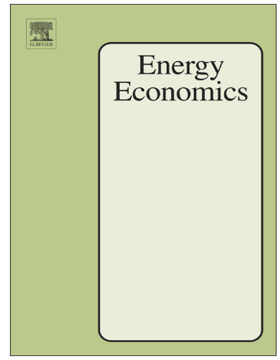
The definitive publisher version is available online at https://doi.org/10.1016/j.eneco.2020.104942

Journal Pre-proof



Prioritizing driving factors of household carbon emissions: An application of the LASSO model with survey data

Xunpeng Shi, Keying Wang, Tsun Se Cheong, Hongwu Zhang

PII: S0140-9883(20)30282-6

DOI: https://doi.org/10.1016/j.eneco.2020.104942

Reference: ENEECO 104942

To appear in: *Energy Economics*

Received date: 22 May 2020

Revised date: 19 July 2020

Accepted date: 27 August 2020

Please cite this article as: X. Shi, K. Wang, T.S. Cheong, et al., Prioritizing driving factors of household carbon emissions: An application of the LASSO model with survey data, *Energy Economics* (2020), https://doi.org/10.1016/j.eneco.2020.104942

Prioritizing driving factors of household carbon emissions: An application of the LASSO model with survey data

Xunpeng Shi [a,b], Keying Wang [b,c*], Tsun Se Cheong [d], Hongwu Zhang [e]

[a] Australia-China Relations Institute, University of Technology Sydney, Australia

[b] Center of Hubei Cooperative Innovation for Emissions Trading System, Hubei University of Economics, China

[c] School of Low Carbon Economics, Hubei University of Economics, China

[d] School of Business, Hang Seng University of Hong Kong, Hong Kong

[e] School of Economics, Zhongnan University of Economics and Law, China

Abstract

The identification of factors that influence household carbon emissions (HCEs)—a key driver of national emissions, is an important step in achieving more accurate predictions, as well as better interpretation and effective policy intervention. In this paper, based on survey data, we first calculated the direct, indirect, and total HCEs per capita for 37,620 households in China in the year of 2012, 2014 and 2016. Then we introduced a LASSO regression model to determine the main driving factors of HCEs and ranked the factors according to their importance. The use of the LASSO regression model addresses the issues of multicollinearity and over-fitting. It also provides two practical benefits: minimizing the number of influencing factors for forecasting and giving more flexibility in policy design. The results showed that fuel type and dwelling type can explain more than 70% of the direct HCEs, while income, urban or rural residency, and fuel type are the three most important influencing factors of the indirect HCEs. To mitigate HCEs while China will

continue its rapid urbanization and fast consumption growth, the government needs to provide affordable clean energy, improve the efficiency of household energy consumption, promote green and low-carbon economic recovery, and guide low-carbon lifestyles.

JEL Classification: Q56; D15

Keywords: total household carbon emissions; driving factors of HCEs; LASSO regression model

1 Introduction

The mitigation of climate change is one of the most critical global concerns. Under the framework of the Paris Agreement, most countries have made great efforts to reduce carbon emissions in recent years, however, the global CO_2 emissions have increased for the three consecutive years of 2017 to 2019, which means that the mismatch between the social demands for action on climate change and the actual rate of progress is increasingly serious (BP, 2020). The longer the carbon emissions continue to grow, the more difficult and costlier it will be to achieve the goal of zero-carbon. We have to rethink the reasons behind the increase (Druckman and Jackson, 2015). In recent years, more studies show that household carbon emissions (HCEs), either from direct household energy consumption or embodied in the goods and services, increase rapidly. Since HCEs are not regulated properly in many countries, they have been one of the main sources of the rebound of global carbon emissions, which have undermined the global emissions mitigation efforts.

Taking measures to control the rising trend of HCEs due to consumption growth is increasingly recognized as critical in the Chinese government's efforts for mitigating climate change (Liu et al., 2011). As the largest carbon emitter in the world, China's efforts to achieve its reduction target is also significant for global climate mitigation. A clearer understanding of the most important driving factors of HCES is critical for policymakers to identify what kind of abatement policies are more effective than others and thus to be prioritized.

In addition, China is launching a new round of stimulating consumption and infrastructure construction, which will have a profound impact on the households' consumption patterns, lifestyle, industrial structure, and further affect households' energy consumption and carbon emissions. Prioritizing the driving factors of HCEs and forecasting the HCEs in the future are key steps for achieving low-carbon development while increasing well-being.

At present, the decomposition analysis and regression analysis are the two widely used paradigms for investigating the driving factors of HCEs. There is a broad consensus that income is a key determinant of HCEs in the long run (Duarte et al., 2010; Lyons et al., 2012). Large-scale household surveys in several countries have generally found that household characteristics such as family size, dwelling type, the location of households, and the age, education level, marital status and occupation of the household's head, could be statistically significant predictors of HCEs (Baiocchi et al., 2010; Choi and Zhang, 2017; Golley and Meng, 2012; Long et al., 2018).

However, the decomposition methods face some difficulties in decomposing per capita HCEs based on survey data and including more influencing factors. While the increases in the total number of relevant variables in the regression model can greatly minimize the deviation of forecasting, but they will also lead to the problems of overfitting and multicollinearity. Moreover, a large number of statistically significant determinants may cause challenges in forecasting. Studies based on ad hoc hypotheses about the driving factors of HCEs cannot shed light on the relative importance of these factors and have limitations in terms of forecasting and policy development. Therefore, it is necessary to prioritize the key driving factors of HCEs in terms of the perspectives of HCEs prediction, decision-making and statistical methodology.

In this paper, by employing the emissions coefficient method and input-output model with survey data, we calculated direct HCEs, indirect HCEs and total HCEs per capita for 37,620 households in China in the year of 2012, 2014 and 2016. Then we used the least absolute shrinkage and selection operator (LASSO) regression model to determine the most important factors influencing the HCEs.

The results showed that the proportion of indirect HCEs was 84.8%, which means that researchers and policy-makers should pay more attention to the emissions embodied in the production of goods and services consumed by households. For both the direct HCEs and indirect HCEs, only three factors (out of a total of 25 factors) could explain more than 70% of the total decline in mean squared error (MSE). The family demographics of “coal as the main fuel”, “firewood/straw as the main fuel” and “living in bungalows” were the most important driving factors for the direct HCEs. For the indirect HCEs, the most important driving factors were “household income”, “urban or rural residency”, and “LNG/natural gas as the main fuel”. The characteristics of the household’s head were found to be not so important for both direct and indirect HCEs. This paper provides a discussion of these findings and accordingly proposes several important policy implications for mitigating HCEs.

Our major contribution to the literature is to introduce the LASSO regression model to investigate the driving factors influencing HCEs. LASSO provides an objective data-driven method for selecting the most important driving factors influencing HCEs. It is appropriate because economic theories currently support not all factors. Powerful predictors are often highly correlated in the prevailing econometrics models used in the literature. This means that the analysis cannot be performed properly due to multicollinearity. By contrast, the LASSO model can pinpoint the most important determinants, thereby enabling the researchers to address the issues of multicollinearity and over-fitting. Moreover, ordinary least squares (OLS) estimates are unbiased but have a large variance, whereas the LASSO estimates, by employing a continuous shrinkage method, sacrifice some bias to reduce the variance, thereby improving the overall accuracy of estimation (Tibshirani, 1996).

Using the LASSO can have two practical benefits. Firstly, in the formulation of carbon reduction policies, accuracy, simplicity, effectiveness and cost optimization must be taken into consideration. This requires that the aim of the research should not be restricted to the verification of whether each variable exerts an impact on HCEs. It should also identify the most influential factors (L. Wang et al., 2019). The LASSO model can set the regression coefficients of relatively unimportant factors to zero by imposing the L1 penalty, thereby minimizing the issue of too many variables in the policy-making process (Zhao and Yu, 2006). Secondly, the importance of the variables in terms of the change of parameters of the LASSO model can be ranked. This gives policy-makers more flexibility in determining policy interventions. In addition, the priority of identifying the driving factors minimizes the challenges of forecasting. For policy development, this is a key step in projecting future HCEs.

The paper is structured as follows. Section 2 briefly explains the driving factors of HCEs which have already been discussed in the literature. Section 3 explains the LASSO regression methodology and estimation of HCEs. Section 4 presents the data sources, descriptive statistics of the HCEs and household demographics. Section 5 provides the empirical results and discusses the key drivers of the HCEs. Section 6 concludes with a discussion of the policy implications.

2 Literature review

This paper is closely related to two research strands in the existing literature. The first strand is concerned with the driving factors of HCEs. The second strand is methodologies for investigating the driving factors of HCEs and the application of LASSO and some machine-learning methods in energy and climate science research.

With HCEs becoming an increasingly important issue in the field of environmental studies, more and more factors have been found to have an impact on HCEs. They can be divided into four categories. The first category is the geographical factors. The second one is household characteristics, such as income, family size, type of dwelling, type of fuel, and. The third one is the household head's demographics, including age, birth generation, educational level, gender, marital status, etc. The last one is the environmental awareness of family members.

Firstly, whether from the macro or micro level, urban or rural areas and geographical location are important factors affecting HCEs. The disparities between urban and rural areas, or among the regions, in terms of socio-economic development, government policies, habits have a definite effect on household consumption patterns and the related HCEs (Maraseni et al., 2016; Mi et al., 2020; Z. Wang et al., 2019). In the case of China, due to the existing urban-rural dual structure, the remarkable difference in income, dwelling, traffic, energy consumption and related CO2 emissions between urban and rural households has always been one of the research focuses (Feng et al., 2011).

Secondly, discussion about the role of household income exists in almost all the literature and it is widely accepted that income significantly promotes the growth of HCEs. In low-income countries especially, there is a disproportionate increase in both direct and indirect HCEs accompanying the expansion of income and consumption

(Meangbua et al., 2019). In addition, some studies have found that the direct HCEs of households with higher incomes are lower because they are more likely to use clean energy (Lyons et al., 2012). Moreover, empirical studies have found that there is an upwards and concave curve between income and HCEs. This indicates that the Environmental Kuznets Curve (EKC) hypothesis can be supported in the HCE level (Chancel, 2014; Grossman and Krueger, 2006; Zhang et al., 2017), including in developing countries, such as China, Philippines and Indonesia (Irfany and Klasen, 2017; Seriño and Klasen, 2015; Zhu et al., 2018).

Several studies have found that family size has a positive effect on HCEs, especially for direct HCEs, due to more rooms occupied, more children to raise, and higher residential heating expenditures (Jones and Kammen, 2014; Meier and Rehdanz, 2010). HCEs per capita, however, decrease with family size in the function of scale economy, and the tendency of small-sized families to place significant pressure on carbon reduction (Longhi, 2015; Qu et al., 2013; Underwood and Zahran, 2015). The type of dwelling is related to space heating, energy efficiency, and the use of household appliances, all of which influence direct and indirect HCEs (Motawa and Oladokun, 2015; Niamir et al., 2020). Household behavior in terms of cooking and heating is important. Undoubtedly, changes in the fossil fuel mix cause direct changes in HCEs (Papathanasopoulou, 2010).

Thirdly, as for the household heads' demographics, age is mentioned in many studies. Survey data has indicated that older people in the US, UK and China tend to generate more carbon emissions in some domains (Golley and Meng, 2012; Meier and Rehdanz, 2010; Murray and Mills, 2011). However, some studies find that the lifestyles of young people are more carbon-intensive than those of older people (Han et al., 2015). Be independent of age, the birth generation of a household's head has an effect on CO2 emissions (Chancel, 2014). Education is another factor considered to have an important impact on HCEs. However, the findings are not consistent. On the one hand, highly educated people are likely to be more eco-conscious (Dai et al., 2012). On the other hand, people with higher education levels tend to have higher incomes and to consume more, causing them to emit more HCEs (Büchs and Schnepf, 2013; Lee and Lee, 2014). Other demographics, such as gender and marital status, have also been found to influence HCEs (Büchs and Schnepf, 2013; Zhang et al., 2019).

Lastly, subjective variables such as awareness, motivation and social learning, have also been found to have significant impacts on HCEs (Li et al., 2019; Niamir et al., 2018). Decomposition analysis has found that the measurable driving factors account for less than half of the HCEs, while lifestyle change is the key driver of intertemporal HCEs and household energy consumption (Schipper, 1989; Zhang et al., 2020).

From the literature, it can be found that the developing countries and the developed countries in the early stages mainly focused on the effect of economic development and social structural changes on the HCEs, such as the increase of household income, the upgrading of consumption structure, urbanization and regional differences. With the growth of income and social development, especially the increasing of household surveys, the exploration of driving factors of HCEs is becoming more micro and in-depth and family characteristics and consumer behaviors attracted more attention.

As for the methodologies for investigating the driving factors of HCEs, two different paradigms, including decomposition analysis and regression analysis, have extensively explored this topic. Structural decomposition analysis (SDA) and index decomposition analysis (IDA) are two popular decomposition techniques to decompose emission aggregates and recently are applied in the driving factors of HCEs with household survey data (Shigetomi et al., 2018; Wei et al., 2020; Xu and Ang, 2014).

On the other hand, based on the theoretical analysis, many regression analyses made full use of the information in the household survey to reveal the influencing factors of HCEs, of which the multivariate ordinary least square (OLS) regression are widely used (Chancel, 2014; Chitnis et al., 2012; Wilson et al., 2013; Ye et al., 2013). Moreover, ridge regression is introduced to solve the multicollinearity problem in multivariate OLS estimation(Ma et al., 2019). In recent years, some studies applied quantile regression to analyze the differential effects of covariates along with the distribution of carbon emissions (Rong et al., 2018; Seriño, 2017; H. Zhang et al., 2019).

Recently, machine-learning models have provided new opportunities for innovative research in energy and climate science due to their superior performance in processing, classifying, predicting and policy analysis with complex large-scale data (Ghoddusi et al., 2019; Lakshmanan et al., 2015). Multiple machine learning methods have been applied to energy consumption at the household level. Using machine-learning methods, some studies have estimated residential gas and electricity based on household characteristics such as dwelling types and household tenure (Viegas et al., 2015; Wei et al., 2015; Zhang et al., 2018). Other studies have conducted economic assessments of photovoltaic battery systems based on household load profiles (Schopfer et al., 2018). Some research has revealed household characteristics from smart meter data and provided personalized and scalable energy efficiency programs for households (Beckel et al., 2014; Humeau et al., 2013). Knittel and Stolper (2019) evaluated the heterogeneous treatment effects of repeated behavioral nudges towards household energy conservation and found that pre-treatment consumption and home value are the strongest predictors of treatment effect.

As for the decomposition analysis, the total number of households in the survey data is inconsistent with the national or sub-national statistics database. Therefore, there may be bias in the estimated number of households based on survey data and population census data. Moreover, the driving factors based on decomposition analysis are usually summarized as the changes of scale, intensity, etc. while the influence of household demographic characteristics, especially the features of household heads, is seldom discussed. However, the regression analysis can introduce as many variables as possible into the model to empirically analyze their impact on the HCEs, which expands the depth in this topic and provides more perspectives for policymakers.

The LASSO methods combining the advantages of machine learning and regression have shown great potential in the prediction and selection of variables, and in policy analysis in the field of energy economy research. The empirical results have indicated that the LASSO provides significant improvements in the forecasting accuracy of oil prices. The selection efficiency of variables under LASSO is also found to be higher than the stepwise method (Miao et al., 2017; Zhang et al., 2019). When the hierarchical group-LASSO regularization was applied to a comprehensive dataset of energy consumption for commercial office and multifamily buildings in New York City, it was found that the method has high prediction accuracy and tends to be useful in modelling the energy consumption of other sectors (Hsu, 2015).

In summary, more and more factors have been included in the empirical models recently developed in the literature. These factors have been verified as having statistically significant impacts on HCEs, and this is conducive to a more in-depth and comprehensive understanding of the causes of the increases in HCEs. However, having too many driving factors hinders the process of policy development and reduces the efficiency of policy implementation. It can also lead to problems of overfitting and multicollinearity. With the LASSO model, we overcome the shortcomings of OLS regression. More importantly, we can identify the most influential factors influencing HCEs and rank their importance, thereby giving policy-makers more flexibility.

3 Methodology

A detailed discussion of the LASSO regression method and the estimation of HCEs is given below in order to make an explanation of the findings easier in the later sections.

3.1 LASSO regression model

One problem with the traditional regression method is that too many independent variables can be selected for running the regression. The number of independent variables is always greater than what is actually required in the regression model. Therefore, most researchers select the independent variables ad hoc by conducting a literature review and making their selection based on economic theories or through the modification of existing specifications employed by other studies. However, this selection process is based mainly on qualitative research that is largely dependent on previous research history, and which is arbitrary to some extent, especially when economic theories do not fully support the variables. Such an ad hoc selection of variables, however, often leads to ambiguous results (Sala-i-Martin, 1997).

Another critical but often neglected issue in constructing a model is sparsity. A sparse model is one in which only a small number of independent variables are involved but all of them play an important role in prediction. The best model is one with only a few variables, but which can provide a forecast with minimal error. Therefore, a procedure is needed through which the most important variables for a regression model can be selected.

The Least Absolute Shrinkage and Selection Operator (LASSO) model is specially designed to prioritize the importance of independent variables. Like other model selection technics, LASSO lets the data identify an optimal model. See for a discussion on the advantage of model selection technics (Tong Zhang et al., 2019). Additionally, LASSO can be used to derive a sparse model that can provide a good forecast with a minimal number of independent variables (Tibshirani, 2011).

For the LASSO model, the objective function for finding the minimum is different from the traditional regression approach and is shown below:

$$minimize_{\beta_0,\beta}\ (\frac{1}{2N}\sum_{i=1}^{N}(y_i-\beta_0-x_i^T\beta)^2+\lambda\sum_{j=1}^{p}|\beta_j|) \quad (1)$$

where N is the total number of observations, λ is a nonnegative regularization parameter corresponding to one value of Lambda, y_i is the dependent variable, p is the number of independent variables $x_i=(x_{i1},\dots,x_{ip})^T$, β_0 is the intercept, and β_j are the other parameters.

It can be observed that the lower the value of λ, the higher the number of non-zero β, and vice versa. By adjusting the value of λ, one can derive a model with a specified level of sparsity with its corresponding value of λ.

If the value of λ is exceptionally large, then all the parameters of the independent variables would have a value of zero. By reducing the value of λ gradually, one can observe that some parameters would turn from a value of zero to non-zero. Therefore, by a continual adjustment of the value of λ, the parameters will turn from zero to non-zero one by one. Therefore, based on the sequence of appearance of the parameters, one can know which independent variable is the most important for making the prediction. The accuracy of the model can be measured by MSE which can be computed as:

$$MSE = \frac{1}{N}\sum_{j=1}^{N}(y_i - \hat{y}_i)^2 \quad (2)$$

where y_i is actual the value of the dependent variable and $\hat{y}_i$ is the predicted value generated by the model.

It is notable that for the traditional regression models, the calculation of the parameter is conducted for one time only, however, multiple iterations were run for the LASSO method which is used in this study. By running multiple iterations based on different values of λ, one can observe the changes in the significance of the independent variables, thereby offering crucial information on the importance of these variables. Moreover, by studying whether the independent variables remained to be significant across different models of different iterations, one can gain an understanding of the robustness of the models as well. These advantages of the LASSO model can complement existing studies based on traditional regression models and provide important information on HCEs.

Knowledge about the importance of each of the independent variables is crucial for policy-makers as this information can help them establish a priority list when resources are scarce. This can help them determine the order in which each issue should be handled and make resources the most important contributing factor.

In this study, a procedure was developed so that different values of λ were employed in the analysis. One hundred sets of regressions were run with different values of λ. Each of one provided a different specification of independent variables and their corresponding degree of freedom (df) and MSE.

In the process of computation, different values of λ were calculated. Then these values were fed into the equation to derive the different specifications with their own non-zero components of β. The largest λ was computed in a way such that it gave a non-null model with at least one non-zero component of β, while the smallest λ was determined by dividing the largest value of λ by 10 000. The 100 specifications, denoted as *sp*, were constructed within the range of the highest and smallest values of λ. Specification 1 denoted the specification with the smallest value of λ, while specification 100 denoted the specification with the largest value of λ. The values of λ of the other specifications within the range were computed according to a geometric sequence.

Therefore, when the coefficient of an independent variable changes from zero to non-zero, the larger the corresponding λ or SP value means that the variable is more important for prediction.

3.2 Calculation of HCEs

The total carbon emissions for an individual household k in our sample consisted of two parts: indirect emissions $E_{indirect\_k}$, and direct emissions E_{direct_k}. The direct carbon emissions for household k were the emissions from the household's final consumption of fossil fuels, i.e., coal, oil and gas (Zhang et al., 2015). We calculated the direct carbon emissions for household *k* with the emissions coefficient method (ECM) from the IPCC (2006), as has been widely done in the literature (Munksgaard et al., 2000; Qu et al., 2013; Wiedenhofer et al., 2017). The direct CO_2 emissions for household k are:

$$E_{direct\_k} = \sum_i f_i Energy_{ik} \tag{3}$$

where f_i is the CO_2 emissions factor of energy source i, and $Energy_{ik}$ is the quantities of energy source i consumed by household k.

The indirect HCEs were estimated with Input-Output modeling (IOM). This has been a widely employed approach (Büchs and Schnepf, 2013; Dai et al., 2012; Ding et al., 2017; Fan et al., 2012; Golley and Meng, 2012; Mi et al., 2020; Wiedenhofer et al., 2017). The calculation of the indirect CO_2 emissions for a specific household k was as follows:

$$E_{indirect\_k} = D(I - A)^{-1} Exp_k \tag{4}$$

where D is the row vector of direct emission intensities for each sector, $(I - A)^{-1}$ was the Leontief inverse matrix, which is the key to the application of input-output analysis; and Exp_k was the column vector of household k's expenditure per capita on goods and services.

4 Data

Our datasets were comprised of three main components: (i) Samples of China households for the latest available years (2012, 2014 and 2016) from the *China Family Panel Studies* (CFPS); (ii) Input-Output table of China from *the World Input-Output Database* (WIOD); and (iii) Carbon emissions for China's 35 sectors in 2007 from the WIOD.

4.1 Household expenditure and demographics

The CFPS is a national representative longitudinal survey that was launched in 2010 and implemented every two years thereafter. The sample in the CFPS covers almost 15 thousand households for one year across all 31 provinces of China, and has been widely used in studies examining economic activities and household behavior (Xie and Hu, 2014), including energy and emissions issues (Shi et al., 2020; Hongwu Zhang et al., 2019; Zhang et al., 2020). We obtained information about the consumption expenditures of each household from the CFPS dataset. This is of critical importance when estimating direct and indirect carbon emissions.

In addition, we were able to convert the data from the CFPS on the three kinds of related expenditures of cooking, heating and personal driving available into direct energy consumption in the form of physical units with the provincial prices for each type of fuel that year, and then calculate the direct HCEs according to equation (3).

Using the LASSO model to explore the most influential indicators of HCEs, and we included the household characteristics that have been studied theoretically or empirically from CFPS, which includes socio-economic factors, household demographics and the features of households' heads. Specifically, besides dummy of urban or rural and regions reflecting the difference in households' lifestyles and consumption patterns, we collected household disposable income, family size, birth generation, marital status, and education level of household heads. We also collected the type of fuel to feature the household's direct energy consumption structure[1]. Because the type of housing has an important impact on households' energy consumption and carbon emissions, we included the variable into the model.

4.2 Input-Output table, and sectorial CO2 emissions intensity

The WIOD (the first version was released in 2013) provides China Input-Output Tables and total CO_2 emissions for 35 sectors.[2] Based on the two datasets, we derived the Leontief inverse matrix induced from the I-O table and per-Yuan CO_2 emissions intensities coefficients for each sector.[3] According to the method of data-matching applied by Zhang et al. (2020), we further aggregated the consumption-side detailed household expenditure items in the CFPS into a production-side Leontief inverse matrix and CO_2 emissions intensities, then estimated the indirect CO_2 emissions for every surveyed household according to equation (4). Combining the related information, we obtained a set of consolidated datasets providing a single record to show Chinese direct HCEs, indirect HCEs, and total HCEs, income, and other demographic characteristics for each household. Finally, to reduce bias caused by outliers, 1% of households with the highest and lowest emissions and income were excluded. The final sample size was 37,620.

---

[1] Firewood/straw is a form of renewable energy in the long-run, and in China's rural areas it is regarded as a form of non-commercial energy. According to the survey results relating to energy consumption and end-use activities for China's rural households (Wu et al., 2017), it was reasonable to assume that the energy expenditure of households using firewood/straw as their main fuel might also purchase some coal. Thus, the calculation of direct carbon emissions from these households was the same as that for households with coal as the main fuel. However, the households with firewood/straw as the main fuel tended to be different from those using coal as the main fuel. Hence, we kept the two types in the model.

[2] An important reason that we applied the I-O table and sectoral CO2 of the WIOD was because the same sectoral classification in two dataset and sectoral carbon emissions provided directly by the WIOD greatly reduces the bias arising from the processing of data-matching and the bias due to inconsistency in the selection of energy emission factors when calculating the sectoral carbon emissions with energy use data.

[3] In RMB according to the exchange rate in 2007.

4.3 Descriptive statistics

Table 1 presents the key summary statistics. The total HCEs per capita of Chinese households in three years was 2.805 tons,[4] of which the direct HCEs per capita was 0.426 tons, accounting for 15.2%, and the indirect HCEs per capita was 2.379 tons, accounting for 84.8%. In contrast, the carbon emissions resulting from the direct energy consumption of households accounted for only a small share of the total. Most of the HCEs were emitted in the process of producing the goods and services consumed by households. Although indirect HCEs are not from the production sectors, but are generated in the processes to produce the needs of households, households should be responsible for them.

Table 1 shows that the proportion of urban households was 52.8%. The proportions of households from the east, west, central, and northeast regions were 36.2%, 25.8%, 25.5%, and 12.6%, respectively. In terms of fuel types used by the households, 43.5% and 23.5% of households used LNG/natural gas and electricity as their main fuel. The proportion of urban households using solar/biogas or coal as their main fuel was very low; 61.1% of them used LNG/natural gas as their main fuel. Of the rural households, 41.9% still used firewood/straw as their main cooking and heating fuel. This was quite high considering the increasing popularity of electricity and LNG in rural areas. According to the dwelling types, 39.5% of urban households lived in apartments, while 25.5% and 21.2% lived in low-rise buildings and bungalows,

[4] The values of per capita household CO_2 emissions in China are not always consistent in the literature due to differences in the samples, time or method: 2.3 tons per capita for urban households in 2005 (Golley and Meng, 2012), 1.43 tons per capita for peasants and herdsmen in northwestern arid-alpine regions of China in 2008 and 2009 (Qu et al., 2013), 1.77 tons per capita in 2011 (Maraseni et al., 2015), 1.72 per capita for China in 2012 (Wiedenhofer et al., 2017), and 1.6 tons in 2007 and 2.0 tons in 2012 (Mi et al., 2020).

respectively. In the rural areas, 53.6% of households lived in bungalows, and 26.0% lived in low-rise buildings. For the entire sample, the proportions of households living in apartments, low-rise buildings, and bungalows were respectively 22.5%, 23.5% and 38.7%.

As for the characteristics of the household heads, 87% were married. The proportions born in the 1950s, 1960s and 1970s all exceeded 20%. The proportions of those born in the 1940s and 1980s exceeded 10%, and those born in the 1910s to 1930s and in the 1990s were less than 5%. There were some differences in the generation of urban households and rural ones. In the urban areas, the proportion of household heads born after 1970 was 6.9% higher than that in the rural areas. This reflects the fact that the degree of aging in rural areas was more serious. In terms of education level, 59.8% of household heads had junior secondary and above education level, while 9.1% of them had tertiary education. The proportion of urban household heads with senior secondary and above education was higher than in the rural areas at 22.9%.

Table 1 Descriptive Statistics

Variable	Mean	Std.Dev.	Variable	Mean	Std.Dev.
Direct HCEs per capita (tons)	0.426	0.755	Dwelling type (share)		
Indirect HCEs per capita (tons)	2.379	2.015	Apartment	0.225	0.418
Total HCEs per capita (tons)	2.805	2.237	Low-rise building	0.235	0.424
Household income per capita (10^4 Yuan in RMB)	1.215	1.248	Bungalow	0.387	0.487
			Others	0.152	0.359
Household size	3.711	1.772	Generation of Head (share)		
Urban (share)			Generation_10S-30S	0.049	0.215

Region (share)	0.528	0.499	Generation_40S	0.119	0.323
East			Generation_50S	0.208	0.406
West	0.362	0.48	Generation_60S	0.261	0.439
Center	0.258	0.437	Generation_70S	0.208	0.406
Northeast	0.255	0.436	Generation_80S	0.116	0.321
Fuel type (share)	0.126	0.332	Generation_90S	0.040	0.195
Solar/Biogas	0.009	0.094	Education of Head (share)		
Electricity	0.236	0.424	Under primary	0.173	0.379
LNG/Natural gas	0.435	0.496	Primary	0.229	0.420
Coal	0.048	0.213	Junior secondary	0.346	0.476
Firewood/Straw	0.257	0.437	Senior secondary	0.161	0.367
Others	0.015	0.123	College and above	0.091	0.287
Marital Status of Head(share)	0.849	0.358			
Observations			37620		

5 LASSO Regression Results

The LASSO regression model needs three steps to analyze the impact of family demographics on HCEs: (1) filtering the family demographics that were most important to HCEs; (2) sorting the importance of demographics; and (3) determining a set of regression coefficients.

5.1 Filtering the most important family demographics to HCEs

The parameter MSE is a reliable judgment standard for variables selection through the LASSO regression model. A smaller value of MSE means a smaller deviation between the estimated value and the actual value, which also means higher goodness of fit of the model. This standard can be regarded as AIC (Akaike Information Criterion), which requires the most accurate model with the smallest degree of freedom (df), namely the minimum number of independent variables.

According to equation (3), we took the direct HCEs per capita, indirect HCEs per capita, and total HCEs per capita as dependent variables, respectively, and applied the LASSO regression model to select the most important influencing variables on HCEs. The trends in the number of independent variables and also in the contribution to the total decline of the MSE with the decrease of specification (horizontal axis) are shown in Figure 1. In the three models, more coefficients of independent variables changed to non-zero with the decrease of *SP.* At the same time, the contribution to the total decline of the MSE increased sharply until the number of variables in the models rose to seven, eight and seven, respectively, after which the curves levelled off.

For direct HCEs, the first explanatory variable lead to a decline of the MSE accounting for 7.9% of the total decline. This can be considered as the contribution to the total decline of the MSE. When the numbers of variables rose to seven, eight and seven for direct HCEs, indirect HCEs and total HCEs, the contributions were 87.1%, 95.8% and 89.0%, respectively. More specifically, when the numbers of variables rose to three, the contributions reached or exceeded 70% for all three models. This means that only a few of all the 25 household demographics were enough to explain the changes in HCEs per capita. It is worthwhile to focus on the most influential variables by taking the sparsity of the regression model into consideration. In that sense, our empirical results showed that policies targeting the three most influential variables were most effective in reducing HCEs.

Contribution to the total decline of MSE (left Y axis) Number of variables with nonzero coefficient (right Y axis)		
Direct HCEs	Indirect HCEs	Total HCEs

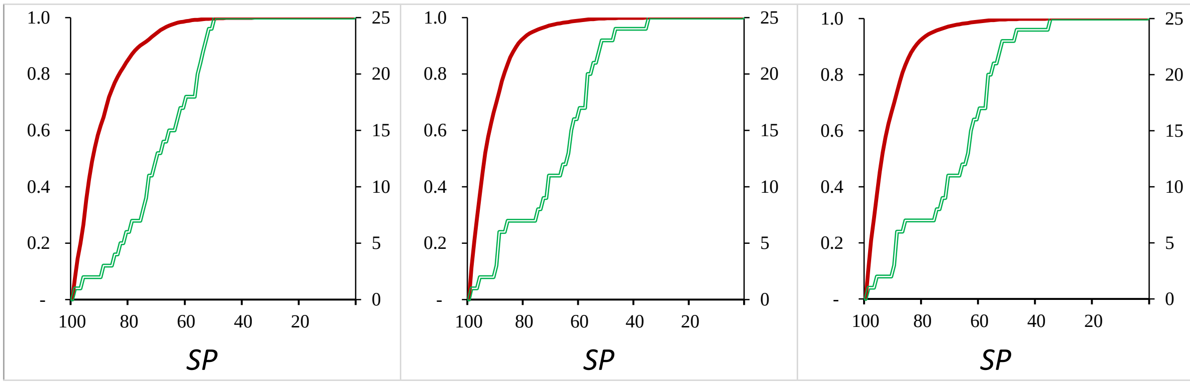


Figure 1 The change of parameters with the decrease of *SP*

5.2 Sorting the importance of demographics

In the LASSO regression model, *SP* was the constraint strength and its range was [0,100]. When the value of *SP* decreased from 100 to 0, the constraint strength gradually decreased. This means that the sparse degree of the optimal solution matrix of the objective function also gradually decreased. Therefore, if the process of value of *SP* decreased from 100 to 0, some coefficients of variables started to change from zero to non-zero. The variable with the first non-zero coefficient had the greatest impact on the HCEs and the corresponding *SP* value was also the biggest.

We used bar charts in Figure 2-4 to demonstrate this more clearly. The importance of each influencing factor to HCEs is seen more intuitively. Figure 2 shows that when the value of *SP* was 99, the coefficients of the variable of "coal as the main fuel" first became non-zero, meaning that the coal variable had the largest impact on direct HCEs. The next variable was "firewood/straw as the main fuel" and the value of *SP* was 96. When the values of *SP* decreased between 80~90, the coefficients of the variables of "family size", "living in low-rise buildings", "living in northeast region" and income became non-zero. When the values of *SP* decreased to 79, the coefficients of the variables of "urban or rural residency" became non-zero.

So far, when the value of *SP* decreased to 79, the coefficients of all seven of the most important variables became non-zero.

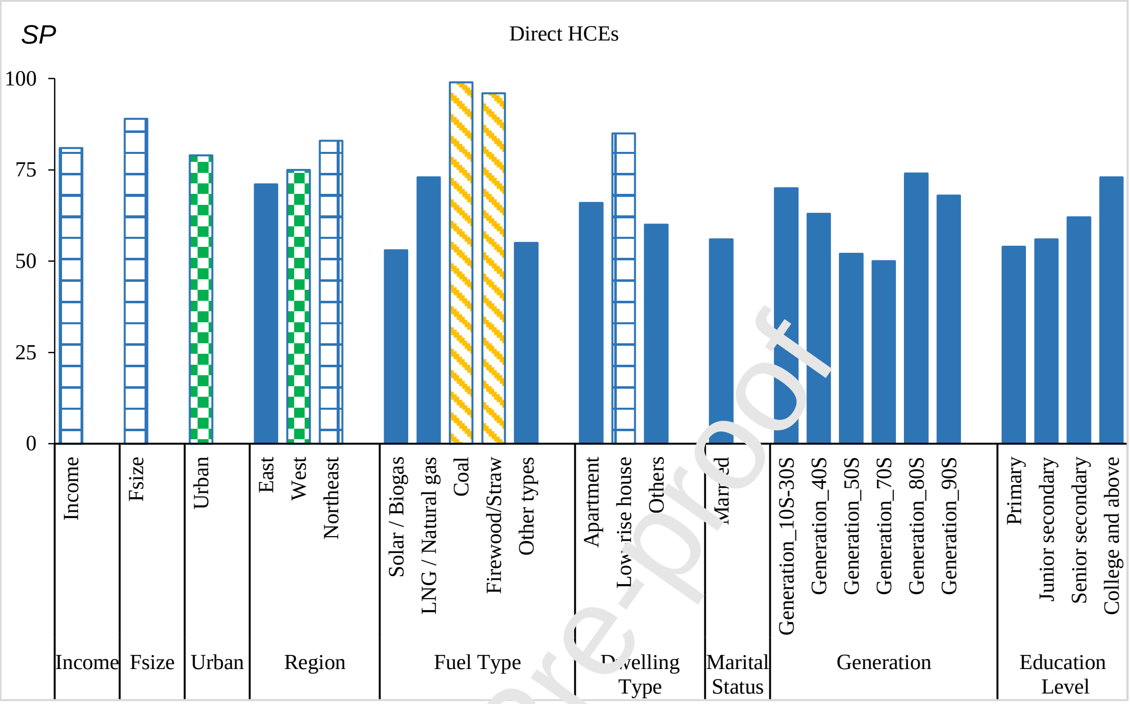


Figure 2 Household demographics of direct HCEs and the corresponding *SP*

As shown in Figure 3, the variable of "household income per capita" was the most important influencing factor for indirect HCEs and the corresponding *SP* was 99. Then the variables of "urban or rural residency" and "the LNG/Natural gas as the main fuel" were the second important driving factors of the indirect HCEs. The importance of family size, "firewood/straw as the main fuel" and "living in apartments" was the same and the corresponding *SPs* were all 89. The features of households' heads of "with education of college and above" and "born in the 1990s or later", had a relatively significant effect on the indirect HCEs, which was different from the direct HCEs.

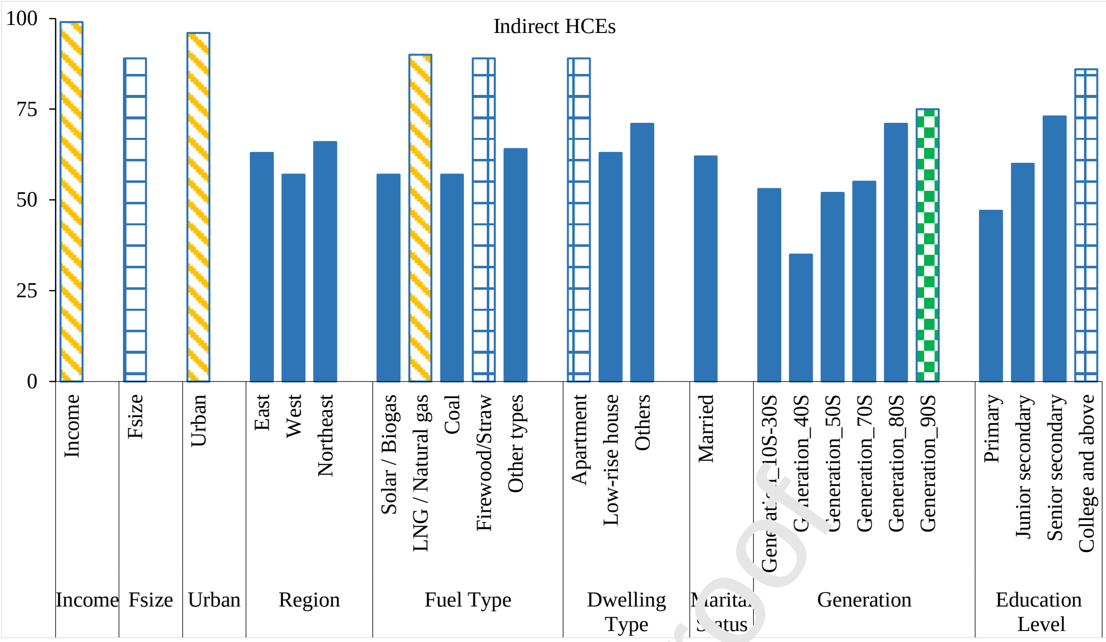


Figure 3 Household demographics of indirect HCEs and the corresponding *SP*

As for total HCEs, Figure 4 shows that regional differences were not an important factor; nor were the marital status and birth generation of the household's head. In addition, when the value of *SP* decreased from 100 to 80, the most influential variable was household income per capita, followed by factors of "urban or rural residency", "family size", "the household's head with the education of college and above", "living in apartments", "LNG/natural gas as the main fuel", and "firewood/straw as the main fuel".

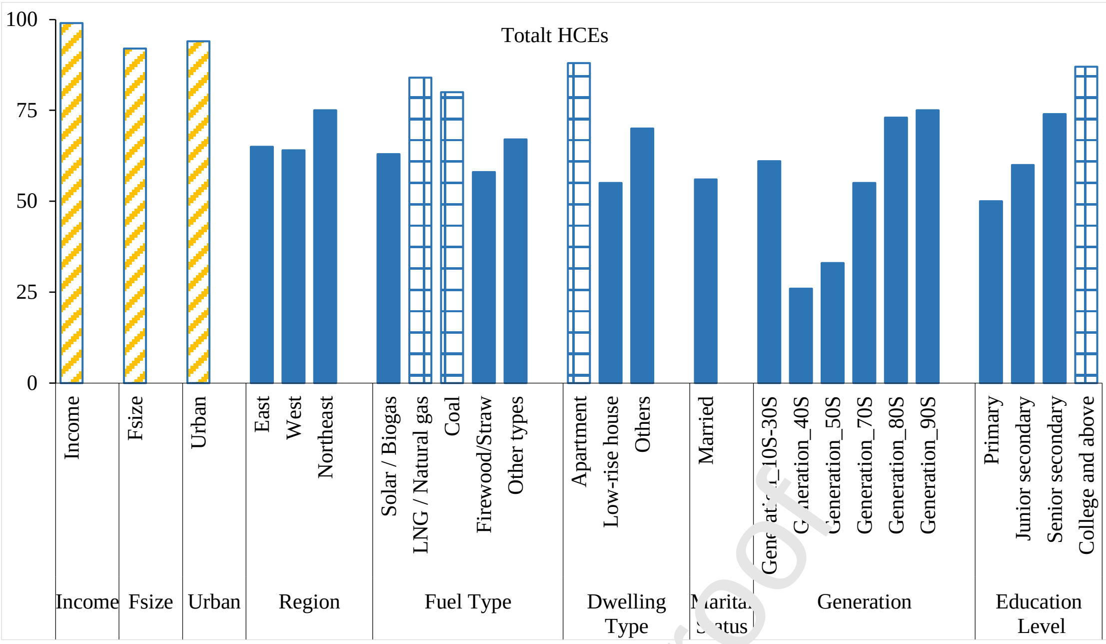


Figure 4 Household demographics of total HCEs and the corresponding *SP*

5.3 Determining a set of regression coefficients

From the above analysis, we can see that the method of observing the change in the MSE enabled us to identify the most important household driving factors of HCEs, while the *SP* value corresponding to the demographics enabled us to further rank the importance of them. However, the LASSO regression model generated a series of coefficients under each *SP* value. Figure 5 describes the trend of the coefficients for the main variables with the decrease in *SP*. We can see these coefficients changed greatly with the decrease in *SP* when the value of *SP* was high and then gradually became stable. It can be seen from Figure 2 and Figure 5 that all three MSEs were stable when *SP* decreased to 79/75/80, but not all coefficients of the important variables were stable before decreases to 50, 40 and 40, for the direct HCEs, indirect HCEs, and total HCEs, respectively. Hence, the set of coefficients under a stable situation can be used to analyze the impact of household demographics on HCEs. In this way, the third purpose of applying a LASSO regression is to obtain more accurate estimated coefficients.

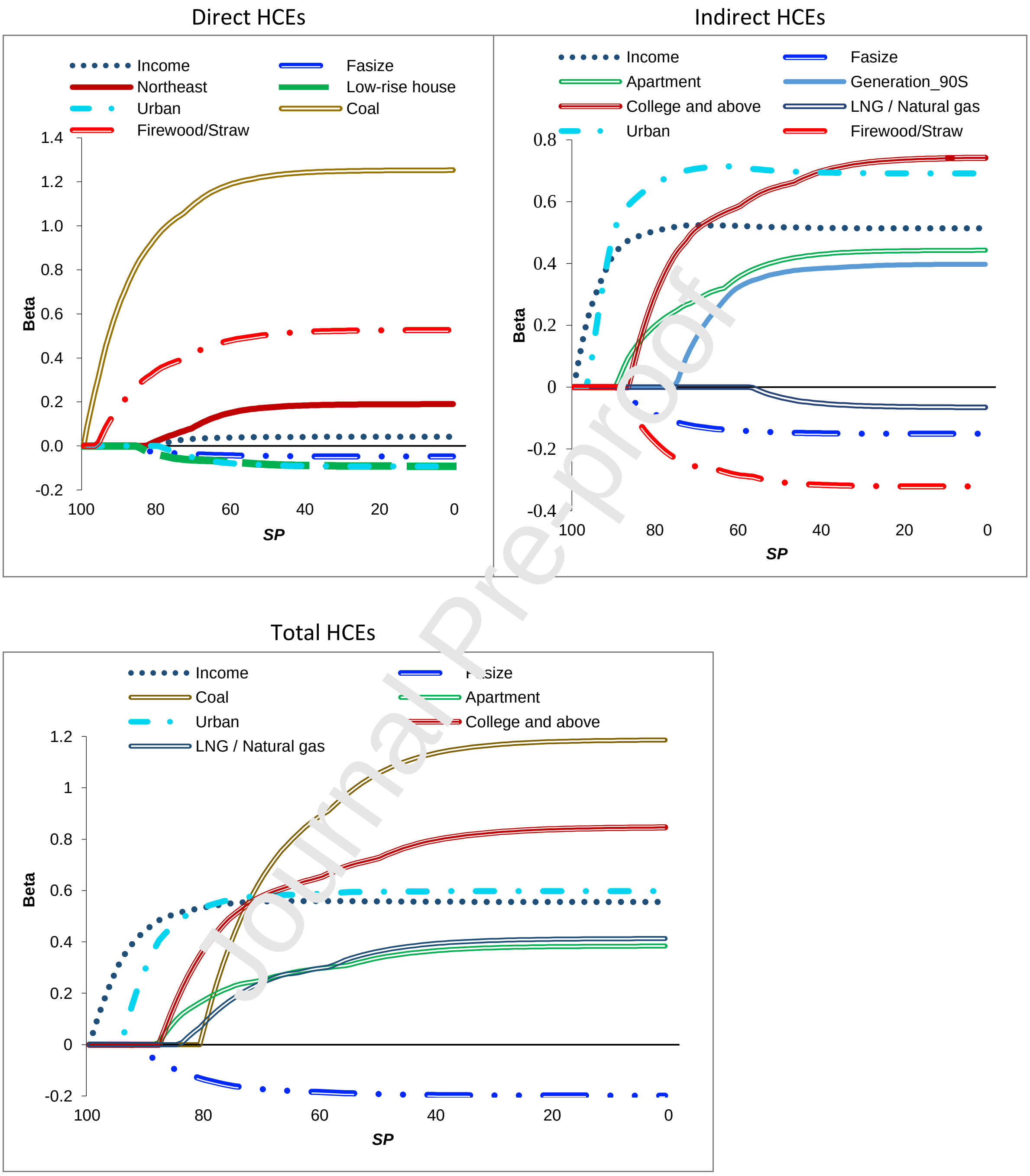


Figure 5 The trend of coefficients with the decrease of *SP*

In Table 2, we list the coefficients of household demographics in three cases in which the MSEs were stable (the values of *SP* were 79, 75 and 80, respectively), all the important coefficients are stable (the values of *SP* were 50, 40 and 40 respectively) and all the values of *SP* were 1. It can be seen that the declines in all MSEs were still small. Even the *SP* decreased to 0, while the absolute value of coefficients increased significantly with *SP* decreasing to 50 or 40, and then remaining stable until *SP* decreased to 0.

Generally speaking, the characteristics of the household heads, such as marital status, birth generation and education level, had no significant impact on the direct HCEs. Moreover, it can be seen from Table 2 that compared to households which used electricity as the main fuel, households which used coal or firewood to heat or cook had considerably larger CO_2 emissions. Compared to households living in low-rise buildings, those living in bungalows emitted more direct HCEs because bungalows are less effective in retaining heat and saving energy. The direct HCEs per capita of rural households was 0.56 tons, while it was 0.37 tons for urban ones. This suggests that rural energy use is less efficient than urban ones. In the Northeast regions of China, the heating season is longer than in other regions due to the longer winters and lower temperatures. Income undoubtedly leads to the expansion of consumption, while the role of family size in reducing carbon emissions per capita reflects the intensive effect of scale.

As for indirect HCEs, Figure 4 and Table 2 show that regional differences were not an important factor. It is no surprise that the higher the income, the larger the consumption scale and the higher indirect HCEs, since indirect HCEs come from households' consumption of all goods and services and income is the most important factor leading to the expansion of consumption scale. Moreover, the indirect HCEs per capita of urban households was 3.24 tons but only 1.52 tons for rural ones. In addition to income, there were still large differences in lifestyle and consumption structure between urban and rural households, leading to a wider gap. In addition, the intensive effect of family size still worked for indirect HCEs.

The dwelling type and fuel type also had an impact on the indirect HCEs, though not as much as on direct HCEs. The indirect HCEs per capita of those "living in apartments" or with "LNG/nature gas as the main fuel" were higher than those for "living in low-rise building" or those using "electricity as the main fuel". Households using "firewood/straw as the main fuel" emitted more direct HCEs, but less indirect HCEs.

Moreover, it can be seen from Table 2 that households whose heads were younger or had higher education emitted more HCEs, especially indirect HCEs. However, only the variable of "college and above" was a more important factor. This may suggest that highly educated people may have more luxurious, and thus more carbon-intensive lifestyles.

In general, the influential directions of income, family size, region, birth generation and education level on the direct and indirect HCEs were the same, while urban or rural residency, dwelling type and fuel type (solar/biogas, coal, firewood/straw) were opposite. Figure 6 demonstrates this.

The trends of more coefficients with the decrease of *SP* are shown in Appendix Figure 1. It shows that for households with higher incomes, fewer family members, living in urban areas, living in northeast regions, living in apartments or low-rise buildings, using LNG/natural gas, or coal, or firewood/straw as the main fuel, the heads were unmarried, younger and with higher education. As such, they ought to be responsible for higher HCEs per capita.

Taken altogether, the results of LASSO regression show that fuel type and dwelling type can explain more than 70% of the direct HCEs, while income, urban or rural residency and fuel type are the three most important factors influencing indirect HCEs. Direct HCEs come from direct energy consumption, specifically the fuel consumption in the activities of transportation, cooking and heating. The relatively small household size and low household income limit the adoption of some appliances such as dishwashers and clothes dryers in China, so the consumptions of gas, coal and even firewood are still the main fuel resources of many households for cooking and heating. Thus it is reasonable that fuel type is the important factor of direct HCEs and that improving the thermal performance of homes is considered an important measure to reduce HCEs by policy-makers in many counties (Druckman and Jackson, 2010; Kurniawan et al., 2018; Zhao et al., 2012). The indirect HCEs embodied in the goods and services consumed by households are related to the scale and structure of consumption. Therefore, households' income and the differences in consumption patterns between urban and rural areas are the most important factors of households' consumption and HCEs(Jiang et al., 2020; Wang and Chen, 2020).

For a developing country like China, households' consumptions for energy and other productions are the determines of HCEs. Some household characteristics, such as age and marital status, have a statistically significant but relatively lower impact on HCEs.

Table 2 The result with LASSO regression model

	Direct HCEs	Direct HCEs	Direct HCEs	Indirect HCEs	Indirect HCEs	Indirect HCEs	Total HCEs	Total HCEs	Total HCEs
SP	**79**	**50**	**0**	**75**	**40**	**0**	**80**	**40**	**0**
MSE	0.520	0.510	0.510	2.404	2.336	2.335	3.394	3.300	3.299
Household income per capita	0.011	0.041	0.042	0.520	0.515	0.514	0.535	0.556	0.556
Household size	-0.029	-0.046	-0.048	-0.115	-0.149	-0.151	-0.137	-0.197	-0.199
Urban (Rural=1)	-0.006	-0.088	-0.093	0.692	0.695	0.691	0.538	0.597	0.598
Region (Center=0)									
East			0.097			0.182			0.279
West			0.137			0.163			0.299
Northeast	0.032	0.176	0.190			0.266			0.457
Fuel Type (Electricity=0)									
Solar/Biogas			0.056			-0.235			-0.179
LNG/Natural gas			0.146	0.258	0.271	0.267	0.090	0.394	0.413
Coal	0.979	1.227	1.233			-0.065	0.085	1.137	1.186
Firewood/Straw	0.357	0.507	0.527	-0.235	-0.317	-0.321			0.204
Others			0.051			-0.553			-0.503
House Type (Bungalow=0)									
Apartment			-0.059	0.251	0.431	0.443	0.177	0.366	0.384
Low-rise building	-0.043	-0.084	-0.092			0.209			0.117
Others			-0.041			0.313			0.272
Marital Status of Head (Unmarried=0)			0.015			-0.068			-0.053
Generation of Head (Generation_60S=0)									
Generation_10S-30S			-0.055			-0.037			-0.092
Generation_40S			-0.006			0.016			0.010
Generation_50S			0.016			-0.002			0.014
Generation_70S			0.011			0.046			0.057
Generation_80S			0.061			0.182			0.243
Generation_90S			0.065	0.022	0.385	0.398			0.463
Education of Head (Under primary=0)									
Primary			0.036			0.080			0.116
Junior secondary			0.040			0.145			0.185
Senior secondary			0.062			0.308			0.370
College and above			0.103	0.441	0.700	0.742	0.379	0.799	0.846

6 Conclusion and policy implications

In recent years, the critical role of HCEs in climate change mitigation is becoming increasingly recognized. In China, with the continuous growth of income and the ongoing improvement in living standards, considerable attention has been given to determining ways to mitigate HCEs. It is necessary to identify the key drivers of HCEs for effective policy intervention. However, the studies to date have been based on ad hoc hypotheses of drivers for HCEs which may end up to ambiguous results and cannot distinguish their relative importance.

In this paper, we calculated direct HCEs, indirect HCEs and total HCEs per capita for 37,620 households in China in the year of 2012, 2014 and 2016 using an emissions coefficient method and input-output model with survey data. Then with a LASSO regression model, we further prioritized the most important factors affecting the direct, indirect and total HCEs. We further estimate the trend of the coefficients for the main variables.

The results showed that cooking and heating with coal or firewood were the most influential source of the direct HCEs. Households with fewer members, living in bungalows, living in northeast regions, having higher income and living in rural areas were likely to emit more direct HCEs. In addition, the variables of fuel type and dwelling type can explain more than 70% of the direct HCEs, while the demographics of the household heads were not important. For indirect carbon emissions, income was the most important influencing factor, followed by urban or rural residency and fuel type.

Compared with the decomposition analysis and general regression model, the LASSO regression not only empirically tests the impact of many household demographic variables on HCEs, but also ranks their importance, thus identifies the most influential driving factors of HCEs. The use of the LASSO regression model in this paper not only addresses the issues of multicollinearity and over-fitting, but also provides policymakers a quantitative basis to determine what kind of abatement actions should be prioritized in order to reduce the emissions effectively, which is the contribution to the existing literature and practical policy-making.

Our study has the following policy implications. First of all, it is essential to promote a transition in household energy consumption structure since the fuel type is the most influential factor of direct HCEs. It is still necessary to continue to promote fuel transition in suburban and rural areas. ''coal to gas" or electrification, that is, changing household cooking and heating fuel from coal or firewood to natural gas, and further electrification in cooking is also desirable (Wu et al., 2017). However, such a transition should keep affordability under control since affordable clean energy is an important means to improve carbon emissions and residents' welfare. Solar PV may be the best choice in that it can reduce both environmental degradation and poverty (Xu et al., 2019), but it may not be able to be popularized in the short term due to financial constraints. If it is not possible to phase out coal, using coal more efficiently and more cleanly is a better option (Chen et al., 2016).

Secondly, the factors of family size, dwelling type and regions have a second important effect on the direct HCEs, which are related to buildings. The higher direct

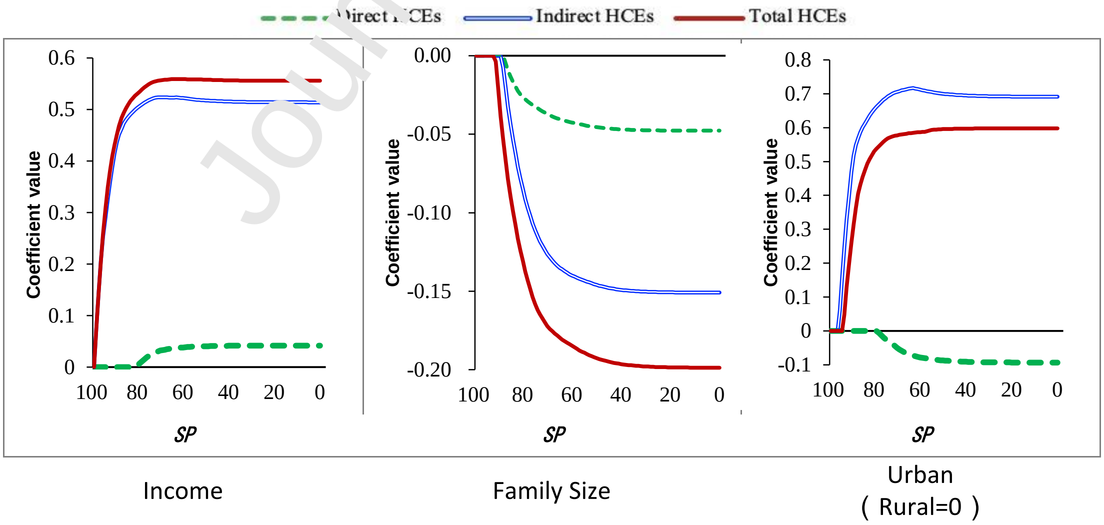
HCEs of households living in bungalows and indirect HCEs of households living in apartments are both relevant to the energy efficiency of residential dwellings, heating and cooking equipment, household appliances, automobiles and other facilities (Khanna et al., 2016). Therefore, it is very necessary to promote energy efficiency in the construction of buildings through renovating existing buildings or build new dwellings that are energy-efficient, and to popularize energy-efficient stoves, improve the efficiency of appliances and the fuel economy of automobiles so as to reduce household energy consumption and related carbon emissions (Firth et al., 2010). China's green building initiative needs to be further improved and promoted (Liu et al., 2019).

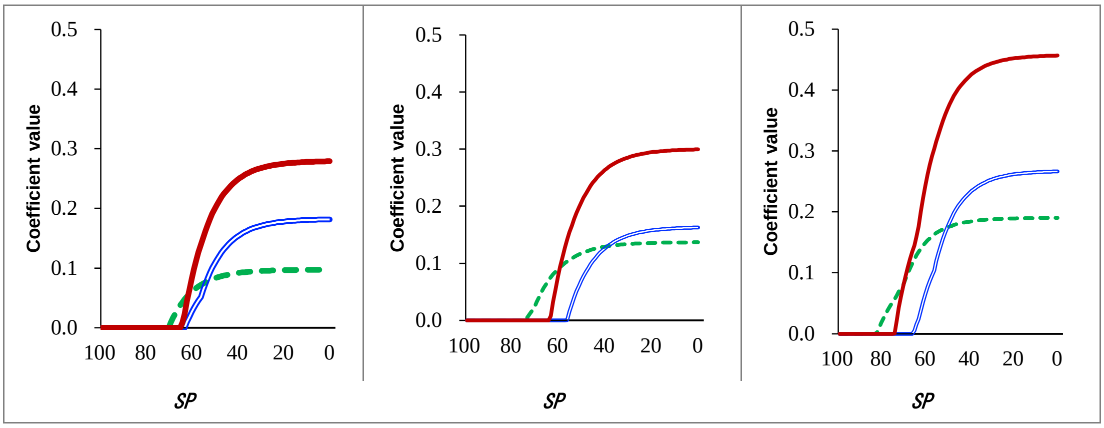
Thirdly, the results show that households' income and the attributes of urban or rural residency have an important impact on both direct and indirect HCEs. Amid the current stimulations for economic recovery, the governments should promote green and low-carbon economic recovery, and align the policies and measures with the Paris Agreement (Schwanen, 2020; Yao et al., 2018). On the one hand, the recovery package needs to serve the low carbon city construction, energy transition, climate action and the structural change to modernize the Chinese economy (Liu et al., 2017; Meng et al., 2018). On the other hand, Policies geared to stimulating consumption must consider the related HCEs. Public campaigns and information dissemination, such as energy efficiency labelling (Shi, 2014) and carbon labelling (Shi, 2013), can be used to inform and assist well-educated households in avoiding higher carbon lifestyles.

It is worth noting that although the employment of LASSO regression can provide a lot of crucial information for policy formulation, it is still subject to the limitations of linear regression models. Therefore, one possible future research direction is to incorporate other non-linear analytical techniques into the analysis, for example, artificial neural network, and distribution dynamics analysis, thereby complementing existing studies.

Acknowledgments

This work was supported by National Social Science Foundation of China (No.15BTJ021; No.14BJY047), Major project of National Social Science Foundation of China (No.18ZDA038), National Natural Science Foundation of China (No.71828401;), China Postdoctoral Science Foundation (No. 2018T110817), and Humanity and Social Science Foundation of the Ministry of Education of China (No.15JZD024; No.19YJAZH079; 20YJC790104).

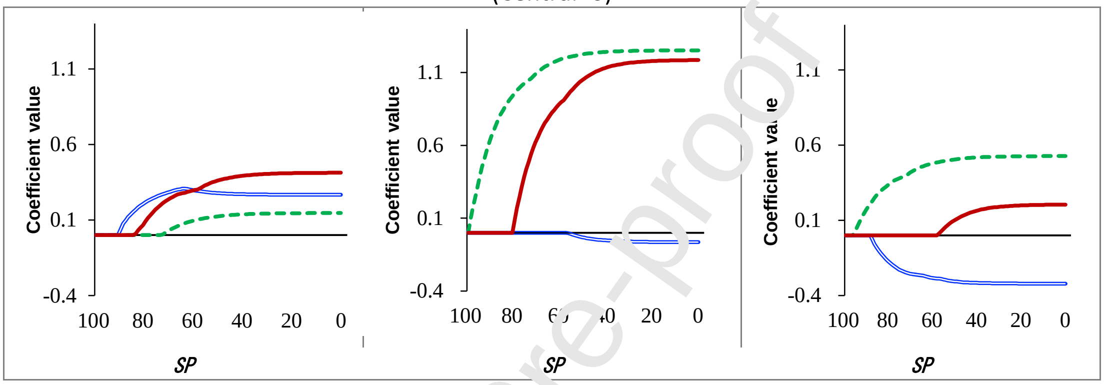




East

West
(Central=0)

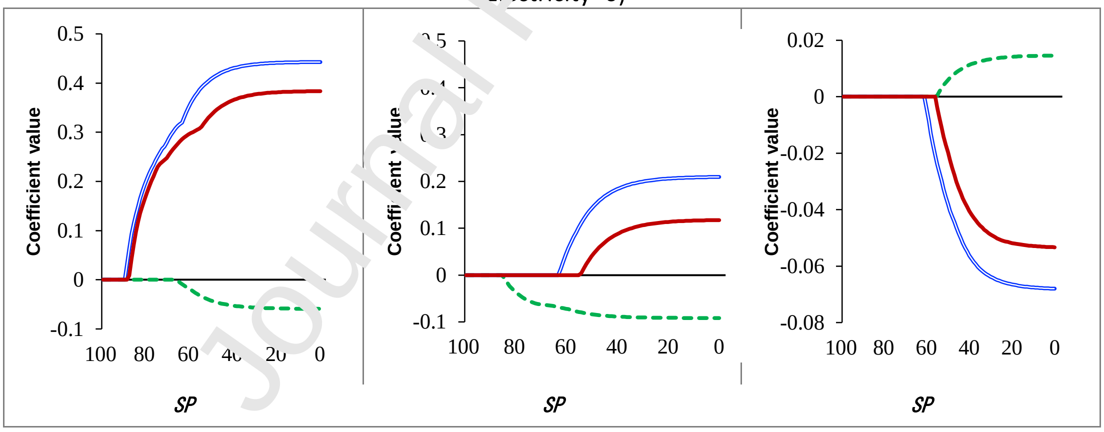
Northeast



LNG / Natural gas

Coal
(Electricity=0)

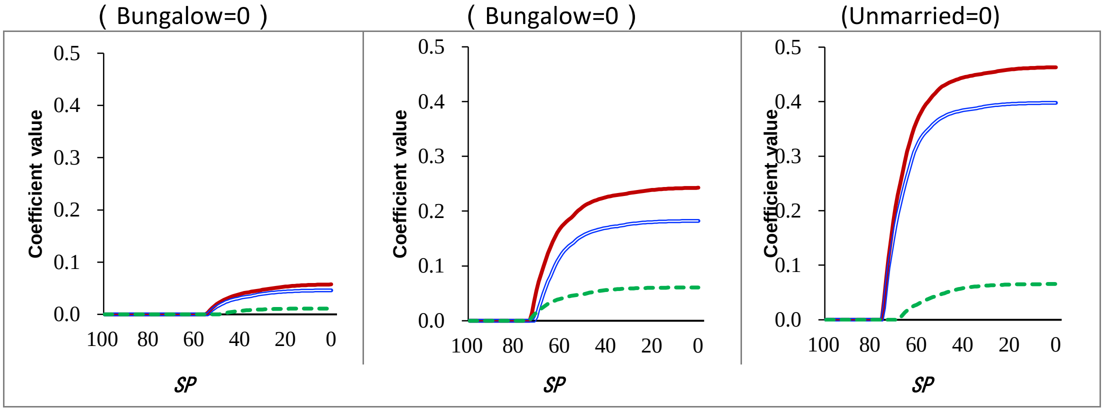
Firewood/Straw



Apartment
(Bungalow=0)

Low-rise building
(Bungalow=0)

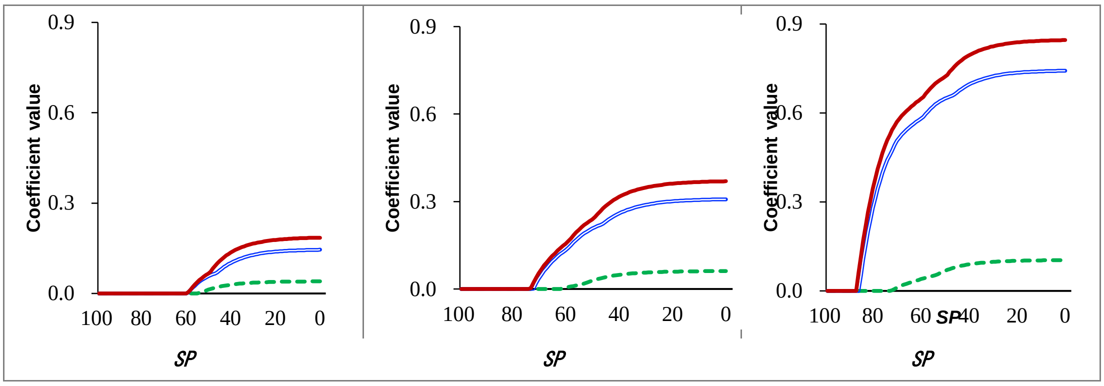
Marital Status
(Unmarried=0)



Generation_70S

Generation_80S
(Generation_60S=0

Generation_90S



Junior secondary | Senior secondary (Under primary=0) | College and above

Appendix Figure 1 The trend of coefficients with the decrease of *SP*

Reference

Baiocchi, G., Minx, J., Hubacek, K., 2010. The impact of social factors and consumer behavior on carbon dioxide emissions in the United Kingdom. J. Ind. Ecol. https://doi.org/10.1111/j.1530-9290.2009.00216.x

Beckel, C., Sadamori, L., Staake, T., Santini, S., 2014. Revealing household characteristics from smart meter data. Energy. https://doi.org/10.1016/j.energy.2014.10.025

BP, 2020. BP Statistical Review of World Energy Statistical Review of World, 69th edition. Ed. BP Stat. Rev. World Energy.

Büchs, M., Schnepf, S. V., 2013. Who emits most? Associations between socio-economic factors and UK households' home energy, transport, indirect and total CO2 emissions. Ecol. Econ. https://doi.org/10.1016/j.ecolecon.2013.03.007

Chancel, L., 2014. Are younger generations higher carbon emitters than their elders? Inequalities, generations and CO2emissions in France and in the USA. Ecol. Econ. https://doi.org/10.1016/j.ecolecon.2014.02.009

Chen, Q., Lv, M., Tang, Z., Wang, H., Wei, W., Sun, Y., 2016. Opportunities of integrated systems with CO2 utilization technologies for green fuel & chemicals production in a carbon-constrained society. J. CO2 Util. https://doi.org/10.1016/j.jcou.2016.01.004

Chitnis, M., Druckman, A., Hunt, L.C., Jackson, T., Milne, S., 2012. Forecasting scenarios for UK household expenditure and associated GHG emissions: Outlook to 2030. Ecol. Econ. https://doi.org/10.1016/j.ecolecon.2012.09.016

Choi, K., Zhang, M., 2017. The impact of metropolitan, county, and local land use on driving emissions in US metropolitan areas: Mediator effects of vehicle travel characteristics. J. Transp. Geogr. https://doi.org/10.1016/j.jtrangeo.2017.09.004

Dai, H., Masui, T., Matsuoka, Y., Fujimori, S., 2012. The impacts of China's household consumption expenditure patterns on energy demand and carbon emissions towards 2050. Energy Policy 50, 736–750. https://doi.org/http://dx.doi.org/10.1016/j.enpol.2012.08.023

Ding, Q., Cai, W., Wang, C., Sanwal, M., 2017. The relationships between household consumption activities and energy consumption in china— An input-output analysis from the lifestyle perspective. Appl. Energy. https://doi.org/10.1016/j.apenergy.2017.06.003

Druckman, A., Jackson, T., 2015. Understanding households as drivers of carbon emissions, in: Taking Stock of Industrial Ecology. https://doi.org/10.1007/978-3-319-20571-7_9

Druckman, A., Jackson, T., 2010. The bare necessities: How much household carbon do we really need? Ecol. Econ. https://doi.org/10.1016/j.ecolecon.2010.04.018

Duarte, R., Mainar, A., Sánchez-Chóliz, J., 2010. The impact of household consumption patterns on emissions in Spain. Energy Econ. https://doi.org/10.1016/j.eneco.2009.08.007

Fan, J., Guo, X., Marinova, D., Wu, Y., Zhao, D., 2012. Embedded carbon footprint of Chinese urban households: Structure and changes. J. Clean. Prod. 33, 50–59. https://doi.org/10.1016/j.jclepro.2012.05.018

Feng, Z.H., Zou, L. Le, Wei, Y.M., 2011. The impact of household consumption on energy use and CO2 emissions in China. Energy. https://doi.org/10.1016/j.energy.2010.09.049

Firth, S.K., Lomas, K.J., Wright, A.J., 2010. Targeting household energy-efficiency measures using sensitivity analysis. Build. Res. Inf. https://doi.org/10.1080/09613210903236706

Ghoddusi, H., Creamer, G.G., Rafizadeh, N., 2019. Machine learning in energy economics and finance: A review. Energy Econ. https://doi.org/10.1016/j.eneco.2019.05.006

Golley, J., Meng, X., 2012. Income inequality and carbon dioxide emissions: The case of Chinese urban households. Energy Econ. https://doi.org/10.1016/j.eneco.2012.07.025

Grossman, G.M., Krueger, A.B., 2006. Economic Growth and the Environment. Q. J. Econ. https://doi.org/10.2307/2118443

Han, Liyan, Xu, X., Han, Lu, 2015. Applying quantile regression and Shapley decomposition to analyzing the determinants of household embedded carbon emissions: Evidence from urban China. J. Clean. Prod. https://doi.org/10.1016/j.jclepro.2014.08.078

Hsu, D., 2015. Identifying key variables and interactions in statistical models ofbuilding energy consumption using regularization. Energy. https://doi.org/10.1016/j.energy.2015.02.008

Humeau, S., Wijaya, T.K., Vasirani, M., Aberer, K., 2013. Electricity load forecasting for residential customers: Exploiting aggregation and correlation between households, in: 2013 Sustainable Internet and ICT for Sustainability, SustainIT 2013. https://doi.org/10.1109/SustainIT.2013.6685208

Ipcc, 2006. 2006 Ipcc Guidelines for National Greenhouse Gas Inventories. Main. https://doi.org/http://www.ipcc-nggip.iges.or.jp/public/2006gl/pdf/2_Volume2/V2_3_Ch3_Mobile_Combustion.pdf
Irfany, M.I., Klasen, S., 2017. Affluence and emission tradeoffs: Evidence from Indonesian households' carbon footprint. Environ. Dev. Econ. https://doi.org/10.1017/S1355770X17000262
Jiang, Y., Long, Y., Liu, Q., Dowaki, K., Ihara, T., 2020. Carbon emission quantification and decarbonization policy exploration for the household sector - Evidence from 51 Japanese cities. Energy Policy. https://doi.org/10.1016/j.enpol.2020.111438
Jones, C., Kammen, D.M., 2014. Spatial distribution of U.S. household carbon footprints reveals suburbanization undermines greenhouse gas benefits of urban population density. Environ. Sci. Technol. https://doi.org/10.1021/es4034364
Khanna, N.Z., Guo, J., Zheng, X., 2016. Effects of demand side management on Chinese household electricity consumption: Empirical findings from Chinese household survey. Energy Policy. https://doi.org/10.1016/j.enpol.2016.04.049
Knittel, C.R., Stolper, S., 2019. Using Machine Learning to Target Treatment: The Case of Household Energy Use. NBER Work. Pap. Ser.
Kurniawan, R., Sugiawan, Y., Managi, S., 2018. Cleaner energy conversion and household emission decomposition analysis in Indonesia. J. Clean. Prod. https://doi.org/10.1016/j.jclepro.2018.08.051
Lakshmanan, V., Gilleland, E., McGovern, A., Tingley, M., 2015. Machine Learning and Data Mining Approaches to Climate Science, Machine Learning and Data Mining Approaches to Climate Science. Springer International Publishing. https://doi.org/10.1007/978-3-319-17220-0
Lee, S., Lee, B., 2014. The influence of urban form on GHG emissions in the U.S. household sector. Energy Policy. https://doi.org/10.1016/j.enpol.2014.01.024
Li, J., Zhang, D., Su, B., 2019. The Impact of Social Awareness and Lifestyles on Household Carbon Emissions in China. Ecol. Econ. https://doi.org/10.1016/j.ecolecon.2019.02.020
Liu, L.C., Wu, G., Wang, J.N., Wei, Y.M., 2011. China's carbon emissions from urban and rural households during 1992-2007, in: Journal of Cleaner Production. pp. 1754–1762. https://doi.org/10.1016/j.jclepro.2011.06.011
Liu, Y., Sun, X., Sun, T., Shi, X. (Roc), Liu, J., 2019. Promoting green residential buildings by increasing homebuyers' willingness to pay: Evidence from Sino-Singapore Tianjin Eco-city in China. J. Clean. Prod. 238. https://doi.org/10.1016/j.jclepro.2019.117884
Liu, Y., Tan, X.J., Yu, Y., Qi, S.Z., 2017. Assessment of impacts of Hubei Pilot emission trading schemes in China – A CGE-analysis using TermCO2 model. Appl. Energy. https://doi.org/10.1016/j.apenergy.2016.05.085
Long, Y., Dong, L., Yoshida, Y., Li, Z., 2018. Evaluation of energy-related household carbon footprints in metropolitan areas of Japan. Ecol. Modell. https://doi.org/10.1016/j.ecolmodel.2018.03.008

Longhi, S., 2015. Residential energy expenditures and the relevance of changes in household circumstances. Energy Econ. https://doi.org/10.1016/j.eneco.2015.03.018

Lyons, S., Pentecost, A., Tol, R.S.J., 2012. Socioeconomic distribution of emissions and resource use in Ireland. J. Environ. Manage. https://doi.org/10.1016/j.jenvman.2012.07.019

Ma, M., Ma, X., Cai, Weiguang, Cai, Wei, 2019. Carbon-dioxide mitigation in the residential building sector: A household scale-based assessment. Energy Convers. Manag. https://doi.org/10.1016/j.enconman.2019.111915

Maraseni, T.N., Qu, J., Yue, B., Zeng, J., Maroulis, J., 2016. Dynamism of household carbon emissions (HCEs) from rural and urban regions of northern and southern China. Environ. Sci. Pollut. Res. https://doi.org/10.1007/s11356-016-7237-5

Maraseni, T.N., Qu, J., Zeng, J., 2015. A comparison of trends and magnitudes of household carbon emissions between China, Canada and UK. Environ. Dev. https://doi.org/10.1016/j.envdev.2015.04.001

Meangbua, O., Dhakal, S., Kuwornu, J.K.M., 2019. Factors influencing energy requirements and CO2 emissions of households in Thailand: A panel data analysis. Energy Policy. https://doi.org/10.1016/j.enpol.2019.02.050

Meier, H., Rehdanz, K., 2010. Determinants of residential space heating expenditures in Great Britain. Energy Econ. https://doi.org/10.1016/j.eneco.2009.11.008

Meng, B., Liu, Y., Andrew, R., Zhou, M., Hubacek, K., Xue, J., Peters, G., Gao, Y., 2018. More than half of China's CO2 emissions are from micro, small and medium-sized enterprises. Appl. Energy. https://doi.org/10.1016/j.apenergy.2018.08.107

Mi, Z., Zheng, J., Meng, J., Ou, J., Hubacek, K., Liu, Z., Coffman, D.M., Stern, N., Liang, S., Wei, Y.M., 2020. Economic development and converging household carbon footprints in China. Nat. Sustain. https://doi.org/10.1038/s41893-020-0504-y

Miao, H., Ramchander, S., Wang, T., Yang, D., 2017. Influential factors in crude oil price forecasting. Energy Econ. https://doi.org/10.1016/j.eneco.2017.09.010

Motawa, I., Oladokun, M.G., 2015. Structural equation modelling of energy consumption in buildings. Int. J. Energy Sect. Manag. https://doi.org/10.1108/IJESM-11-2014-0004

Munksgaard, J., Pedersen, K.A., Wien, M., 2000. Impact of household consumption on CO2 emissions. Energy Econ. https://doi.org/10.1016/S0140-9883(99)00033-X

Murray, A.G., Mills, B.F., 2011. Read the label! Energy Star appliance label awareness and uptake among U.S. consumers. Energy Econ. https://doi.org/10.1016/j.eneco.2011.04.013

Niamir, L., Filatova, T., Voinov, A., Bressers, H., 2018. Transition to low-carbon economy: Assessing cumulative impacts of individual behavioral changes. Energy Policy. https://doi.org/10.1016/j.enpol.2018.03.045

Niamir, L., Ivanova, O., Filatova, T., Voinov, A., Bressers, H., 2020. Demand-side solutions for climate mitigation: Bottom-up drivers of household energy

behavior change in the Netherlands and Spain. Energy Res. Soc. Sci. https://doi.org/10.1016/j.erss.2019.101356
Papathanasopoulou, E., 2010. Household consumption, associated fossil fuel demand and carbon dioxide emissions: The case of Greece between 1990 and 2006. Energy Policy. https://doi.org/10.1016/j.enpol.2010.03.043
Qu, J., Zeng, J., Li, Y., Wang, Q., Maraseni, T., Zhang, L., Zhang, Z., Clarke-Sather, A., 2013. Household carbon dioxide emissions from peasants and herdsmen in northwestern arid-alpine regions, China. Energy Policy. https://doi.org/10.1016/j.enpol.2012.12.065
Rong, P., Zhang, L., Qin, Y., Chen, S., Sun, Y., 2018. The influencing factors of urban household embedded carbon emissions based on quantile regression, in: Energy Procedia. https://doi.org/10.1016/j.egypro.2018.09.238
Sala-i-Martin, X.X., 1997. I Just Ran Two Million Regressions. Am. Econ. Rev. https://doi.org/10.1126/science.151.3712.867-a
Schipper, L., 1989. Linking Life-Styles And Energy Use: A Matter Of Time. Annu. Rev. Energy Environ. https://doi.org/10.1146/annurev.energy.14.1.273
Schopfer, S., Tiefenbeck, V., Staake, T., 2018. Economic assessment of photovoltaic battery systems based on household load profiles. Appl. Energy. https://doi.org/10.1016/j.apenergy.2018.03.185
Schwanen, T., 2020. Low-Carbon Mobility in London: A Just Transition? One Earth. https://doi.org/10.1016/j.oneear.2020.01.013
Seriño, M.N. V., 2017. Effects of affluence on rising household carbon emission in the Philippines: An application using quantile regression approach. DLSU Bus. Econ. Rev.
Seriño, M.N. V., Klasen, S., 2015. Estimation and determinants of the Philippines' household carbon footprint. Dev. Econ. https://doi.org/10.1111/deve.12065
Shi, X., 2014. Setting effective mandatory energy efficiency standards and labelling regulations: A review of best practices in the Asia Pacific region. Appl. Energy 133, 135–143. https://doi.org/10.1016/j.apenergy.2014.07.084
Shi, X., 2013. Spillover effects of carbon footprint labelling on less developed countries: The example of the east asia summit region. Dev. Policy Rev. 31, 239–254. https://doi.org/10.1111/dpr.12005
Shi, X., Yu, J., Cheong, T.S., 2020. Convergence and distribution dynamics of energy consumption among China's households. Energy Policy 142, 111496. https://doi.org/10.1016/j.enpol.2020.111496
Shigetomi, Y., Matsumoto, K., Ogawa, Y., Shiraki, H., Yamamoto, Y., Ochi, Y., Ehara, T., 2018. Driving forces underlying sub-national carbon dioxide emissions within the household sector and implications for the Paris Agreement targets in Japan. Appl. Energy. https://doi.org/10.1016/j.apenergy.2018.07.057
Tibshirani, R., 2011. Regression shrinkage and selection via the lasso: A retrospective. J. R. Stat. Soc. Ser. B Stat. Methodol. https://doi.org/10.1111/j.1467-9868.2011.00771.x
Tibshirani, R., 1996. Regression Shrinkage and Selection Via the Lasso. J. R. Stat. Soc. Ser. B. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Underwood, A., Zahran, S., 2015. The carbon implications of declining household scale economies. Ecol. Econ. https://doi.org/10.1016/j.ecolecon.2015.04.028

Viegas, J.L., Vieira, S.M., Sousa, J.M.C., Melício, R., Mendes, V.M.F., 2015. Electricity demand profile prediction based on household characteristics, in: International Conference on the European Energy Market, EEM. https://doi.org/10.1109/EEM.2015.7216746

Wang, L., Wang, Y., Cheng, H., Cheng, J., 2019. Identifying the driving factors of black bloom in lake bay through bayesian LASSO. Int. J. Environ. Res. Public Health. https://doi.org/10.3390/ijerph16142492

Wang, X., Chen, S., 2020. Urban-rural carbon footprint disparity across China from essential household expenditure: Survey-based analysis, 2010–2014. J. Environ. Manage. https://doi.org/10.1016/j.jenvman.2020.110570

Wang, Z., Cui, C., Peng, S., 2019. How do urbanization and consumption patterns affect carbon emissions in China? A decomposition analysis. J. Clean. Prod. https://doi.org/10.1016/j.jclepro.2018.11.272

Wei, L., Li, C., Wang, J., Wang, X., Wang, Z., Cui, C., Peng, S., Liu, Y., Yu, S., Wang, L., Shi, Z., 2020. Rising middle and rich classes drove China's carbon emissions. Resour. Conserv. Recycl. https://doi.org/10.1016/j.resconrec.2020.104839

Wei, L., Tian, W., Silva, E.A., Choudhary, R., Meng, Q., Yang, S., 2015. Comparative Study on Machine Learning for Urban Building Energy Analysis, in: Procedia Engineering. https://doi.org/10.1016/j.proeng.2015.08.1070

Wiedenhofer, D., Guan, D., Liu, Z., Meng, J., Zhang, N., Wei, Y.M., 2017. Unequal household carbon footprints in China. Nat. Clim. Chang. https://doi.org/10.1038/nclimate3165

Wilson, J., Tyedmers, P., Spinney, J.E.L., 2013. An exploration of the relationship between socioeconomic and well-being variables and household greenhouse gas emissions. J. Ind. Ecol. https://doi.org/10.1111/jiec.12057

Wu, S., Zheng, X., Wei, C., 2017. Measurement of inequality using household energy consumption data in rural China. Nat. Energy. https://doi.org/10.1038/s41560-017-0003-1

Xie, Y., Hu, J., 2014. An introduction to the China Family Panel Studies (CFPS). Chin. Sociol. Rev. https://doi.org/10.2753/CSA2162-0555470101

Xu, X.Y., Ang, B.W., 2014. Analysing residential energy consumption using index decomposition analysis. Appl. Energy. https://doi.org/10.1016/j.apenergy.2013.07.052

Yao, X., Kou, D., Shao, S., Li, X., Wang, W., Zhang, C., 2018. Can urbanization process and carbon emission abatement be harmonious? New evidence from China. Environ. Impact Assess. Rev. https://doi.org/10.1016/j.eiar.2018.04.005

Ye, H., Qiu, Q., Zhang, G., Lin, T., Li, X., 2013. Effects of natural environment on urban household energy usage carbon emissions. Energy Build. https://doi.org/10.1016/j.enbuild.2013.06.001

Zhang, H., Shi, X., Wang, K., Xue, J., Song, L., Sun, Y., 2020. Intertemporal lifestyle changes and carbon emissions: Evidence from a China household survey. Energy Econ. 86, 104655. https://doi.org/10.1016/j.eneco.2019.104655

Zhang, Hongwu, Zhang, L., Wang, K., Shi, X., 2019. Unveiling key drivers of indirect carbon emissions of Chinese older households. Sustain. https://doi.org/10.3390/su11205740
Zhang, H., Zhang, L., Wang, K., Shi, X., 2019. Unveiling key drivers of indirect carbon emissions of Chinese older households. Sustain. 11. https://doi.org/10.3390/su11205740
Zhang, N., Yu, K., Chen, Z., 2017. How does urbanization affect carbon dioxide emissions? A cross-country panel data analysis. Energy Policy. https://doi.org/10.1016/j.enpol.2017.03.072
Zhang, Tong, Shi, X., Zhang, D., Xiao, J., 2019. Socio-economic development and electricity access in developing economies: A long-run model averaging approach. Energy Policy 132, 223–231. https://doi.org/10.1016/j.enpol.2019.05.031
Zhang, W., Robinson, C., Guhathakurta, S., Garikapati, V.M., Dilkina, B., Brown, M.A., Pendyala, R.M., 2018. Estimating residential energy consumption in metropolitan areas: A microsimulation approach. Energy. https://doi.org/10.1016/j.energy.2018.04.161
Zhang, X., Luo, L., Skitmore, M., 2015. Household carbon emission research: An analytical review of measurement, influencing factors and mitigation prospects. J. Clean. Prod. https://doi.org/10.1016/j.jclepro.2015.04.024
Zhang, Yaojie, Ma, F., Wang, Y., 2019. Forecasting crude oil prices with a large set of predictors: Can LASSO select powerful predictors? J. Empir. Financ. https://doi.org/10.1016/j.jempfin.2019.08.007
Zhao, P., Yu, B., 2006. On model selection consistency of Lasso. J. Mach. Learn. Res.
Zhao, X., Li, N., Ma, C., 2012. Residential energy consumption in urban China: A decomposition analysis. Energy Policy. https://doi.org/10.1016/j.enpol.2011.11.027
Zhu, H., Xia, H., Guo, Y., Peng, C., 2018. The heterogeneous effects of urbanization and income inequality on CO2emissions in BRICS economies: evidence from panel quantile regression. Environ. Sci. Pollut. Res. https://doi.org/10.1007/s11356-018-1900-y

Credit Author Statement

Xunpeng Shi: Conceptualization, Validation, Writing- Reviewing

Keying Wang: Data curation, Visualization, Validation, Writing- Reviewing and Editing

Tsun Se Cheong: Methodology, Software, Validation, Writing- Reviewing

Hongwu Zhang: Writing- Original draft preparation, Validation, Writing- Reviewing and Editing

Highlights

- Identified the most influential driving factors of HCEs with the LASSO model and survey data.
- Three factors can explain more than 70% of the total decline in MSE of HCEs.
- Fuel type and dwelling type are the most influential factors of direct HCEs.
- Income, urban residency, and fuel type are the most influential factors of indirect HCEs.
- The results with the LASSO model offer more flexibility in emission mitigation policies.