that were unaccounted for in our model, such as the neighborhood level effect of EB in our earlier spatial lag model. There may be omitted variables that contribute to EB and are common between neighboring locations. The spatial error model assumes that these omitted variables are not correlated with other independent variables.

Spatial regression model with a spatially-autocorrelated error term and spatially-lagged dependent and independent variables (MSARSAR – Modified SARSAR)

We adopt a model with spatially-lagged independent and dependent variables and spatially-correlated error terms similar to the foundational spatial interaction model by Manski (1993). In our model, the dependent variable of EB in each location *i* depends not only on the own-location-specific independent variables but also those same independent variables spatially weighted over its neighboring locations *j* with a spatially-correlated error term capturing unobserved characteristics. This model was statistically significant, $F(67,3075) = 71.744$, $p < .001$. With an adjusted R-squared value of 0.593, this model dramatically improves upon the predictive capacity of the other models. Model fit is improved as the AIC (13,474.60) was comparatively smaller than the other specified models. These test statistics led us to adopt this model as our final one to understand EB based on the concentrated disadvantage framework. The spatial error parameter, lambda, is −0.099 and indicates that the unexplained variation in EB is weakly and negatively correlated. All independent variables, excluding *PerSSI*, *PerFemHead*, *PercCase20*, and *PercDeat20* were statistically significant at $p < 0.05$. These findings describe the direct effects of our selected independent variables to understand EB. In contrast, indirect effects can be quantified by the change in the spatial lag of each independent variable. Most spatially lagged terms, excluding *PerPoverty*, are statistically significant at $p < 0.05$ (Table 3).