

# Household fuel choice in urban China: evidence from panel data

XIAO-BING ZHANG

*School of Economics, Renmin University of China, Beijing 100872, China.*  
*Email: [xiaobing.zhang@ruc.edu.cn](mailto:xiaobing.zhang@ruc.edu.cn)*

SIED HASSEN

*Environment and Climate Research Center, Ethiopian Development Research Institute; and Department of Economics, Mekelle University, Ethiopia.*  
*Email: [seidy2004@gmail.com](mailto:seidy2004@gmail.com)*

*Submitted 28 November 2013; revised 9 December 2014, 23 February 2015, 15 July 2016; accepted 10 February 2017; first published online 20 April 2017*

**ABSTRACT.** Using eight rounds of household survey data that span two decades, this paper analyzes the determinants of household fuel choice in urban China. Using the correlated random effects generalized ordered probit model, the authors find that household fuel choice in urban China is related to fuel prices, households' economic status and size and household head's gender and education. The results suggest that policies and interventions that increase households' income, reduce the price advantage of dirty fuels (e.g., taxing coal) and empower women in the household are of great significance in encouraging the use of clean energy sources.

## 1. Introduction

Half of the world's population and up to 95 per cent of people in developing countries rely on solid fuels (biomass fuels and coal) to meet

The authors gratefully thank an associate editor and two anonymous referees for their helpful comments and suggestions on the preliminary draft of this paper, according to which the content was improved. They would also like to thank the Carolina Population Center for providing the data sets of the China Health and Nutrition Survey, funded by the NIH (R01-HD30880, DK056350 and R01-HD38700), and Yonas Alem, Jessica Coria, Thomas Sterner, Subhrendu K. Pattanayak and Gunnar Köhlin for their helpful discussions and comments on this paper. All errors and omissions remain the sole responsibility of the authors. Financial support from the National Natural Science Foundation of China (No. 71603267) and the Swedish International Development Cooperation Agency (Sida) through the Environmental Economics Unit at the University of Gothenburg is gratefully acknowledged.

their energy needs (IEA, 2011). Household dependence on solid fuels for cooking has health and environmental impacts. Conservative estimates document that exposure to indoor smoke produced by household solid fuel combustion is responsible for about 2 million premature deaths per year globally, which is 3.3 per cent of the global burden of disease. About 548,900 of these deaths occurred in China alone in 2004 (Smith *et al.*, 1994, 2004; WHO, 2008).

With increasing household wellbeing and income in China, especially in urban areas, more and more households have shifted from the traditional firewood or coal to modern energy, such as liquefied natural gas (LNG) or electricity. China's experience makes it a good example with which investigate the economic and social determinants of household cooking fuel choices. Understanding these determinants will be helpful in finding ways to accelerate the transition to cleaner fuels in developing countries more generally.

Traditionally, the 'energy ladder' hypothesis has been used to explain households' fuel choices and switching strategies in developing countries. This hypothesis describes income as the sole factor in determining these decisions. However, fuel choice behavior of households is not as simple as is prescribed by the traditional energy ladder hypothesis (Masera *et al.*, 2000; Heltberg, 2005), and the simple association between income and fuel demand (choice) has been criticized in recent literature because fuel choice can be affected by a multitude of demographic and socio-economic factors (Masera *et al.*, 2000; Heltberg, 2005; Mekonnen and Köhlin, 2008).

A number of previous studies analyzed the determinants of households' fuel choice in the developing world. However, many of these studies are based on cross-sectional data (e.g., Hosier and Dowd, 1987; Leach, 1992; Farsi *et al.*, 2007) and studies employing panel data are rare (e.g., Mekonnen and Köhlin, 2008). Moreover, previous studies in China are mainly based on aggregate statistics or on surveys conducted in certain provinces or counties (e.g., ESMAP, 1996; Wang and Feng, 1997; Chen *et al.*, 2006). As far as we are aware, very few studies have examined household energy choices in China through longitudinal data from a nationwide household survey. This paper tries to fill this gap by using eight-round panel data from the China Health and Nutrition Survey (CHNS). These panel data enable us to control for unobserved individual heterogeneity and time trends in the analysis of household fuel choice.

This study focuses on households' primary cooking fuel choices in urban areas of China, only because the fuel choices of rural households are largely determined by fuel availability and opportunity costs for fuel collection rather than by budget constraints, which complicate the modeling of household fuel choice in such circumstances (Farsi *et al.*, 2007). Further, while fuel choices for cooking and heating are the main interests of the literature on household energy choices, this paper concentrates on households' cooking fuel choices given the fact that heating in urban China is mainly provided at the district level, which means that households have little freedom to choose the type of heating energy used. Finally, we focus on households' choices of primary cooking energy. Primary cooking energy

is the type that is most frequently used by a household.<sup>1</sup> *Farsi et al. (2007)* used a similar definition in their study of fuel choice in India.

In line with *Farsi et al. (2007)*, ordered probit models were employed to take into account the potential ordering of different fuels in terms of efficiency or convenience to use. In addition, we contribute two extensions to the application of an ordered discrete choice model to the fuel choice issue. First, this paper employed a more flexible empirical framework through generalized ordered probit models rather than the standard ordered probit model, which is based on a restrictive assumption of the parallel regression or the same slope coefficients across the different fuel categories, implying a homogeneous effect of the explanatory variables across the distribution of fuel categories. Secondly, to explore the panel structure of the data set, the random effects generalized ordered probit model with Mundlak transformation was adopted to analyze the household fuel choices. The Mundlak approach enables us to address the potential bias that would come from possible correlation between the unobserved heterogeneity and the explanatory variables.

Our results suggest that policies and interventions that raise households' income, reduce the price advantage of dirty fuels (e.g., taxing coal) and empower women in the household are of great significance in encouraging the adoption of clean energy sources. The results also show the importance of other sociodemographic factors such as education in determining the choices of primary cooking fuels in urban Chinese households.

This paper is organized as follows. Section 2 provides a brief review of related literature. Section 3 describes the empirical strategy used in this study. Section 4 presents the data and some descriptive statistics. The estimation results of our econometric model are illustrated and discussed in section 5. Finally section 6 summarizes the conclusions and policy implications.

## 2. Review of literature

A growing body of empirical studies is attempting to investigate the energy choices and switching strategies of households in developing countries. These studies focus on the effect of household characteristics, income and prices on fuel choices and also on the validity of the 'energy ladder' hypothesis. In the paragraphs below, we present a brief review of previous studies, focusing on households' fuel choices and switching strategies, and highlighting existing knowledge gaps.

The 'energy ladder' hypothesis is based on the assumption that households are exposed to a number of fuel choices, which can be ranked in order of increasing efficiency and technological sophistication, and that households make the transition to the higher ranked fuel as their income rises

<sup>1</sup> The survey identifies the cooking fuel used most frequently by a household. Due to lack of information on the proportion of each fuel for the households that use multiple fuels, it is difficult to effectively incorporate the multiple fuel choices into our estimation.

(Hosier and Dowd, 1987). Electricity, natural gas and other commercial fossil fuels are ranked higher than the traditional biomass fuels. The energy choice of a household will move 'up' the energy ladder to higher ranked fuels as its income increases. A few earlier studies provide evidence for this hypothesis (e.g., Alam *et al.*, 1985; Sathaye and Tyler, 1991). Alam *et al.* (1985) found that income has a direct effect on household fuel choice decisions. The higher the income level, the greater the tendency for households to choose commercial fuels over biomass fuels. Using a cross-section of 1,000 sample households from Bangalore, India, Reddy (1995) examined household energy choices through a series of binomial logit models for different pairs of energy carriers. He confirmed the hypothesis of the energy ladder and the importance of income in household energy choices.

However, Reddy (1995: 936) also argued that 'as times change, societies become more egalitarian and this energy-ladder concept based on income may disappear'. In fact, the simple association between income and fuel choice has been criticized in recent literature. Fuel choice can be affected by a multitude of demographic and socio-economic factors (Barnett, 2000; Masera *et al.*, 2000; Heltberg, 2005; Mekonnen and Köhlin, 2008). A limited but increasing number of studies from the largest developing countries, such as India and China, provide evidence of the multiple factors that determine household fuel choice (e.g., Jiang and O'Neill, 2004; Farsi *et al.*, 2007; Pachauri and Jiang, 2008). Household size, education, and gender of the household head are found to be among the key determinants of fuel choice and transition.

Despite these findings, the existing empirical research in this area documents mixed results for some of these factors. For example, while Hosier and Dowd (1987) found that large households tend to move away from wood and toward kerosene, the finding of Ouedraogo (2006) indicates that small households are more likely to use liquefied petroleum gas (LPG) and less likely to use firewood. Unlike these two studies, Heltberg (2004) found insignificant effects of household size on fuel transition (switching). Similarly, the empirical evidence on the effect of prices or relative prices is also mixed. Leach (1992) found that relative fuel prices are less important for households' substitution of traditional biomass fuels by modern energy sources. Likewise, Zhang and Kotani (2012) found that coal and LPG prices do not exhibit substitution effects. Nonetheless, Heltberg (2005) and Gupta and Köhlin (2006) found significant cross-price effects between different fuel types. Regarding the effect of education, most studies found a positive effect of education on the transition to high-quality fuel (e.g., Heltberg, 2004; 2005; Jiang and O'Neill, 2004, Farsi *et al.*, 2007). Some studies also looked at effect of locations or regions (Hosier and Dowd, 1987; Leach, 1992; Heltberg, 2004; 2005). Contrary to the energy ladder hypothesis, recent literature documents that 'fuel switching' in developing countries is often not complete and is, in fact, a gradual process with many households often using multiple fuels; the reasons for multiple fuel use are varied, from supply security to cultural, social or taste preferences (Masera *et al.*, 2000; Heltberg, 2005; Farsi *et al.*, 2007; Mekonnen and Köhlin, 2008).

Many studies have analyzed the determinants of fuel choices in the developing world based on cross-sectional data (e.g., [Hosier and Dowd, 1987](#); [Leach, 1992](#); [Farsi et al., 2007](#)). Few studies have employed panel data (e.g., [Mekonnen and Köhlin, 2008](#)). Regarding the studies on China, most of them are based on aggregate statistics, on surveys conducted in certain provinces or counties, or on rural households ([ESMAP, 1996](#); [Wang and Feng, 1997](#); [Chen et al., 2006](#); [Peng et al., 2010](#); [Zhang and Kotani, 2012](#)). Based on aggregate statistics and descriptive statistical tests, [Cai and Jiang \(2008\)](#) tested the energy ladder hypothesis by comparing the energy consumption pattern of rural households with that of urban households. Their results show that urban households use fuel that is more convenient, cleaner and more efficient than that used in rural areas, where biomass and coal are common fuel. [Peng et al. \(2010\)](#) studied household-level fuel switching using cross-sectional data from rural Hubei. They found that fuel use varies enormously across geographic regions due to disparities in availability of different energy sources. Their results indicate that rural households do switch to commercial energy sources, with coal being the principal substitute for biomass. Using household survey data from rural Beijing, ([Zhang and Kotani, 2012](#)) found that coal and LPG prices do not exhibit substitution effects. While many of the studies are based on surveys conducted in certain provinces or villages, very few examined household energy choices through a nationwide household survey. [Jiang and O'Neill \(2004\)](#) explored patterns of residential energy use in rural China by using a nationally representative rural household survey and various sources of aggregate statistics.

From the above empirical evidence on fuel choices in China and other developing countries, we observe the following knowledge gaps. First, many of the existing studies in China are based on surveys in a certain province or county. Due to the large regional variations across China, the experience from one region may not be perfectly applicable in another region, which highlights the importance of controlling regional variations in relevant studies. Secondly, most of the previous studies on household fuel choice in developing countries are based on cross-sectional data in which it is difficult to control for the unobserved individual heterogeneity and the potential bias from the correlation of unobserved heterogeneity and explanatory variables. Our study seeks to fill these gaps.

### 3. Empirical strategy

Our choice of empirical strategy is based on the ordinal ranking of different fuel types in terms of their convenience to use, comfort and modernity,<sup>2</sup> which is also consistent with the transition of households in urban China from solid fuel sources to clean energy over a couple of decades, as discussed in the section below. For instance, among the three fuel types considered in this study (firewood, coal and LNG), LNG is the most convenient

<sup>2</sup> See, for example, figure 2 in the WHO report, *Fuel for Life: Household Energy and Health* (WHO, 2006: 9).

to use and efficient fuel source while firewood is the least efficient and least convenient to use. Considering this, an ordered discrete choice framework is used in this study.<sup>3</sup> Ordered probit models have already been applied to household cooking fuel choices in the recent literature (e.g., *Farsi et al., 2007; Mensah and Adu, 2013; Nlom and Karimov, 2014*). However, to our knowledge, none of the studies using ordered probit models has analyzed household fuel choice by employing the ordered probit model with panel data application, where the unobserved individual heterogeneity can be better dealt with by exploring the panel structure of the data set.

Given the above ordinal fuel choice structure and assuming that individual households' fuel choices are based on a latent variable ( $E_{it}^*$ ), the random effects ordered response model can be written as:

$$E^* = x'_{it}\beta + \alpha_i + \varepsilon_{it}, (\varepsilon_{it}|x_{it}) \sim N(0, 1), (\alpha_i|x_{it}) \sim N(0, \sigma_\alpha^2) \tag{1}$$

$$E_{it} = j \text{ if } \mu_{j-1} < E_{it}^* < \mu_j \text{ for } j = 0, 1, 2, \dots, J \text{ and } \mu_{-1} = -\infty, \mu_J = +\infty, \tag{2}$$

where  $E_{it}$  represents the observed cooking fuel choice for household  $i = 1, \dots, n$  in time period  $t = 1, \dots, T$ , which can be ordered in terms of efficiency or convenience (e.g.,  $E = 0$  for firewood,  $E = 1$  for coal, and  $E = 2$  for LNG, etc.).  $x_{it}$  denotes a vector of explanatory variables, including income and other household characteristics.  $\beta$  is the vector of parameter estimates for explanatory variables, and the  $\mu$ s denote the unknown threshold values to be estimated with  $\beta$ .  $\varepsilon_{it}$  is the time-varying unobserved heterogeneity which is assumed to be normally distributed with mean zero and variance one.  $\alpha_i$  is the time-invariant unobserved heterogeneity. Conditional on  $x_{it}$ ,  $\alpha_i$  is normally distributed with mean zero and variance  $\sigma_\alpha^2$ , and is assumed to be independent of  $\varepsilon_{it}$  and  $x_{it}$  (something to be relaxed later on). Under these assumptions, it can be shown that the correlation between the composite error terms ( $\alpha_i + \varepsilon_{it}$ ) across any two time periods is given by  $\rho = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\varepsilon^2}$ , which can also be considered as a measure of the relative importance of the unobserved effect (*Wooldridge, 2010: 608–662*). Conditional on  $x_{it}$  and  $\alpha_i$ , the random effects ordered probit model is:

$$P(E_{it} = j|x_{it}, \alpha_i) = \Phi(\mu_j - \alpha_i - x'_{it}\beta) - \Phi(\mu_{j-1} - \alpha_i - x'_{it}\beta) \text{ for } j = 0, 1, 2, \dots, J; \tag{3}$$

where  $\Phi$  is the standard normal cumulative distribution function with  $\Phi(\mu_{-1}) = 0$  and  $\Phi(\mu_J) = 1$ . From equation (3), we can see that the random

<sup>3</sup> However, one may argue that the real process of decision making is not known to econometricians and, hence, the ranking of the fuels may not be obvious. Alternatively, one may use a multinomial logit model rather than the ordered regression models. Nonetheless, (*Anderson, 1984*) argued that, in cases where ordering is not 'a priori obvious', a generalized ordered probit/logit model (which will be discussed later in this section) is preferable for the interpretation of coefficient estimates. Also, a multinomial logit model will result in inefficient estimates if the ordering is inherent in the household fuel preference (*Boes and Winkelmann, 2010*).

effects ordered probit model takes into account the unobserved heterogeneity, which cannot be handled in the standard ordered probit analysis using cross-sectional data.

However, the standard random effects ordered probit models implicitly impose the parallel regression assumption. This implies a homogeneous effect of the explanatory variables across the cumulative distribution of cooking fuel types, i.e., single crossing of marginal probability effects or constant relative effects (Maddala, 1983; Boes and Winkelmann, 2006). To relax this rather restrictive assumption, we can employ a more flexible framework through a generalized ordered probit model, where the effects of explanatory variables across the cumulative distribution of the dependent variable are unrestricted (Boes and Winkelmann, 2006). This can be carried out by making the threshold values linear functions of the explanatory variables, i.e.,  $\mu_{ij} = k_j + x'_{it}\lambda_j$  (Terza, 1985). Substituting  $\mu_{ij} = k_j + x'_{it}\lambda_j$  in equation (3) gives the generalized random effects model:

$$P(E_{it} = j | x_{it}, \alpha_i) = \Phi(k_j - \alpha_i - x'_{it}\beta_j) - \Phi(k_{j-1} - \alpha_i - x'_{it}\beta_j) \text{ for } j = 0, 1, 2, \dots, J, \quad (4)$$

where the estimated coefficients are  $\beta_j = \beta - \lambda_j$ . Thus, we can see that the heterogeneity in the generalized model makes the vector of parameter estimates,  $\beta_j$ , become category specific. The standard random effects ordered probit model can be considered as a special case of the generalized model with the imposition of the restriction  $\beta_1 = \dots = \beta_J$ .

Both the standard and general random effects ordered probit models assume that time-invariant unobserved heterogeneity ( $\alpha_i$ ) are independent of the explanatory variables. However this is a rather restrictive assumption as some of the time-invariant unobserved heterogeneity, such as motivation, may be correlated with some of the regressors in the model, such as education and income, which in turn may introduce bias in the coefficient estimates. Following (Boes and Winkelmann, 2010), (Chamberlain, 1980) and (Mundlak, 1978), it is possible to estimate more precise estimates for the generalized ordered probit framework by allowing for possible correlation between  $\alpha_i$  and  $x_{it}$ , which involves including the averages of time-varying regressors in the model.<sup>4</sup> This approach relies on the assumption that the time-invariant unobserved effects are linearly

<sup>4</sup> The application of the standard fixed effects method to generalized discrete choice models, however, is not easy, as there are no simple transformations (first difference or within transformation). This is also difficult for the standard ordered probit due to its functional form complexity. As a result, the application of the fixed effects method to ordered probit models is rare. Nonetheless, the application of the standard fixed effects method is relatively easier for ordered logit models and there are a growing number of empirical studies based on fixed effects ordered logit models (e.g., Booth and van Ours, 2008; Kassenboehmer and Haisken-DeNew, 2009). We also present in the appendix the results from two fixed effects estimates of ordered logit models that are commonly used in the literature: the Ferrer-i-Carbonell and Frijters (FF) estimator (Ferrer-i-Carbonell and Frijters, 2004) and the 'Blow-up and Cluster' (BUC) estimator (Baetschmann *et al.*, 2011)

correlated with explanatory variables, as specified by:

$$\alpha_i = \bar{x}_i\gamma + v_i, \tag{5}$$

where  $\bar{x}_i$  is the vector of the averages of  $x_{it}$  over time,  $\gamma$  is the parameter vector, and  $v_i$  is an orthogonal error with  $v_i|\bar{x}_i \sim N(0, \sigma_v^2)$ . Replacing  $\alpha_i$  in equation (4) with equation (5), we obtain:

$$P(E_{it} = j|x_{it}, \bar{x}_i, v_i; \gamma, \beta) = \Phi(k_j - v_i - x'_{it}\beta_j - \bar{x}'_i\gamma) - \Phi(k_{j-1} - v_i - x'_{it}\beta_{j-1} - \bar{x}'_i\gamma) \text{ for } j = 0, 1, 2, \dots, J. \tag{6}$$

The resulting modification (6) is the so-called ‘correlated random-effects generalized ordered probit model’ or ‘random-effects generalized ordered probit with Mundlak transformation’. The two names are used interchangeably hereinafter. The joint distribution of  $E_i = (E_{i1}, \dots, E_{iT})$  can then be obtained by integrating equation (6) over  $v_i$ , formally:

$$f(E_i|x_{it}, \bar{x}_i; \gamma, \beta_j, \sigma_v^2) = \int_{-\infty}^{\infty} \prod_{t=1}^{T_i} \prod_{j=0}^J P(E_{it} = j|x_{it}, \bar{x}_i, v_i; \gamma, \beta_j)^{\mathbf{1}(E_{it}=j)} \frac{1}{\sigma_v} \phi\left(\frac{v_i}{\sigma_v}\right) dv_i, \tag{7}$$

where  $f(\cdot)$  is the joint distribution function and  $\mathbf{1}(\cdot)$  is the indicator function. The integral in equation (7) does not have a closed form solution; however, it can be numerically approximated by the Gauss–Hermite quadrature method and the parameters can then be estimated by maximum likelihood (Boes and Winkelmann, 2010).

**4. Data**

The longitudinal data used in this study come from the CHNS, which is one of the most widely used surveys for micro-level research in China. The CHNS was designed as a time-cohort survey. Up to now, eight waves of CHNS data have been collected (for the years 1989, 1991, 1993, 1997, 2000, 2004, 2006 and 2009). The survey employed a multi-stage, random cluster design to draw a sample of households, covering both rural and urban areas of nine Chinese provinces that vary substantially in socio-economic indicators.<sup>5</sup> Because this study concentrates on urban China, the rural households have been dropped from our sample.

In the CHNS survey, households are asked what kind of fuel(s) they normally use for cooking. Although households may not rely on just one type

(which is a modified version of Das and Van Soest’s estimator, see Das and van Soest, 1999).

<sup>5</sup> The provinces are Liaoning, Heilongjiang, Jiangsu, Shandong, Henan, Hubei, Hunan, Guangxi and Guizhou. Further details on the survey design are available at <http://www.cpc.unc.edu/projects/china>.



of cooking fuel, this study focused on the choice of primary cooking fuel, which is the fuel most often used, as stated in each household's response to the survey. Firewood (including wood, sticks, straw, etc.), coal and LNG are found to be three of the most commonly used primary cooking fuels among the urban households and they account for up to 78 per cent of the pooled sample. Thus, we focus on the choice among these three cooking fuels. Consequently, our dependent variable – household cooking fuel choice – in the ordered probit model would be choice of firewood, coal and LNG, which are in order of efficiency, convenience to use and modernity. Therefore, we assign  $E = 0$  for firewood,  $E = 1$  for coal, and  $E = 2$  for LNG as the dependent variable. This is similar to the order in *Farsi et al. (2007)*, which considered the cooking fuels in urban India in the order of firewood, kerosene and LPG.

Given that our study focused on analyzing the determinants of primary cooking energy choice, we eliminated the observations for which important variables (including household income, energy price data, characteristics of household head, etc.) were not available. This produced our final sample of 3,859 pooled observations, with a total number of 1,640 households in this sample.<sup>6</sup>

From this sample, we find that the percentage of households using LNG as their primary cooking fuel increased dramatically over time, while the percentage of households using firewood or coal shows a clear tendency of decreasing. This implies households' transition toward more efficient sources of primary cooking energy in urban China. For instance, the data indicate that, in 1989, 75.5 per cent of urban households used coal as their primary cooking fuel. This figure decreased dramatically to 19.4 per cent in 2009. At the same time, the proportion of households using LNG as their primary cooking fuel increased from 13.2 per cent in 1989 to 44.7 per cent in 1997 and rose further to 71.2 per cent in 2009. These facts seem to suggest a tendency of switching in households' preference of primary cooking fuels to more efficient energy.

We studied the effect of household income because that has been found in the literature to be an important determinant of household fuel choice. The household income used in this study is already inflated to the year 2009, using Consumer Price Indices, to make income comparable over different time periods (waves) and is included in the CHNS data set. The relationship between primary cooking fuel choices and household income level is presented in figure 1. It can be seen that, as household income increases, people are more likely to choose the fuels that are ranked higher in terms of efficiency or modernity, i.e., LNG, as their primary cooking fuels.

<sup>6</sup> Specifically, 3,661 observations were dropped due to missing energy price data (coal or LNG prices), and 216 further observations were dropped due to missing values of other important control variables including household income, household head characteristics (gender, marriage status, occupation, education) and household size. Also, it can be noted that households may appear in some of the waves only. This could raise the issue of attrition which, however, is not found systematic according to a standard test on the probit regression of attrition. The details can be obtained from the authors upon request.

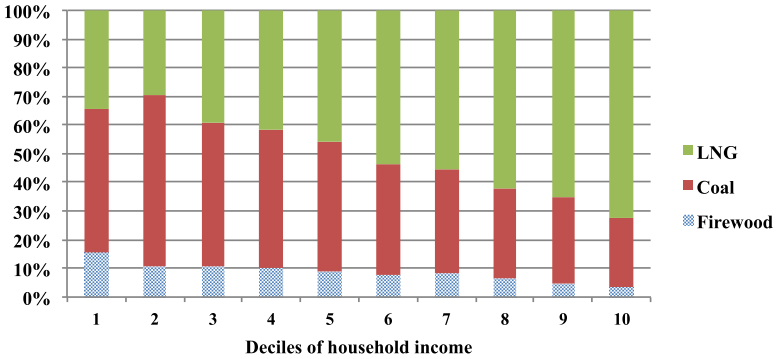


Figure 1. Primary cooking fuel by deciles of household income

Table 1 presents the descriptive statistics for socio-economic variables by cooking fuel choices. It can be observed from table 1 that some trends exist in households' choice of primary cooking fuel. As well as household income, gender, education and job characteristics of household heads are associated with distinct differences between households who choose firewood as their primary cooking fuel and those who choose LNG as their primary cooking fuel. For instance, the average income of the households who choose LNG as their primary cooking fuel is around 80 per cent higher than of those that prefer firewood. Further, the proportion of female household heads in the 'LNG' category is larger than in the 'firewood' category or in the 'coal' category.

Three dummy variables were used to represent the highest education level attained by household heads: primary school degree, secondary school degree, and university (or higher) degree. As shown in table 1, only 0.3 per cent of the heads of the households choosing firewood as their primary cooking fuel have a university or higher degree, while the figure is about 9.1 per cent for those choosing LNG. This seems to suggest that better educated people tend to choose more efficient fuels.

Table 1. Means of major variables by primary cooking fuel choice

	Firewood <i>E</i> = 0	Coal <i>E</i> = 1	LNG <i>E</i> = 2
Household income	16, 367.54	19, 196.87	28, 917.86
Birth year of household head	1948.349	1944.009	1948.494
HHs with a female head	0.164	0.256	0.273
HH head married	0.857	0.836	0.850
HH head employed in public sector	0.104	0.343	0.342
HH head with primary education	0.218	0.217	0.159
HH head with secondary education	0.406	0.394	0.557
HH head with university (or higher) education	0.003	0.020	0.091
HH size (number of persons in HH)	3.979	3.596	3.248

Table 2. *Descriptive statistics*

<i>Variable</i>	<i>Mean</i>	<i>SD</i>	<i>Minimum</i>	<i>Maximum</i>
Household income	23,797.900	27,286.500	5.160	425,816.500
Coal price	0.539	0.747	0.035	4.674
LNG price	79.408	18.833	26.275	127.800
Birth year of household head	1946.622	14.241	1898	1988
HHs with a female head	0.257	0.437	0	1
HH head married	0.845	0.362	0	1
HH head employed in public sector	0.322	0.467	0	1
HH head with primary education	0.188	0.391	0	1
HH head with secondary education	0.476	0.499	0	1
HH head with university (or higher) education	0.054	0.226	0	1
HH size (number of persons in HH)	3.455	1.399	1	12

The job characteristic of a household head is represented by a dummy variable indicating whether the head was employed in the public sector<sup>7</sup> at the time of the survey. It can be observed from table 1 that there is a possible relationship between public sector employment and household cooking fuel choice. In terms of household size, the descriptive statistics show that the households choosing LNG as their primary cooking fuel tend to be smaller on average, compared with those choosing firewood or coal as their primary cooking fuel.

In the literature, fuel prices have also been found to be potential determinants of household cooking fuel choices. Because the prices of firewood (wood, straw, etc.) are not available in the survey, only the prices of coal and LNG are considered in the final analysis. Moreover, the fuel prices collected in the survey are at the community level, which implies that all the households in one community face the same price.<sup>8</sup> In addition, we inflate the prices of coal and LNG in the data set to the year 2009 (by using the community-level inflation indexes which are included the data set) to make them comparable over different time periods (waves). Furthermore, it is likely that income grows with experience (age) and therefore the age variable may capture some of the income effect in the regressions. To avoid

<sup>7</sup> The public sector here includes government departments, state service/institutes, and state-owned enterprises.

<sup>8</sup> As stated in the CHNS community survey questionnaire, the prices are obtained by 'asking the community leader or appropriate sellers in the shopping malls/free market'.

this, we replace the age variable with the birth year of household head.<sup>9</sup> The variables used in the econometric analysis that follows are listed in table 2, where the descriptive statistics are for the entire sample.

## 5. Empirical results

Results from the maximum likelihood estimation for random effects generalized ordered probit (RE) and correlated random effects generalized ordered probit (CORE) models<sup>10</sup> are presented in table 3. To control for the time trend of the panel data and regional differences in fuel choices, we include wave and area (province) dummies in all the regressions. From table 3, it can be seen that, in the generalized models, two parameter-vectors ( $\beta_1$  and  $\beta_2$ ) are estimated. The parameter vector  $\beta_1$  refers to estimated coefficients of the determinants for the coal category compared to the base category (firewood). Vector  $\beta_2$  is for LNG. The explanatory variables used in all these regressions are displayed in table 2. The marginal effects of the variables are presented in table 4.

A Wald test on the generalized ordered probit models against the standard models suggests that we can reject the parallel regression assumption ( $\chi^2_{25} = 523.95$  ( $p$ -value = 0.000) for RE and  $\chi^2_{25} = 528.00$  ( $p$ -value = 0.000) for CORE).<sup>11</sup>

As mentioned above, the potential unobserved individual heterogeneity generally leads to inefficient and inconsistent estimation results in the cross-sectional models, which can be addressed by panel data models.<sup>12</sup> As shown in table 3, the estimated correlation coefficient of the composite error term,  $\rho$  ( $\rho = 0.457$ ) in the RE model is statistically significant, which implies significant unobserved heterogeneity in the households' fuel choices and the need to control for this unobserved heterogeneity.

While RE accounts for individual unobserved heterogeneity, its implicit assumption of no correlation between unobserved heterogeneity and the explanatory variables is often unlikely to hold. For example, a household head who is born with persistent personal motivation may work more hours and earn money (income) that enables him/her to purchase more efficient, modern and convenient (to use) fuels. Following Boes and Winkelmann (2010), Chamberlain (1980) and Mundlak (1978), it is possible to control the bias from the possible correlation between the unobserved

<sup>9</sup> Our thanks to an anonymous reviewer for pointing this out and inspiring us to do such a transformation.

<sup>10</sup> The RE model and the CORE model were estimated by the maximum likelihood estimation using the Stata command 'regoprobit'.

<sup>11</sup> In addition, following Pfarr (2010), a Wald test is applied on each variable to identify which variables have heterogeneous distributional impacts. The null hypothesis of equal coefficients can be rejected in 20 out of the 26 variables (including year and province dummies) at the 10 per cent level of significance.

<sup>12</sup> The results of standard ordered probit and cross-sectional generalized ordered probit models are omitted here to save space (following the reviewers' suggestions). They are available from the authors upon request.

Table 3. *Estimation results of random effects generalized ordered probit (RE) and correlated random effects generalized ordered probit (CORE)*

Variables	RE		CORE	
	$\beta_1$	$\beta_2$	$\beta_1$	$\beta_2$
Household income (log)	0.362*** (0.048)	0.210*** (0.034)	0.196*** (0.061)	0.055 (0.042)
Coal price (log)	-0.199*** (0.072)	0.172*** (0.041)	-0.211*** (0.074)	0.174*** (0.042)
LNG price (log)	-0.376 (0.344)	-0.279 (0.189)	-0.351 (0.354)	-0.191 (0.193)
Year of birth	-0.013*** (0.004)	-0.006* (0.003)	-0.012*** (0.005)	-0.007** (0.003)
HHs with a female head	0.685*** (0.143)	0.469*** (0.093)	0.603*** (0.148)	0.436*** (0.094)
HH head married	0.013 (0.157)	-0.019 (0.107)	-0.102 (0.162)	-0.061 (0.109)
HH head employed in public sector	0.804*** (0.138)	0.628*** (0.087)	0.211 (0.176)	0.126 (0.111)
HH head with primary education	0.138 (0.142)	0.150 (0.105)	0.069 (0.211)	-0.069 (0.155)
HH head with secondary education	0.655*** (0.141)	0.666*** (0.103)	0.699*** (0.255)	0.260 (0.165)
HH head with university (or higher) education	1.892*** (0.554)	1.405*** (0.200)	0.126 (1.323)	1.129*** (0.394)
HH size (number of persons in HH)	-0.151*** (0.036)	-0.065** (0.027)	0.069 (0.055)	0.042 (0.038)
.Iwave_1991	-0.192 (0.284)	0.757*** (0.177)	-0.192 (0.290)	0.804*** (0.180)
.Iwave_1993	0.166 (0.226)	1.143*** (0.181)	0.260 (0.232)	1.163*** (0.184)
.Iwave_1997	-0.230 (0.253)	1.806*** (0.190)	-0.166 (0.261)	1.851*** (0.195)
.Iwave_2000	0.431** (0.219)	2.020*** (0.183)	0.529** (0.227)	2.049*** (0.187)
.Iwave_2004	0.944*** (0.242)	2.061*** (0.193)	1.042*** (0.250)	2.064*** (0.197)
.Iwave_2006	0.983*** (0.238)	2.362*** (0.202)	1.101*** (0.243)	2.378*** (0.205)
.Iwave_2009	0.746*** (0.235)	2.535*** (0.208)	0.874*** (0.240)	2.605*** (0.213)
Constant	23.549*** (8.395)	8.589 (6.094)	19.093** (8.757)	8.773 (6.213)
Province dummies	YES		YES	
Rho	0.457*** (0.032)		0.448*** (0.033)	
Overall Wald test for parallel regression assumption: $\chi^2$ (p-value)	523.95(0.000)		528.00(0.000)	

(continued)

Table 3. Continued

Variables	RE		CORE	
	$\beta_1$	$\beta_2$	$\beta_1$	$\beta_2$
Joint test for the mean of time-varying variables: $\chi^2$ (p-value)			133.88(0.000)	
Observations	3,859	3,859	3,859	3,859

Notes: Standard errors in parentheses. \*\*\*  $p < 0.01$ ; \*\*  $p < 0.05$ ; \*  $p < 0.1$ .

heterogeneity and the explanatory variables by including the averages of time-varying regressors in the model. According to Ferrer-i-Carbonell (2005), this Mundlak transformation approach will yield similar results to the standard fixed effects approaches that factor out the fixed effects from the estimation.

Results of the RE with Mundlak transformation or the CORE are presented in table 3 (columns 3 and 4). As shown in table 3, the mean values of time-varying covariates are jointly significant, implying the relevance of the Mundlak approach in controlling the bias from unobserved heterogeneity. In addition, comparing the estimated coefficients of RE and CORE, we can observe the differences in size and significance for the following variables: income, public sector employment, all education dummies and household size. Considering the importance of allowing the correlation between the unobserved heterogeneity and explanatory variables, we refer mainly to the CORE results and the corresponding marginal effects in our discussion that follows.

From the CORE regression results, it is evident that household economic status, head's characteristics, fuel prices and year (wave) dummy variables play an important role in determining the household's primary cooking fuel choice in urban China. Beginning our analysis with the economic status of the household, it can be seen from table 3 that this variable is positive in the two parameter vectors ( $\beta_1$  and  $\beta_2$ ) of the CORE columns. Likewise, the marginal effect of household income is positive for the probability of choosing LNG, when evaluated at the sample mean. This positive estimate conveys the message that, as the income of households rises, they prefer to use LNG rather than firewood as their primary cooking fuels, although the marginal effect estimate is not statistically significant. Nevertheless, as can be seen from table 4, the marginal probability effect of income on the choice of firewood is negative and statistically significant. In general, a one-unit increase in household income (log) will on average decrease the probability of choosing firewood by 0.6 per cent. Compared with the results from CORE, the marginal effect results from RE also suggest the negative effects of income on the choice of firewood. Moreover, the negative effect of income on the choice of coal and the positive effect of income on LNG is more evident in RE, as can be seen in table 4. The general finding that households with higher income are less likely to choose

Table 4. Marginal effects from random effects generalized ordered probit model and correlated random effects generalized ordered probit model

Variables	RE – marginal effects			CORE – marginal effects		
	Firewood	Coal	LNG	Firewood	Coal	LNG
Household income (log)	-0.013*** (0.002)	-0.049*** (0.010)	0.062*** (0.010)	-0.006*** (0.002)	-0.010 (0.012)	0.016 (0.012)
Coal price (log)	0.007*** (0.003)	-0.058*** (0.012)	0.051*** (0.012)	0.006*** (0.002)	-0.058*** (0.012)	0.052*** (0.012)
LNG price (log)	0.014 (0.012)	0.068 (0.054)	-0.082 (0.056)	0.010 (0.010)	0.046 (0.056)	-0.056 (0.057)
Year of birth	0.000*** (0.000)	0.001 (0.001)	-0.002* (0.001)	0.000** (0.000)	0.002* (0.001)	-0.002** (0.001)
HHs with a female head	-0.020*** (0.004)	-0.117*** (0.026)	0.137*** (0.026)	-0.014*** (0.004)	-0.114*** (0.026)	0.128*** (0.027)
HH head married	-0.000 (0.006)	0.006 (0.030)	-0.006 (0.032)	0.003 (0.004)	0.015 (0.031)	-0.018 (0.032)
HH head employed in public sector	-0.025*** (0.005)	-0.158*** (0.025)	0.183*** (0.025)	-0.006 (0.005)	-0.031 (0.032)	0.037 (0.033)
HH head with primary education	-0.005 (0.005)	-0.039 (0.030)	0.044 (0.031)	-0.002 (0.006)	0.022 (0.045)	-0.021 (0.046)
HH head with secondary education	-0.024*** (0.006)	-0.170*** (0.028)	0.194*** (0.029)	-0.021** (0.008)	-0.056 (0.047)	0.077 (0.049)
HH head with university (or higher) education	-0.024*** (0.004)	-0.336*** (0.037)	0.361*** (0.037)	-0.003 (0.032)	-0.303*** (0.092)	0.307*** (0.088)
HH size (number of persons in HH)	0.005*** (0.001)	0.014* (0.008)	-0.019** (0.008)	-0.002 (0.002)	-0.010 (0.011)	0.012 (0.011)
Iwave_1991	0.008 (0.013)	-0.222*** (0.047)	0.215*** (0.046)	0.006 (0.011)	-0.235*** (0.047)	0.229*** (0.047)
Iwave_1993	-0.005 (0.007)	-0.305*** (0.041)	0.311*** (0.041)	-0.006 (0.005)	-0.312*** (0.041)	0.318*** (0.042)
Iwave_1997	0.010 (0.012)	-0.456*** (0.033)	0.446*** (0.032)	0.005 (0.009)	-0.464*** (0.032)	0.458*** (0.032)
Iwave_2000	-0.013** (0.005)	-0.478*** (0.029)	0.491*** (0.029)	-0.012*** (0.004)	-0.489*** (0.029)	0.501*** (0.029)
Iwave_2004	-0.023*** (0.005)	-0.478*** (0.031)	0.501*** (0.031)	-0.019*** (0.004)	-0.488*** (0.031)	0.507*** (0.032)
Iwave_2006	-0.023*** (0.005)	-0.522*** (0.027)	0.545*** (0.027)	-0.020*** (0.004)	-0.533*** (0.028)	0.553*** (0.028)
Iwave_2009	-0.018*** (0.004)	-0.520*** (0.022)	0.538*** (0.022)	-0.016*** (0.004)	-0.535*** (0.021)	0.550*** (0.022)
Province dummies		YES			YES	
Observations	3,859	3,859	3,859	3,859	3,859	3,859

Notes: Standard errors in parentheses. \*\*\*  $p < 0.01$ ; \*\*  $p < 0.05$ ; \*  $p < 0.1$ .

the low-ranked energy sources is consistent with the findings from earlier studies (Heltberg, 2004, 2005; Farsi *et al.*, 2007; Mekonnen and Köhlin, 2008).

Consistent with previous studies (Heltberg, 2004, 2005; Farsi *et al.*, 2007; Mekonnen and Köhlin, 2008; Gebreegziabher *et al.*, 2012), fuel prices are also found to be important in determining cooking energy choices in urban China. As we can see from the results of CORE in table 3 and its corresponding marginal effect (table 4), a higher coal price decreases the probability of choosing coal as the primary cooking energy and increases

the probability of choosing firewood and LNG. More specifically, a one-unit increase in the coal price (log) will decrease the average share of households choosing coal as their primary cooking fuel by 5.8 per cent, while increasing the share of those choosing firewood and LNG by 0.6 per cent and 5.2 per cent, respectively. We can see similar results from the RE marginal effect estimate of coal price.

An increase in LNG price, on the other hand, is not found to have a significant effect on households' cooking fuel choice. It can be observed from the marginal effects (table 4) that the effect of (higher) LNG price on choosing LNG is negative but statistically insignificant at the conventional level in both generalized models (RE and CORE). The estimated results for the fuel price variables imply that the policies that increase the prices of dirty fuels (e.g., taxing coal) can have a positive effect in helping the transition to cleaner energy.

In many developing countries, it is more common to see women than men cooking, and hence they are more likely to be exposed to the health hazards of indoor air pollution from using dirty fuel sources. Our expectation is that, compared to the male-headed households, the decision makers in female-headed households better understand the health risks and inconveniences of cooking with unclean fuel sources. Consistent with our expectation, all four models suggest that female-headed households are less likely to choose firewood or coal as their primary cooking fuel, and more likely to choose LNG. Referring to the marginal effect results of the CORE model, female-headed households are, on average, 12.8 per cent more likely to choose LNG as their primary cooking fuel, and 11.4 per cent less likely to choose coal, than are households with male heads. This implies that greater empowerment of women in the household can be helpful in increasing the usage of cleaner household energy in China.

In addition, the job characteristics of a household head may also affect the household's preference for cooking fuel choices. The estimated coefficient and marginal effect of having a public sector employed head in the CORE model is positive for choosing LNG; however, it is found statistically insignificant in this model while significant in the RE model. This difference may be due to the control of the correlation between the explanatory variables and the time-invariant unobserved heterogeneity in the CORE model.

Education is an important policy tool to raise households' awareness about the benefits of clean energy sources and the risks of dirty fuel sources. This implies that a household head with a higher education level is expected to be more likely to choose clean energy sources. In this study, we use three dummy variables to represent the highest education attained by household heads: primary school degree, secondary school degree, and university (or higher) degree. Our result indicates lower education levels (primary) are insignificant in the determinant of fuel choices. However, it can be observed that household heads having a university or higher degree are more likely to choose higher ranked energy sources than those who do not have a university degree. From the marginal effect results of CORE displayed in table 4, it can be seen that household heads having a university



degree or higher are 30.7 per cent more likely to choose LNG, compared with those without a university degree. This effect of education is consistent with earlier studies on fuel demand (Heltberg, 2004, 2005; Jiang and O'Neill, 2004; Farsi *et al.*, 2007; Mekonnen and Köhlin, 2008).

Previous studies find mixed results on the effect of household size on fuel choice and fuel switching (Heltberg, 2004, 2005; Ouedraogo 2006; Farsi *et al.* 2007). For example, Ouedraogo (2006) suggests that, in urban Burkina Faso, households with fewer members are more likely to adopt LPG and less likely to use firewood for cooking. However, Heltberg (2004) found insignificant effect of household size on fuel switching. In this study, we find that the effect of household size is positive for LNG but insignificant (see CORE columns in table 4), which is consistent with Heltberg (2004). However, according to the marginal effect results from RE (table 4), increasing the household size will be associated with a negative effect on the probability of choosing LNG (significant at the 5 per cent level). Again, the difference between the two models suggests the importance of controlling the potential correlation between the explanatory variables and the time-invariant unobserved heterogeneity.

As can be seen from table 4, the marginal effect of the wave dummies shows that more and more people are shifting from firewood and coal toward LNG. These wave dummies account for the effects of policy changes or other phenomena over time (other than the change in socio-economic characteristics) which could make households shift their energy choice. For example, the shift to LNG may be associated with an increased access to LNG over time. Yang *et al.* (2014) documented an increasing trend of the LNG pipeline networks in urban China since the year 1998. Furthermore, the reduction in the usage of firewood in urban China over time may also be related to the introduction of more restrictive forest policies such as the Natural Forest Conservation Program (NFCP) from 1998 onward (Zhang, 2000), which stipulates the protection of existing natural forests from excessive cutting, thereby reducing the supply of firewood.

Ordered probit models assume the ordering of different fuels in terms of efficiency and convenience to use. As discussed in the empirical strategy section (see footnote 3), it can be argued that the households' real ranking of the fuels may not be obvious and hence the multinomial logit model may be used alternatively. Nonetheless, the generalized ordered probit/logit model is preferable in cases where ordering is not 'a priori obvious' (Anderson, 1984) and the multinomial logit model can result in inefficient estimates if the ordering is inherent in the household fuel preference (Boes and Winkelmann, 2010). Therefore, we stick to the generalized ordered models (which are also better justified with the previous literature) in this paper. Yet, for comparison purposes, we also provide the marginal effects from the multinomial logit model in the appendix (table A1).<sup>13</sup> It can

<sup>13</sup> It should be mentioned that considering the random effects in the multinomial logit model would lead to computation difficulties in the estimation. However, an approximation assuming random effects are equal to their means, i.e., zeros, seems satisfactory from a policy point of view. This reduces the model to a simple

be seen from table A1 that the multinomial logit model generally supports our previous conclusions on the positive effects of raising households' income, increasing coal price, empowering women, and higher education in helping the transition to clean energy. Meanwhile, the results from the multinomial model also suggest the importance of reducing LNG price (e.g., subsidizing LNG) in encouraging the usage of clean energy, which, however, is not well supported by the (RE and CORE) generalized ordered models above.

## 6. Conclusion and policy implications

Households' transition to modern energy sources can reduce the health and environmental impacts caused by the usage of traditional energy. Understanding the determinants of household fuel choice can provide policy implications for encouraging the adoption of cleaner and more efficient fuels in households. As the largest developing country in the world, China's evidence on this issue is of great interest. In contrast to previous studies on China, most of which are based on aggregate statistics or cross-sectional data from household surveys in a certain province or county, this paper employed the panel data from a nationwide survey (CHNS) to study the determinants of household fuel choice in urban China. Ordered probit models were employed in this study to take into account the potential ordering of different fuels in terms of efficiency or convenience to use, as in *Farsi et al. (2007)*. Moreover, as mentioned above, we also made extensions to the applications of ordered discrete choice models to household fuel choice in terms of relaxing the parallel regression assumption and controlling the potential correlation between the explanatory variables and unobserved heterogeneity.

Our results indicate the heterogeneous effects of the explanatory variables across the distribution of different cooking fuels, which supports the use of the generalized ordered probit model. Also, the RE with Mundlak transformation approach generates results that are significantly different from those that are based on the standard random effects methods.

Furthermore, the results indicate that higher income leads to a lower probability of choosing low-ranked cooking fuel sources. Meanwhile, the results also show that, in addition to income, sociodemographic factors such as gender and education of the household heads are also important in determining the choices of primary cooking fuels in urban Chinese households. Thus, consistent with other recent studies (*Masera et al., 2000; Heltberg, 2005; Farsi et al., 2007*), our results suggest that household fuel choice is not determined merely by a household's economic condition.

Coal price was also found to be important in household choices of primary cooking energy. An increase in the coal price is associated with a statistically significant decrease in the probability of choosing coal but a significant increase in the probability of choosing LNG or firewood as

multinomial logit model. We thank an anonymous reviewer for this suggestion which inspired us to do such an approximation.

the primary cooking fuel. However, the effect of LNG price is found to be insignificant. From a policy point of view, these results indicate that interventions that reduce the price advantage of dirty fuels (e.g., taxing coal) may encourage households to use cleaner energy as their primary cooking fuel.

The tendency of households with female heads to be more likely to choose LNG as the primary cooking fuel and to reduce the usage of firewood implies that greater empowerment of women in the household can be helpful in increasing the usage of clean energy in urban China. In addition, the results that more education for household heads increases the probability of choosing LNG suggest that promotion of higher levels of education can be an effective way to encourage households to choose clean energy as the primary cooking fuel.

The estimated results of the year dummies indicates that more and more people are shifting over time from firewood and coal to LNG. The shift to LNG may also be associated with increased access to LNG over time, such as an increase in pipeline networks in urban China. In addition, the reduction in the usage of firewood in urban China over time may also be related to the introduction of more restrictive forest policies such as the NFCP.

However, this paper is not without limitations. For example, due to lack of information on the proportion of each fuel for the households who use multiple fuels, this study focused on the choice of primary cooking fuels and did not analyze multiple fuel use. Therefore, a direction for future research can be more comprehensive modeling of households' decision making on cooking fuels with consideration of all the fuels that a household can use. This will require richer data on households' cooking fuel choices.

## References

- Alam, M., J. Dunkerley, and A.K. Reddy (1985), 'Firewood use in the cities of the developing countries: two case studies from India', *Natural Resources Forum* **9**(1): 205–213.
- Anderson, J.A. (1984), 'Regression and ordered categorical variables', *Journal of the Royal Statistical Society Series B (Methodological)* **46**(1): 1–30.
- Baetschmann, G., K. Staub, and R. Winkelmann (2011), 'Consistent estimation of the fixed effects ordered logit model', IZA Discussion Paper No. 5443, Institute for the Study of Labor (IZA), Bonn.
- Barnett, A. (2000), 'Energy and the fight against poverty', *Livelihood Sector Report*, Department for International Development (Dfid), London.
- Boes, S. and R. Winkelmann (2006), 'Ordered response models', *Allgemeines Statistisches Archiv* **90**: 165–179.
- Boes, S. and R. Winkelmann (2010), 'The effect of income on general life satisfaction and dissatisfaction', *Social Indicators Research* **95**: 111–128.
- Booth, A.L. and J.C. van Ours (2008), 'Job satisfaction and family happiness: the part-time work puzzle', *Economic Journal* **118**: F77–F99.
- Cai, J. and Z. Jiang (2008), 'Changing of energy consumption patterns from rural households to urban households in China: an example from Shaanxi Province, China', *Renewable and Sustainable Energy Reviews* **12**(3): 1667–1680.

- Chamberlain, G. (1980), 'Multivariate regression models for panel data', *Journal of Econometrics* **18**: 5–45.
- Chen, L., N. Heerink, and M. van den Berg (2006), 'Energy consumption in rural China: a household model for three villages in Jiangxi Province', *Ecological Economics* **58**(2): 407–420.
- Das, M. and A. van Soest (1999), 'A panel data model for subjective information on household income growth', *Journal of Economic Behaviour and Organization* **40**: 409–426.
- ESMAP (1996), *Energy for Rural Development in China: An Assessment Based on a Joint Chinese/ESMAP Study in Six Counties*, Report No. 183/96, Washington, DC: World Bank, Energy Sector Management Assistance Programme.
- Farsi, M., M. Filippini, and S. Pachauri (2007), 'Fuel choices in urban Indian households', *Environmental and Development Economics* **12**(3): 757–774.
- Ferrer-i-Carbonell, A. (2005), 'Income and well-being: an empirical analysis of the comparison income effects', *Journal of Public Economics* **89**: 997–1019.
- Ferrer-i-Carbonell, A. and P. Frijters (2004), 'How important is methodology for the estimates of the determinants of happiness?', *Economic Journal* **114**: 641–659.
- Gebreegziabher, Z., A. Mekonnen, M. Kassie, and G. Köhlin (2012), 'Urban energy transition and technology adoption: the case of Tigray, northern Ethiopia', *Energy Economics* **34**(2): 410–418.
- Gupta, G. and G. Köhlin (2006), 'Preferences for domestic fuel: analysis with socio-economic factors and rankings in Kolkata, India', *Ecological Economics* **57**(1): 107–121.
- Heltberg, R. (2004), 'Fuel switching: evidence from eight developing countries', *Energy Economics* **26**(2): 869–887.
- Heltberg, R. (2005), 'Factors determining household fuel choice in Guatemala', *Environmental and Development Economics* **10**(1): 337–361.
- Hosier, R. and J. Dowd (1987), 'Household fuel choice in Zimbabwe – an empirical test of the energy ladder hypothesis', *Resources and Energy* **9**(4): 347–361.
- IEA (International Energy Agency) (2011), *World Energy Outlook 2011*, Paris: OECD.
- Jiang, L. and B.C. O'Neill (2004), 'The energy transition in rural China', *International Journal of Global Energy Issues* **21**(1–2): 2–26.
- Kassenboehmer, S.C. and J.P. Haisken-DeNew (2009), 'You're fired! The causal negative effect of unemployment on life satisfaction', *Economic Journal* **119**: 448–462.
- Leach, G. (1992), 'The energy transition', *Energy Policy* **20**(2): 116–123.
- Maddala, G.S. (1983), *Limited-Dependent and Qualitative Variables in Economics*, New York: Cambridge University Press.
- Masera, O., B. Saatkamp, and D. Kammen (2000), 'From linear fuel switching to multiple cooking strategies: a critique and alternative to the energy ladder model', *World Development* **28**(12): 2083–2103.
- Mekonnen, A. and G. Köhlin (2008), 'Determinants of household fuel choice in major cities in Ethiopia', Discussion Paper Series No. EFD DP 08-18, Environment for Development, Gothenburg.
- Mensah, J.T. and G. Adu (2013), 'An empirical analysis of household energy choice in Ghana', Working Paper, Department of Economics, Swedish University of Agricultural Sciences, Uppsala.
- Mundlak, Y. (1978), 'On the pooling of time series and cross section data', *Econometrica* **46**: 69–85.
- Nlom, J.H. and A. Karimov (2014), 'Modeling fuel choice among households in northern Cameroon', Working Paper No. 2014/038, United Nations University, Tokyo.
- Ouedraogo, B. (2006), 'Household energy preferences for cooking in urban Ouagadougou, Burkina Faso', *Energy Policy* **34**(18): 3787–3795.

- Pachauri, S. and L. Jiang (2008), 'The household energy transition in India and China', *Energy Policy* 36(11): 4022–4035.
- Peng, W., Z. Hisham, and J. Pan (2010), 'Households level fuel switching in rural Hubei', *Energy for Sustainable Development* 14(1): 238–244.
- Pfarr, C., A. Schmid, and U. Schneider (2010), 'Estimating ordered categorical variables using panel data: a generalized ordered probit model with an autofit procedure', Discussion Paper in Economics No. 02/10, Department of Law and Economics, University of Bayreuth.
- Reddy, B.S. (1995), 'A multilogit model for fuel shifts in the domestic sector', *Energy* 20(9): 929–936.
- Sathaye, J. and S. Tyler (1991), 'Transitions in household energy use in urban China, India, the Philippines, Thailand, and Hong Kong', *Annual Review of Energy and the Environment* 16: 295–335.
- Smith, K.R., M.G. Apte, Y. Ma, W. Wathana, and A. Kulkarni (1994), 'Air pollution and the energy ladder in Asian cities', *Energy* 19(2): 587–600.
- Smith, K.R., S. Mehta, and M. Maeusezahl (2004), 'Indoor air pollution from household use of solid fuels', in M. Ezzati *et al.* (eds), *Comparative Quantification of Health Risks: Global and Regional Burden of Disease Attributable to Selected Major Risk Factors*, Geneva: World Health Organization.
- Terza, J. (1985), 'Ordinal probit: a generalization', *Communications in Statistics – Theory and Methods* 14: 1–11.
- Wang, X. and Z. Feng (1997), 'Rural household energy consumption in Yangzhong County of Jiangsu Province in China', *Energy* 22(12): 1159–1162.
- WHO (2006), *Fuel for Life: Household Energy and Health*, [Available at] <http://www.who.int/indoorair/publications/fuelforlife/en/>.
- WHO (2008), 'Global estimates of burden of disease caused by the environment and occupational risks', [Available at] [http://www.who.int/quantifying\\_ehimpacts/global/en/](http://www.who.int/quantifying_ehimpacts/global/en/).
- Wooldridge, J. (2010), *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MA: MIT Press.
- Yang, C.J., Y.P. Zhou, and R.B. Jackson (2014), 'China's fuel gas sector: history, current status, and future prospects', *Utilities Policy* 28: 12–21.
- Zhang, J. and K. Kotani (2012), 'The determinants of household energy demand in rural Beijing: can environmentally friendly technologies be effective?', *Energy Economics* 34(2): 381–388.
- Zhang, P., G. Shao, G. Zhao, M.D. Le, G.R. Parker, J.B. Dunning, and Q. Li (2000), 'China's forest policy for the 21st century', *Science* 288: 2135–2136.

## Appendix

Table A1. *Marginal effect from the multinomial logit model*

Variables	Multinomial logit – marginal effects		
	Firewood	Coal	LNG
Household income (log)	–0.012*** (0.003)	–0.063*** (0.012)	0.075*** (0.012)
Coal price (log)	0.004* (0.002)	–0.062*** (0.014)	0.058*** (0.014)

(continued)

Table A1. Continued

Variables	Multinomial logit – marginal effects		
	Firewood	Coal	LNG
LNG price (log)	0.033** (0.014)	0.133** (0.064)	−0.166** (0.065)
Year of birth	0.001*** (0.000)	0.000 (0.001)	−0.001 (0.001)
HHs with a female head	−0.021*** (0.005)	−0.070** (0.030)	0.091*** (0.030)
HH head married	−0.008 (0.009)	0.036 (0.036)	−0.029 (0.036)
HH head employed in public sector	−0.036*** (0.006)	−0.170*** (0.029)	0.206*** (0.029)
HH head with primary education	−0.004 (0.005)	−0.045 (0.032)	0.050 (0.033)
HH head with secondary education	−0.013** (0.007)	−0.178*** (0.032)	0.191*** (0.032)
HH head with university (or higher) education	−0.027*** (0.005)	−0.308*** (0.038)	0.334*** (0.038)
HH size (number of persons in HH)	0.005*** (0.002)	0.022** (0.009)	−0.028*** (0.009)
Iwave_1991	−0.018*** (0.006)	−0.177*** (0.038)	0.195*** (0.039)
Iwave_1993	−0.017*** (0.005)	−0.239*** (0.039)	0.256*** (0.041)
Iwave_1997	−0.013** (0.006)	−0.380*** (0.029)	0.394*** (0.031)
Iwave_2000	−0.026*** (0.005)	−0.412*** (0.027)	0.438*** (0.027)
Iwave_2004	−0.031*** (0.006)	−0.430*** (0.028)	0.461*** (0.028)
Iwave_2006	−0.034*** (0.006)	−0.469*** (0.025)	0.503*** (0.025)
Iwave_2009	−0.028*** (0.005)	−0.471*** (0.019)	0.499*** (0.019)
Province dummies		YES	
Observations	3,859	3,859	3,859

Notes: Standard errors in parentheses. \*\*\*  $p < 0.01$ ; \*\*  $p < 0.05$ ; \*  $p < 0.1$ .