Research Institute for Environmental Economics and Management

Discussion Paper Series No.1910

# Subjective Probabilistic Expectations, Indoor Air Pollution, and Health: Evidence from cooking fuel use patterns in India

Mriduchhanda Chattopadhyay, Toshi H. Arimura, Hajime Katayama,

Mari Sakudo, and Hide-Fumi Yokoo

September 2019

WASEDA

# Subjective Probabilistic Expectations, Indoor Air Pollution, and Health: Evidence from cooking fuel use patterns in India


Mriduchhanda Chattopadhyay[a,*], Toshi H. Arimura[b, c], Hajime Katayama[c, d], Mari Sakudo[c, e] and Hide-Fumi Yokoo[c, f]

[a.] Graduate School of Economics, Waseda University, 1-6-1 Nishiwaseda, Shinjuku-ku, Tokyo 169-8050; E-mail: mriduchhanda29@gmail.com

[b.] Faculty of Political Science and Economics, Waseda University, 1–6–1 Nishiwaseda, Shinjuku-ku, Tokyo 169– 8050, Japan; E-mail:toshi.arimura@gmail.com

[c.] Research Institute for Environmental Economics and Management, Waseda University, 1–6–1 Nishiwaseda, Shinjuku-ku, Tokyo 169–8050, Japan

[d.] Faculty of Commerce, Waseda University, 1-6-1 Nishiwaseda, Shinjuku-ku, Tokyo 169-8050, Japan; E-mail: hajikat@gmail.com

[e.] Faculty of Engineering, Information and Systems, University of Tsukuba, Tennodai 1-1-1, Tsukuba, Ibaraki 305-8573, Japan; E-mail: marisakudo@gmail.com

[f.] Graduate School of Economics, Hitotsubashi University, Naka 2-1, Kunitachi, Tokyo 186-8601, Japan; E-mail: hidefumi.yokoo@r.hit-u.ac.jp



## Abstract

An increasing number of empirical studies have investigated the determinants of cooking fuel choice in developing countries, where health risk from indoor air pollution is one of the most important issues. We contribute to this stream of literature by examining individuals' subjective probabilistic expectations about health risks when using different types of fuel and their influence on cooking fuel usage patterns. We also explore how these patterns, in turn, affect health status. Using data collected from 557 rural Indian households, we find that subjective probabilistic expectations of becoming sick from dirty fuel usage have a negative influence on the fraction of days with dirty fuel usage in the household. The results also show that dirty fuel usage degrades the health of the individual. We then examine the effectiveness of information provision regarding the health risks of dirty/clean fuel usage. Our simulation demonstrates that although the provision of information results in statistically significant changes in the households' cooking fuel usage patterns and in the individuals' health status, the changes may be small in size.



**Keywords:** subjective probabilistic expectations, indoor air pollution, cooking fuel usage pattern, health, developing country

**JEL classification code**: I10, Q40, C83


---


[*]corresponding author. Graduate School of Economics, Waseda University, 1-6-1 Nishiwaseda, Shinjuku-ku, Tokyo 169-8050; E-mail: mriduchhanda29@gmail.com

## 1. Introduction

Indoor air pollution (hereafter, IAP), caused mainly by the incomplete combustion of dirty cooking fuels[1], is a global health threat accounting for more than 1.45 million annual premature deaths worldwide (Alem et al., 2016). In particular, developing countries suffer from exposure to IAP, accounting for approximately 3.7% of the loss of disability-adjusted life years (Bonjour et al., 2007). The main victims seem to be women due to their primary responsibility for cooking activities in the household (Pitt et al., 2006).

To lower exposure to IAP in developing countries and thereby prevent the related health hazards, it is important to design effective policies that help to reduce the use of dirty cooking fuel. It thus becomes necessary to understand why a large population of developing countries continue to use dirty cooking fuels, which has motivated researchers to explore the role of different factors on the choice of cooking fuel used in developing countries. For example, previous studies observe that other than economic and demographic factors such as income (Heltberg, 2005) and education (Farsi et al., 2007), energy access significantly affects the cooking fuel choice of households in developing countries (Pachauri and Jiang, 2008). Jeuland et al. (2015) also note that the relative cost advantage and availability of dirty fuels may be drivers in the use of such fuels in developing countries. Gupta and Köhlin (2006) argue that individuals in developing countries tend to choose dirty cooking fuels due to dietary preferences because they believe that food tastes better when cooked with such fuels.

However, despite much evidence in the literature, there remains an unexplored aspect of the choice of household cooking fuel, specifically, the role of expectations about health risks from IAP. In the absence of precise estimates of and/or information about the health risks related to IAP, individuals may have different expectations about these health risks and

[1] Cooking fuels such as firewood, solid biomass fuels and coal are categorized as dirty cooking fuels because of high smoke emission during combustion. Similarly, fuels such as electricity, LPG and kerosene are categorized as clean cooking fuels for low smoke emission.

therefore make different decisions regarding the use of dirty fuel, regardless of personal preferences. Furthermore, researchers typically have no information about individuals' expectations for IAP's health risks. As a result, it is difficult to infer whether preferences or expectations motivate the choice of cooking fuel because different combinations of expectations and preferences can lead to the same observed choice (Delavande, 2014; Manski, 2004). This identification problem limits the ability to devise effective behavioral interventions (Delavande and Kohler, 2015).

Although neglected in studies on the choice of household cooking fuel, the role of expectations in other choice situations has drawn increasing attention in economics. Examples of choices studied include purchases of water treatment products (Brown et al., 2017), multiple sexual partners (Delavande and Kohler, 2015), mental health and labor supply (Baranov et al., 2015), migration (McKenzie et al., 2013) and portfolio allocation (Kézdi and Willis, 2011). These studies confirm that expectations play a certain role in various choice situations.

This study attempts to bridge the existing gap in the literature on the choice of household cooking fuel. In particular, we investigate the impact of individuals' expectations of becoming sick with diseases typically observed from IAP exposure (specifically, dry cough, sore or runny eyes, and difficulty breathing) on their cooking fuel usage pattern and health status. By cooking the fuel usage pattern, we refer to the fraction of days in which dirty fuel is used over a 30-day period. We explore the direct impact of individuals' subjective probabilistic expectations (SPEs, hereafter) of falling sick with such diseases on their cooking fuel usage patterns. Concurrently, we investigate how the individual's cooking fuel usage pattern, in turn, affects his or her probability of suffering from common physical symptoms. This helps us determine the degree to which individuals' expectations about the health risks related to IAP can influence their health status indirectly by influencing their cooking fuel usage pattern.

We analyze a unique dataset on individuals' expectations elicited in probabilistic form from survey respondents in rural India. To elicit the individuals' SPEs about health risks related to IAP, we adopted an interactive elicitation method using visual aids developed by Delavande and Kohler (2009). Our results show that SPEs influence cooking fuel usage patterns, which in turn affect the health status of individuals. Given the significant role played by SPEs in cooking fuel usage patterns, we then explore the effectiveness of providing information on the health risks of dirty/clean fuel usage. Our simulation demonstrates that although the provision of information results in statistically significant changes in households' cooking fuel usage patterns and in individuals' health status, the changes may be small in size.

The remainder of the paper is organized as follows. Section 2 describes the survey methodology and variables considered for the study along with their summary statistics. It also elaborates the methodology used to elicit the SPEs. The next section presents our empirical model and results. Section 4 discusses the policy simulation results, and Section 5 concludes by discussing the directions of future research.

## 2. Data

### 2.1 Survey design

To collect data from rural Indian households on expectations, health, and cooking-related behaviors, we conducted a survey in 17 villages under the Dhapdhapi-II village council in the state of West Bengal, India. Located approximately 40 kilometers from the state capital, Kolkata, the survey site had a population of approximately 14,000 and a population density of 0.7 thousand per square kilometer as of January 2016. Although the survey site does not have a large population size, its population density is high in comparison with the average in India (approximately 0.4 thousand/sq. km). Due to its proximity to Kolkata, the survey site has access to modern amenities but retains the typical traits of a rural area in any developing country.

Out of the total households in those villages, 600 were randomly selected as a sample for our analysis. The enumerators visited the selected households from December 2016 to January 2017. Our respondents are the individuals primarily responsible for cooking in the household; consequently, all of the respondents were female. The survey was conducted via the door-to-door interview method, thus ensuring a high response rate (approximately 99%), and thus the sample size at the end of the first round was 596. From December 2017 to January 2018, the data collectors made second visit to those households surveyed in the first round and were able to elicit responses from 588 out of the 596 original respondents (representing an attrition rate of 1.34%). For our analysis, we exclude from the sample respondents who had no spouse or provided no information about their spouse. This reduces our sample size to 557. Table 1 presents descriptive statistics of the variables used in this study; these will be explained in the next subsection.

## 2.2 Descriptions of the variables and their summary statistics

### Fuel usage pattern

In this study, the fuel usage pattern is a variable of interest and represents the fraction of days dirty fuel was used for cooking. It was elicited by the following survey instrument: "*How many days in the 30 days prior to last month have you used the following kinds of fuels for cooking: electricity, LPG, kerosene, coal/charcoal, solid biomass fuels like animal dung cake and agricultural crop residue, firewood, or any other variety?*". We compute the fraction of days of dirty fuel usage using information on the number of days the respondents used coal/charcoal, solid biomass fuels and firewood. This variable was collected in the second round (2017−18) of the survey for reasons we will explain in the next section.

The fraction is found to be 0.68 on average (see Table 1), suggesting the prevalence of dirty fuel usage in rural areas of India. To better understand this variable, we also draw its distribution in Figure 1. Although a certain portion of households are observed at the end-point of the fuel usage pattern variable, the use of both clean and dirty cooking fuel seems to be common among households. This property of the variable motivated us to model it by using the fractional response variable framework, as will be explained later.

### Self-reported health status

The self-reported health status of the respondents refers to whether the respondent has experienced at least one of three common physical symptoms caused by IAP—dry cough, sore or runny eyes, and difficulty breathing—in the last 30 days. These three minor yet common symptoms were chosen for analysis based on the study by Hanna et al. (2016) in rural India. Indeed, these symptoms are found to be prevalent among the respondents; 76 percent of the respondents experienced at least one of the three symptoms (see Table 1). As with the fuel usage pattern, this variable was collected in the second round (2017−18) of the survey.

### Methodology to elicit SPEs and the elicited SPEs

Partly because the survey targets households in a rural area of India, it was not assumed that the respondents would have a good understanding of the meaning of the word "probability" or of probability concepts. Therefore, to facilitate elicitation of SPEs from the respondents, we used an interactive method with visual aids, which is described in detail by Delavande (2014) and Delavande and Kohler (2015). This method makes it cognitively easier for the respondents to answer questions involving probability concepts than other methods (Brown et al., 2017).

In this elicitation method, we explicitly asked the respondent to link the number of candies placed in front of her to her perceived likelihood of the occurrence of an event. During the elicitation, the enumerators first provided the following instructions to the respondents: *"There are ten candies in front of you. Each candy denotes one chance for the occurrence of* *any event out of 10. To express how likely you think that a specific event will occur, please choose and put aside some candies from the lot. If you are sure that the event will not occur, please do not put any candies aside. If you think the event is more likely to occur, please put more candies. If you think, the event is less likely to occur, please put fewer candies. If you are sure that the event will occur, please put all the candies.".* The respondent's perceived likelihood of an event is then obtained by dividing the number of candies by 10.

Following Godlonton and Thornton (2013), who examine individuals' beliefs regarding the prevalence of HIV, we elicited responses about the SPEs of different health situations for a hypothetical individual. We presume that the respondent expects her likelihood of being sick in the next three months to be dependent only on her current health status and cooking fuel usage; in other words, the health status is assumed to follow a first-order Markov process conditional on cooking fuel usage.

In the survey, we asked for four SPEs about the transition probabilities of health status. The respondents were first asked about two SPEs conditional on dirty fuel usage, denoted as $SP(s_{t+1} = 1|s_t = 1, d)$ and $SP(s_{t+1} = 1|s_t = 0, d)$, where $s$ represents an indicator for being sick and $d$ symbolizes dirty fuel usage. The former (the latter) represents the expected likelihood that the hypothetical individual will remain (become) sick in the next three months given that she is currently sick (not sick) and uses dirty fuels only. By "sick," we refer only to the individuals having suffered from at least one of the three symptoms in the last 30 days—dry cough, sore or runny eyes, and difficulty breathing. Using $SP(s_{t+1} = 1|s_t = 1, d)$, we can compute the expected likelihood of transition from "sick" to "not sick," i.e., $SP(s_{t+1} = 0|s_t = 1, d)$, as $1 - SP(s_{t+1} = 1|s_t = 1, d)$. Likewise, we can compute the expected likelihood of transition from "not sick" to "not sick," i.e., $SP(s_{t+1} = 0|s_t = 0, d)$, as $1 - SP(s_{t+1} = 1|s_t = 0, d)$. We also asked the respondents about two SPEs conditional on clean fuel usage, $SP(s_{t+1} = 1|s_t = 1, c)$ and $SP(s_{t+1} = 1|s_t = 0, c)$, where $c$ symbolizes clean fuel usage.

The other two transition probabilities, $SP(s_{t+1} = 0|s_t = 1, c)$ and $SP(s_{t+1} = 0|s_t = 0, c)$, can be computed in a similar manner to their counterparts on dirty fuel usage. Table 2 is presented to facilitate understanding of these transition probabilities.

For elicitation, we asked each respondent to consider a hypothetical individual identical to her in all respects except her current health status and cooking fuel usage. The respondent is then asked, using a survey instrument, to state how likely she thinks it is that each of the following events will occur: *"Suppose that the individual is currently sick (not sick). How likely is it that she will remain (become) sick in the next 3 months if she uses: a) LPG/kerosene and b) coal/solid biomass fuels/firewood*?*"*. During elicitation, we carefully avoided mentioning the terms "clean" and "dirty," as that would have acted as a signal to the respondents and thereby induced social desirability response bias. All SPE variables were collected in the first round of the survey, unlike the fuel usage pattern and the self-reported health status.

**Characteristics of the elicited SPEs**

We plot the distributions of $SP(s_{t+1} = 1|s_t = 1, c)$ and of $SP(s_{t+1} = 1|s_t = 1, d)$ in panels I and II of Figure 2, respectively. These figures show that a majority of the respondents expressed a low (high) expected probability of transition from "sick" to "sick" in the next three months by using clean (dirty) fuels only. It is also revealed that approximately six percent of the respondents expressed surety about remaining sick from dirty fuel usage, while such a pattern was not observed for clean fuel usage. Based on these results, the respondents seem to recognize the possible health benefits from clean fuel usage. This is also evident in the difference between these two SPEs. As presented in Table 1, the mean of the difference is found to be -0.47: the respondents on average think that clean fuel usage is 47 percent more likely to ease the symptom of the disease, suggesting that a certain number of respondents recognize that sickness is linked to the type of cooking fuel they use.

Panel III displays the distribution of $SP(s_{t+1} = 1|s_t = 0, c)$. Comparing Panel III with Panel I, the distribution of $SP(s_{t+1} = 1|s_t = 0, c)$ differs in shape from that of $SP(s_{t+1} = 1|s_t = 1, c)$; the respondents seem to feel that health status in the next period depends on that in the current period. Panel III also shows that a majority of the respondents assigned a low likelihood to falling sick in the next three months when using clean fuels only.

In Panel IV, we present the distribution of $SP(s_{t+1} = 1|s_t = 0, d)$. Approximately 70 percent of the respondents assigned a moderately high probability (i.e., 0.4 to 0.6) to falling sick in the next period from dirty fuel usage, and a higher likelihood was expressed by approximately 20 percent of the respondents. These results seem to be consistent with the respondents tending to associate dirty fuel usage with the deterioration of health. This observation is also supported by the mean difference of -0.4 between $SP(s_{t+1} = 1|s_t = 0, c)$ and $SP(s_{t+1} = 1|s_t = 0, d)$, as shown in Table 1; the respondents on average think that dirty fuel usage is 40 percent more likely to degrade their health than clean fuel usage.

Using the elicited SPEs conditional on dirty fuel usage, we calculate the equilibrium distribution of the Markov process, denoted as $SP(s = 1|\ d)$[2]. This represents the expectation about the long-term fraction of periods during which the respondent would be sick provided that she uses dirty fuels only. It can therefore be interpreted as the perceived risk from dirty fuel usage on health. Likewise, we derive the equilibrium distribution of the Markov process conditional on clean fuel usage, i.e., $SP(s = 1|c)$, which can be interpreted as the perceived health risk from clean cooking fuel usage.

Panels I and II of Figure 3 present the distributions of $SP(s = 1|\ c)$ and $SP(s = 1|\ d)$, respectively. The two distributions differ greatly in shape. The mean of the latter (0.73) is much larger than that of the former (0.17), as presented in Table 1; the respondents, on average,

[2] $SP(s = 1|d) = SP(s_{t+1} = 1|s_t = 0, d)/1 + [SP(s_{t+1} = 1|s_t = 0, d) - SP(s_{t+1} = 1|s_t = 1, d)]$ . We can obtain the other equilibrium SPE, SP($s = 1|c$ ) using this relationship, replacing the SPEs conditional on dirty fuel usage to the corresponding SPEs conditional on clean fuel usage.

expect that dirty fuel usage will result in considerably longer periods of sickness than clean fuel usage. Furthermore, $SP(s = 1| c)$ has a right-skewed distribution, while $SP(s = 1| d)$ has a left-skewed distribution; for a majority of the respondents, the perceived risk from dirty (clean) fuel usage on health is larger (smaller) than the mean value implies.

**Control variables**

In addition to the SPEs, individual and household specific factors may influence the respondents' cooking fuel usage pattern and health status. In our models, we therefore control for a set of factors including number of cooks (surrogate for household size), total monthly household expenditure, respondents' age and years of schooling, dummy for the occupation of the respondent (respondent is housewife), dummies for the occupation of the spouse (spouse works in the informal sector and in the agricultural sector), dummy for religion (respondent is Hindu), time needed to reach the nearest market on foot (in minutes), dummy for the location of the kitchen (kitchen is located within the dwelling space), dummy for ventilation (cooking area has ventilation facility), and dummies for the ownership of a television and for access to internet.

In rural areas, one can access cooking fuels without incurring any monetary cost. For example, individuals, particularly the women and children in households, spend long hours daily collecting fuels (IEA, 2017), such as firewood from forests, common lands, roadsides, and private fields; crop residues from farms; and dung gathered from domestic animals (Das and Srinivasan, 2012). Such access may influence the respondents' cooking fuel usage pattern. We therefore asked the respondents the following question: "*Do you have an opportunity to collect/get cooking fuels for free*?". Fifty-nine percent of the respondents were found to have access to free cooking fuels (see Table 1).

## 3. Estimation models and results

We address whether SPEs influence individuals' cooking fuel usage patterns, and in turn whether usage patterns affect individuals' health. If so, SPEs have a direct effect on cooking fuel usage patterns and an indirect effect on individuals' health. For this purpose, we take a two-step approach. In the first stage, we estimate the effect of SPEs on individuals' cooking fuel usage patterns. Second, we estimate the effect of cooking fuel usage patterns on individuals' health. This allows us to infer the extent to which SPEs influence individuals' health conditions indirectly through cooking fuel usage.

For this estimation, we need to address several econometric issues. First, the dependent variable in the first stage, that is, the individual's choice of fuel usage pattern, has a particular feature; specifically, it is restricted to values in the unit interval $[0,1]$. In addition, a non-negligible fraction of observations are located within the interval, as mentioned earlier. To fully account for this characteristic of the data, we use a fractional response variable framework (Papke and Wooldridge, 1996).

Second, the fuel usage pattern may be endogenous in the second stage regression of a binary health status variable. To address the nonlinearity of the model as well as the presence of the fractional endogenous regressor, we use the two-stage residual inclusion method (Terza et al., 2008).

Third, the simultaneous elicitation of responses related to cooking fuel usage patterns and SPEs may lead to the issue of reverse causality between them. To address bias due to reverse causality, we use lagged values of SPEs in modeling cooking fuel usage patterns.
For this purpose, we conducted the survey in two rounds with a gap of one year, whereby SPE variables were elicited in the first round (2016−17) and responses related to cooking fuel usage behaviors were collected in the second round (2017−18), as mentioned earlier. In the next sub-sections, we describe the estimation model and results in detail.

### 3.1. Effect of SPEs on cooking fuel usage patterns

#### Estimation model

We assume that individual $i$ chooses fuel usage pattern ($w_i$) based upon her SPEs of falling sick (denoted by the vector $\boldsymbol{spe_i}$), individual and household characteristics ($\boldsymbol{z_i}$), and opportunity to access free cooking fuels ($free_i$). In particular, we specify the conditional mean of $w_i$, given the observed characteristics, in the following manner:

$$E(w_i|\boldsymbol{spe_i}, \boldsymbol{z_i}, free_i) = \Phi(\beta_0 + \boldsymbol{\beta_1 spe_i} + \boldsymbol{\beta_2 z_i} + \beta_3 free_i), \tag{1}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal random variable that ensures that predicted values are in the interval $[0,1]$, as required by the data. For later use, we re-express equation (1) in the following form:

$$w_i = \Phi(\beta_0 + \boldsymbol{\beta_1 spe_i} + \boldsymbol{\beta_2 z_i} + \beta_3 free_i) + r_i, \tag{2}$$

where $r_i$ is an idiosyncratic error term with $E(r_i|\boldsymbol{spe_i}, \boldsymbol{z_i}, free_i) = 0$.

We estimate equation (1) (or, equivalently, equation (2)) by the Bernoulli quasi-maximum likelihood because the obtained estimator is robust to distributional assumptions. It can be shown that this estimator is consistent as long as the conditional mean function is correctly specified (Wooldridge, 2010). We compute robust standard errors, as recommended by Papke and Wooldridge (1996).

#### Estimation results

We estimate equation (1) by including the combinations of the four elicited SPEs as covariates. Specifically, we include the difference in the SPEs conditional on health status "sick" in period $t$ (i.e., $SP(s_{t+1} = 1|s_t = 1, c) - SP(s_{t+1} = 1|s_t = 1, d)$ ) and the difference conditional on health status "not sick" in period $t$ (i.e., $SP(s_{t+1} = 1|s_t = 0, c) - SP(s_{t+1} = 1|s_t = 0, d)$). The former (latter) difference refers to the expected reduction in the likelihood of being sick due to clean fuel usage, given that the current health status is "sick" ("not sick").

Column 1 of Table 3 presents the estimation results. In line with the previous literature, the results show that individuals with higher income (proxied by household expenditure), higher levels of education, better access to information (via internet access), and an affiliation with the Hindu religion tend to reduce their dirty fuel usage. It is also found that proximity to the market induces households to decrease dirty fuel usage, while access to free cooking fuels motivates them to increase dirty fuel usage. The difference in the elicited SPEs conditional on current health status “not sick” is positively associated with how often dirty fuel is used, but the level of significance is marginal ($p<0.1$). This suggests that an increase in the expected reduction of health risks by clean fuel usage given that the current health status is “not sick” will lower the dirty fuel usage. However, the differences in the elicited SPE variables are not found to play an important role in cooking fuel usage patterns when separately including the difference of the SPEs conditional on “not sick” (Column 2) or conditional on “sick” (Column 3).

To the extent that individuals are concerned about the long-term fraction of time during which they would be sick, the equilibrium SPEs may play a role in cooking fuel usage patterns. To explore this possibility, we examine the equilibrium SPEs as covariates for equation (1). As presented in Column 1a of Table 4, the results show that although $SP(s = 1|\, c)$ is not associated with cooking fuel usage patterns, $SP(s = 1|\, d)$ is ($p < 0.05$). The average marginal effect of $SP(s = 1|\, d)$ is found to be -0.188 (Column 1b), indicating that if an individual’s perceived health risk from dirty fuel usage increases by 10 percentage points, she will lower the fraction of days of dirty fuel usage by approximately 1.9 percentage points.

We also examine whether the difference between the equilibrium SPEs (i.e., $SP(s = 1|\, c) - SP(s = 1|\, d)$) and their ratio (i.e., $SP(s = 1|\, c)/SP(s = 1|\, d)$) matter to cooking fuel usage patterns. Both measures refer to the individuals’ perceived reduction in health risk from using clean cooking fuel instead of dirty fuel. As presented in Column 2a, the difference in the equilibrium SPEs is positively and significantly associated with the fraction of days of dirty fuel usage ($p < 0.05$), suggesting that an increase in the expected reduction of health risks through clean fuel usage will lower dirty fuel usage. According to the average marginal effect (Column 2b), a reduction of 10 percentage points in the perceived health risk due to clean cooking fuel usage is associated with a 1.8 percentage point decrease in the fraction of days of dirty fuel usage. Similar results are obtained for the ratio of the equilibrium SPEs (Columns 3a and 3b), although the significance is marginal ($p < 0.1$). Overall, our results show that individuals' SPEs have some influence on their cooking fuel usage patterns, but the magnitude of the impact may be small.

To account for the distribution of the cooking fuel usage pattern, we used the fractional response variable framework. To examine the robustness of our results to estimation methods, we re-estimated equation (1) by fitting a linear regression. As Table 5 presents, the results are qualitatively and quantitatively similar to those based on the fractional response models. Our main results therefore do not seem to be driven by the fractional variable framework.

### 3.2. Effect of cooking fuel usage pattern on health

**Estimation model**

We assume that the underlying health status of individual $i$ ($s_i^*$) is unobservable and depends on the fuel usage pattern ($w_i$), individual and household specific characteristics, and individuals' SPEs ($\boldsymbol{x_i} = [\boldsymbol{z_i}\ \boldsymbol{spe_i}]$):

$$s_i^* = \gamma_0 + \gamma_1 w_i + \boldsymbol{\gamma_2 x_i} + \gamma_3 r_i + u_i, \tag{3}$$

where $u_i$ is an idiosyncratic error term that is uncorrelated with $w_i$ and $\boldsymbol{x_i}$. In equation (3), the presence of $r_i$ defined in equation (2) is a potential cause of endogeneity issues; if not controlled for, $r_i$ could be absorbed by the error term, thereby inducing a correlation between the error term and fuel usage pattern ($w_i$). This is not the case only when $\gamma_3 = 0$.

The observed self-reported health status of the individual ($s_i$) is an indicator variable that takes the value of one if the individual has suffered from at least one of the aforementioned disease symptoms in the last 30 days. We assume that $s_i$ and $s_i^*$ are associated in the following manner: $s_i = 1 \; if \; s_i^* \geq 0$ and $s_i = 0 \; if \; s_i^* < 0$. We also assume that $u_i$ follows a standard normal distribution. Under these assumptions, the response probability can be derived as follows:

$$\Pr(s_i = 1 | w_i, \boldsymbol{\gamma_2 x_i}, r_i) = \Phi(\gamma_0 + \gamma_1 w_i + \boldsymbol{\gamma_2 x_i} + \gamma_3 r_i). \tag{4}$$

Equation (4) suggests a two-step estimation procedure, which is an application of the two-stage residual inclusion (Terza et al., 2008). The first stage is to estimate equation (2) (which we have already done) and compute a residual $\hat{r}_i$ for each $i$. The second stage involves replacing $r_i$ with $\hat{r}_i$ in equation (4) and estimating the model using maximum likelihood. We compute standard errors using bootstrapping (with 500 replications) to account for the fact that $\hat{r}_i$ is a generated regressor. For identification, we need an instrumental variable; at least one regressor in equation (2) should not be included in equation (4) because the fuel usage pattern variable may be correlated with the omitted variable $r_i$. In our estimation, the opportunity to access free cooking fuels ($free_i$) plays a role as an instrument.

## Estimation results

Columns 1a and 1b in Table 6 present the estimated coefficients and corresponding average marginal effects, respectively, where the first stage model is specified as in Column 1a of Table 4. We find that the coefficient on the first stage residual is not significant at the ten percent level; in other words, there is little evidence that the fuel usage pattern is endogenous in equation (4). We also observe that the fuel usage pattern is positively and significantly associated with the likelihood of being sick with at least one of the physical symptoms ($p <$ 0.01). Based on the average marginal effect, the likelihood increases by approximately six percentage points with an increase in dirty fuel usage of ten percentage points. These results seem to remain largely unchanged even when the first stage model is specified as in either Columns 2a or Column 3a of Table 4. Overall, the results for equation (4), along with those for equation (1), support the finding that the SPEs indirectly affect health status by influencing cooking fuel usage patterns.

## 4. Simulation of policies

Because SPEs were most likely made based on very limited information or even no information except past experience, respondents could update their SPEs upon obtaining information regarding dirty/clean fuel usage. This provides a rationale for information provision policies, given our evidence that SPEs matter to cooking fuel usage patterns. Several studies have shown that information provision may reduce health and environmental risks in developing countries (Somanathan, 2010). For example, the disclosure of information on water contamination has induced households to switch their drinking water source (Barnwal et al., 2017) or adopt related preventive measures (Brown et al., 2017). Likewise, Dendup and Arimura (2019) established that access to information can influence clean cooking fuel usage in their study of rural Bhutanese households. However, contradictory evidence is also available in the literature. For instance, Delavande and Kohler (2015) concluded that information provision about the transmission risk of HIV disease has limited impact on risky sexual behavior in rural Malawi. Correspondingly, the provision of information about health seems to be ineffective in the adoption of preventive measures for malarial diseases in Kenya (Dupas, 2009).

In this section, we consider and evaluate different information provision alternatives by simulating the extent to which they influence the fuel usage patterns and consequently the health of individuals. We assume that in response to information provision about the health risks associated with different cooking fuel categories, all individuals update their SPEs to a preset benchmark level. It would be ideal to examine how individuals respond when information is provided on the "true" risk of dirty/clean fuel usage, as it could serve as a benchmark. Unfortunately, such information is not available to us. We therefore consider cases in which individuals are informed of the proportion of sick individuals among those who predominantly use clean (dirty) fuels as the benchmark level. In our sample, 47 (397) respondents out of 130 (458) who primarily used clean (dirty) fuels were found to be sick; 0.36 (0.87) is therefore set as the benchmark probability of becoming sick from clean (dirty) fuel usage.

We explore three alternative outcomes: first, when only information about health risk from clean cooking fuel is provided; second, when only information about health risks associated with dirty cooking fuel is provided; and finally, when both types of information are provided simultaneously. For the policy simulation analysis, we use the estimated models specified in Column 1a of Table 4 and Table A1. The model in Table A1 is a restricted version of that reported in Table 6, where the first stage residuals are excluded from a set of regressors. Table 7 presents summary statistics for the predicted fuel usage patterns and the proportions of sick individuals under the policy alternatives. The table also provides the average changes in predicted probabilities and proportions of individuals whose predicted dirty fuel usage increased or decreased from the baseline scenario.

In response to information provision about the health risks from clean cooking fuels, 547 (41) individuals who had an $SP(s = 1|c)$ below (above) than 0.36 update to the benchmark level. If the objective of information provision is to reduce (increase) dirty (clean) fuel usage among households, the policy maker is likely to have an undesirable outcome. Most of the individuals (approximately 93%) become pessimistic about clean cooking fuels and increase their usage of dirty fuel compared to the no-policy situation; specifically, the average fraction of days of dirty fuel usage increases from 67.8% to 70.9%. This increase in dirty fuel usage also adversely affects the health of the individuals. On average, the predicted proportion of sick individuals increases to 83.5% from the baseline (82.3%).

When only information about health risks associated with dirty cooking fuels is provided, 510 (78) individuals who had an $SP(s = 1|d)$ below (above) than 0.87 updated to that level. Under this policy alternative, the majority of the individuals become aware of the health risks associated with dirty fuel usage. On average, the fraction of days of dirty fuel usage reduces to 65.3% from the baseline situation (67.8%). This reduction possibly occurs because approximately 87% of the individuals reduce their dirty fuel usage compared with the baseline situation, as they turn pessimistic about dirty cooking fuels. As a result of this reduction in dirty fuel usage, the associated health conditions among the individuals improve. The average predicted proportion of sick individuals reduces to 80.5% compared to the no-policy situation (82.3%). Although the average change in predicted probabilities of becoming sick is quite small (1.8%) in our sample, this result suggests that providing information about the health risks related to dirty cooking fuel may lead to a reduction in dirty fuel usage among the households and consequently improve the related health status.

Because the impact of information works in the opposite direction for clean fuel versus dirty fuel usage, the effect of the policy becomes nullified when information on the health risks associated with both clean and dirty cooking fuels is provided simultaneously. We observe that the mean fraction of days of dirty fuel usage remains almost unchanged at 68% of days under this policy alternative. Accordingly, simultaneous information provision also does not affect the proportion of sick individuals; the mean predicted proportion of sick individuals remains almost unchanged at 82%. However, in comparison with the situation in which only information about the health risks from clean cooking fuels is provided, approximately 40% of the individuals decrease their dirty fuel usage under this policy alternative.

Summarizing the impacts of different policy scenarios on cooking fuel usage patterns and health risks, we can conclude that information provision on the risks of dirty cooking fuel may lead to a reduction of dirty fuel usage and consequently improve health status, but the magnitude of the change is limited in our analysis. On the other hand, information provision about clean cooking fuels may have an undesired effect on the fuel usage pattern and consequently on health.

## 5. Conclusion

In this study, we analyze a unique dataset from rural Indian households to examine individuals' SPEs of becoming sick with the diseases typically observed from IAP exposure and their impact on cooking fuel usage patterns. Simultaneously, we also investigate how individuals' cooking fuel usage patterns, in turn, affect their health status. Our results support the hypothesis that individuals' expectations have some influence on cooking fuel usage patterns, although the magnitude of the impact seems to be rather small. The results also show that the increased usage of dirty cooking fuels is likely to increase individuals' likelihood of suffering from physical symptoms related to IAP. Based on the estimated coefficients, the policy simulation analysis suggests that providing information on the health risks of dirty cooking fuels will reduce dirty fuel usage among the individuals and consequently enhance the associated health status. However, the magnitude of the change is limited in our analysis.

At least two limitations of our analyses should be mentioned. First, we do not have detailed knowledge on how individuals update beliefs. In our policy simulation, we presumed that individuals update their beliefs immediately and accurately when they receive information. However, this may not be the case. Our simulation is based upon this assumption, and therefore, our results would represent the upper (or lower) bound of the effects of information provision. A study on the mechanism of belief updating is required. Second, examination of the content of information provision is not sufficient. We used, for simplicity, information on the proportion of sick individuals disaggregated by the primary fuel used, which can be different from the objective probability. More detailed analyses on “true” risk are required, including evaluation of heterogeneity regarding individual characteristics.

Although we have confined our attention to the health status of the respondents, this may be an oversimplification. To analyze the health impact of IAP holistically, we need to investigate the health status of the other household members, particularly that of children. Focusing only on the respondents’ expectations may be another oversimplification because cooking fuel usage is a household decision that may involve a trade-off between the perceptions of the respondent and her family members regarding cooking fuels. One future research avenue is the extension of our analysis by incorporating intra-household expectations. Finally, to comprehensively assess the role of SPEs on health, we plan to focus on the long-term health effects of IAP, which may demand the involvement of professional medical teams during the elicitation of responses.

**Acknowledgment:**

We thank Global Climate Change, Jadavpur University for their infrastructural assistance with our field survey. This research was supported by the Research Institute for Environmental Economics and Management (RIEEM), Waseda University [JSPS KAKENHI Grant Numbers 16K13364].

## References


1. Alem, Y., Beyene, A. D., Köhlin, G., & Mekonnen, A. (2016). Modeling household cooking fuel choice: a panel multinomial logit approach. *Energy Economics*, *59*, 129-137.https://doi.org/10.1016/j.eneco.2016.06.025
2. Baranov, V., Bennett, D., & Kohler, H. P. (2015). The indirect impact of antiretroviral therapy: Mortality risk, mental health, and HIV-negative labor supply. *Journal of health economics*, *44*, 195-211. https://doi.org/10.1016/j.jhealeco.2015.07.008
3. Barnwal, P., van Geen, A., von der Goltz, J., & Singh, C. K. (2017). Demand for environmental quality information and household response: Evidence from well-water arsenic testing. *Journal of Environmental Economics and Management*, *86*, 160-192.https://doi.org/10.1016/j.jeem.2017.08.002
4. Bonjour, S., Prüss-Üstün, A., & Rehfuess, E. (2007). *Indoor Air Pollution: National Burden of Disease Estimates*. World Health Organization.
5. Brown, J., Hamoudi, A., Jeuland, M., & Turrini, G. (2017). Seeing, believing, and behaving: Heterogeneous effects of an information intervention on household water treatment. *Journal of Environmental Economics and Management*, *86*, 141-159. https://doi.org/10.1016/j.jeem.2016.08.005
6. Das, D., & Srinivasan, R. (2012). Income levels and transition of cooking fuel among rural poor in India. *Energy Science and Technology*, *4*(2), 85-91. https://doi.org/10.3968/j.est.1923847920120402.226
7. Delavande, A. (2014). Probabilistic expectations in developing countries. *Annual Review of Economics*, *6*(1), 1-20. https://doi.org/10.1146/annurev-economics-072413-105148
8. Delavande, A., & Kohler, H. P. (2009). Subjective expectations in the context of HIV/AIDS in Malawi. *Demographic research*, *20*, 817.-874. https://dx.doi.org/10.4054/DemRes.2009.20.31
9. Delavande, A., & Kohler, H. P. (2015). HIV/AIDS-related expectations and risky sexual behaviour in Malawi. *The Review of Economic Studies*, *83*(1), 118-164 https://doi.org/10.1093/restud/rdv028
10. Dendup, N., & Arimura, T. H. (2019). Information leverage: The adoption of clean cooking fuel in Bhutan. *Energy Policy*, *125*, 181-195. https://doi.org/10.1016/j.enpol.2018.10.054
11. Dupas, P. (2009). What matters (and what does not) in households' decision to invest in malaria prevention?. *American Economic Review*, *99*(2), 224-30. https://doi.org/10.1257/aer.99.2.224

12. Farsi, M., Filippini, M., & Pachauri, S. (2007). Fuel choices in urban Indian households. *Environment and Development Economics*, *12*(6), 757-774. https://doi.org/10.1017/S1355770X07003932
13. Godlonton, S., & Thornton, R. L. (2013). Learning from Others' HIV Testing: Updating Beliefs and Responding to Risk. *American Economic Review*, *103*(3), 439-44. https://doi.org/10.1257/aer.103.3.439
14. Gupta, G., & Köhlin, G. (2006). Preferences for domestic fuel: analysis with socio-economic factors and rankings in Kolkata, India. *Ecological Economics*, *57*(1), 107-121. https://doi.org/10.1016/j.ecolecon.2005.03.010
15. Hanna, R., Duflo, E., & Greenstone, M. (2016). Up in smoke: the influence of household behavior on the long-run impact of improved cooking stoves. *American Economic Journal: Economic Policy*, *8*(1), 80-114. https://doi.org/10.1257/pol.20140008
16. Heltberg, R. (2005). Factors determining household fuel choice in Guatemala. *Environment and development economics*, *10*(3), 337-361. https://doi.org/10.1017/S1355770X04001858
17. International Energy Agency. (2017). Energy Access Outlook 2017: From Poverty to Prosperity. IEA.
18. Jeuland, M., Pattanayak, S.K. and Bluffstone, R. (2015). The Economics of Household Air Pollution, *Annual Review of Resource Economics*, *7*(1), 81-108 https://doi.org/10.1146/annurev-resource-100814-125048
19. Kézdi, G., & Willis, R. J. (2011). *Household stock market beliefs and learning* (No. w17614). National Bureau of Economic Research. https://doi.org/10.3386/w17614
20. Manski, C. F. (2004). Measuring expectations. *Econometrica*, *72*(5), 1329-1376. https://doi.org/10.1111/j.1468-0262.2004.00537.
21. McKenzie, D., Gibson, J., & Stillman, S. (2013). A land of milk and honey with streets paved with gold: Do emigrants have over-optimistic expectations about incomes abroad?. *Journal of Development Economics*, *102*, 116-127. https://doi.org/10.1016/j.jdeveco.2012.01.001
22. Pachauri, S., & Jiang, L. (2008). The household energy transition in India and China. *Energy policy*, *36*(11), 4022-4035. https://doi.org/10.1016/j.enpol.2008.06.016
23. Papke, L. E., & Wooldridge, J. M. (1996). Econometric methods for fractional response variables with an application to 401 (k) plan participation rates. *Journal of applied econometrics*, *11*(6),619-632.

https://doi.org/10.1002/(SICI)1099-1255(199611)11:6%3C619::AID-JAE418%3E3.0.CO;2-1

24. Pitt, M. M., Rosenzweig, M. R., & Hassan, M. N. (2006, June). Sharing the burden of disease: gender, the household division of labor and the health effects of indoor air pollution in Bangladesh and India. In *Stanford Institute for Theoretical Economics Summer Workshop* (Vol. 202).
25. Somanathan, E. (2010). Effects of Information on Environmental Quality in Developing Countries. *Review of Environmental Economics and Policy*, *4*(2), 275–292. https://doi.org/10.1093/reep/req012
26. Terza, J. V., Basu, A., & Rathouz, P. J. (2008). Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. *Journal of health economics*, *27*(3), 531-543. https://doi.org/10.1016/j.jhealeco.2007.09.009
27. Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

## Table 1. Descriptive statistics (N = 557)

| Variable | Mean | SD | Min | Max |
|---|---|---|---|---|
| *Sick in last 30 days with at least one physical symptom* (binary) | 0.76 | 0.43 | 0 | 1 |
| *Fraction of days of dirty fuel usage in 30 days prior to previous month* | 0.68 | 0.38 | 0 | 1 |
| *Subjective probabilistic expectations (SPE) variables* | | | | |
| $SP(s_{t+1} = 1 \| s_t = 0, c) - SP(s_{t+1} = 1 \| s_t = 0, d)$ | -0.40 | 0.15 | -0.8 | 0.5 |
| $SP(s_{t+1} = 1 \| s_t = 1, c) - SP(s_{t+1} = 1 \| s_t = 1, d)$ | -0.47 | 0.16 | -0.8 | 0.7 |
| $SP(s = 1 \| c)$ | 0.17 | 0.12 | 0 | 0.91 |
| $SP(s = 1 \| d)$ | 0.73 | 0.14 | 0.11 | 1 |
| $SP(s = 1 \| c) - SP(s = 1 \| d)$ | -0.56 | 0.17 | -1 | 0.52 |
| $SP(s = 1 \| c)/SP(s = 1 \| d)$ | 0.25 | 0.24 | 0 | 4 |
| *Control Variables* | | | | |
| Number of cooks | 1.13 | 0.41 | 1 | 4 |
| Age | 37.78 | 10.79 | 17 | 76 |
| Years of schooling | 4.83 | 4.13 | 0 | 17 |
| Hindu (binary) | 0.68 | 0.47 | 0 | 1 |
| Housewife (binary) | 0.97 | 0.17 | 0 | 1 |
| Spouse works in informal sector (binary) | 0.43 | 0.50 | 0 | 1 |
| Spouse works in agricultural sector (binary) | 0.30 | 0.46 | 0 | 1 |
| Kitchen located inside dwelling unit (binary) | 0.16 | 0.36 | 0 | 1 |
| Access to ventilation in cooking area (binary) | 0.97 | 0.16 | 0 | 1 |
| Time to market (minutes) | 17.52 | 13.60 | 1 | 120 |
| Expenditure (in INR 1,000) | 7.51 | 3.74 | 2.3 | 55 |
| Access to internet (binary) | 0.24 | 0.43 | 0 | 1 |
| Owns television (binary) | 0.86 | 0.35 | 0 | 1 |
| Opportunity to collect cooking fuels for free (binary) | 0.59 | 0.49 | 0 | 1 |

Note: In subjective probabilistic expectation (SPE) variables, $SP(\cdot \mid \cdot)$, $s$ denotes the state of being sick with at least one of the physical symptoms (dry cough, sore or runny eyes, and difficulty breathing) and $c$ $(d)$ represents clean (dirty) fuel usage.

**Table 2. SPEs about transition probabilities of health status conditional on dirty cooking fuel usage**

| Health status in period $t$ | Health status in period $t$+1 | |
|---|---|---|
| | $sick$ | $not\ sick$ |
| $sick$ | $SP(s_{t+1} = 1 \mid s_t = 1, d)$ | $SP(s_{t+1} = 0 \mid s_t = 1, d)$ |
| $not\ sick$ | $SP(s_{t+1} = 1 \mid s_t = 0, d)$ | $SP(s_{t+1} = 0 \mid s_t = 0, d)$ |

Note: This table is provided to facilitate understanding our notations for the SPE variables. Replace $d$ (dirty fuel usage) with $c$ (clean fuel usage) for the SPEs about transition probabilities of health status conditional on clean cooking fuel usage.

**Table 3 First stage results with the elicited SPEs (average marginal effects)**

| | (1) | (2) | (3) |
|---|---|---|---|
| $SP(s_{t+1} = 1 \mid s_t = 0, c) - SP(s_{t+1} = 1 \mid s_t = 0, d)$ | 0.161* | 0.137 | |
| | (0.095) | (0.092) | |
| $SP(s_{t+1} = 1 \mid s_t = 1, c) - SP(s_{t+1} = 1 \mid s_t = 1, d)$ | -0.071 | | -0.021 |
| | (0.084) | | (0.081) |
| Number of cooks | 0.079** | 0.078** | 0.083** |
| | (0.032) | (0.032) | (0.032) |
| Age | -0.002* | -0.002* | -0.002* |
| | (0.001) | (0.001) | (0.001) |
| Years of schooling | -0.014*** | -0.014*** | -0.014*** |
| | (0.003) | (0.003) | (0.003) |
| Hindu | -0.154*** | -0.153*** | -0.157*** |
| | (0.029) | (0.029) | (0.029) |
| Housewife | -0.09 | 0.087 | -0.098 |
| | (0.064) | (0.064) | (0.066) |
| Spouse works in informal sector | 0.012 | 0.013 | 0.007 |
| | (0.03) | (0.03) | (0.03) |
| Spouse works in agricultural sector | 0.082** | 0.082** | 0.078** |
| | (0.035) | (0.035) | (0.035) |
| Kitchen located inside dwelling unit | 0.018 | 0.021 | 0.022 |
| | (0.038) | (0.038) | (0.038) |
| Access to ventilation | 0.086 | 0.088 | 0.081 |
| | (0.082) | (0.055) | (0.082) |
| Time to market | 0.003*** | 0.003*** | 0.003*** |
| | (0.001) | (0.001) | (0.001) |
| Expenditure | -0.017*** | -0.018*** | -0.018*** |
| | (0.004) | (0.005) | (0.005) |
| Access to internet | -0.13*** | -0.129*** | -0.13*** |
| | (0.028) | (0.028) | (0.028) |
| Owns television | -0.08 | -0.083*** | -0.082*** |
| | (0.042) | (0.042) | (0.041) |
| Opportunity to access free cooking fuel | 0.192*** | 0.193*** | 0.189*** |
| | (0.024) | (0.024) | (0.024) |
| *Pseudo log-likelihood* | -274.059 | -274.244 | -274.897 |
| *Pseudo* $R^2$ | 0.216 | 0.216 | 0.214 |
| $\chi^2$ | 282.67 | 284.09 | 284.59 |

Note: This table provides estimation results for equation (1), where the dependent variable is the fraction of days of dirty fuel usage. Average marginal effects of the variables are presented. The sample size is 557. ***, ** and * indicate statistical significance at the one, five and ten percent levels, respectively. Standard errors in parentheses are computed by the delta method with robust standard errors for the parameters.

## Table 4. First stage results with the equilibrium SPEs

| | (1a) | (1b) | (2a) | (2b) | (3a) | (3b) |
|---|---|---|---|---|---|---|
| | Coef. | AME | Coef. | AME | Coef. | AME |
| $SP(s=1\|c)$ | 0.585 | 0.161 | | | | |
| | [0.424] | (0.117) | | | | |
| $SP(s=1\|d)$ | -0.682** | -0.188** | | | | |
| | [0.345] | (0.095) | | | | |
| $SP(s=1\|c)-SP(s=1\|d)$ | | | 0.641** | 0.177** | | |
| | | | [0.285] | (0.079) | | |
| $SP(s=1\|c)/SP(s=1\|d)$ | | | | | 0.396* | 0.109* |
| | | | | | [0.223] | (0.062) |
| Number of cooks | 0.287** | 0.079** | 0.286** | 0.079** | 0.288** | 0.079** |
| | [0.116] | (0.032) | [0.116] | (0.032) | [0.117] | (0.032) |
| Age | -0.008* | -0.002* | -0.008* | -0.002* | -0.008* | -0.002* |
| | [0.005] | (0.001) | [0.05] | (0.001) | [0.005] | (0.001) |
| Years of schooling | -0.052*** | -0.014*** | -0.052*** | -0.014*** | -0.051*** | -0.014*** |
| | [0.012] | (0.003) | [0.012] | (0.003) | [0.012] | (0.003) |
| Hindu | -0.531*** | -0.146*** | -0.532*** | -0.147*** | -0.558*** | -0.154*** |
| | [0.110] | (0.029) | [0.11] | (0.029) | [0.109] | (0.029) |
| Housewife | -0.304 | -0.084 | -0.305 | -0.084 | -0.336 | -0.092 |
| | [0.229] | (0.063) | [0.228] | (0.063) | [0.229] | (0.063) |
| Spouse works in informal sector | 0.071 | 0.019 | 0.071 | 0.02 | 0.045 | 0.012 |
| | [0.111] | (0.031) | [0.111] | (0.031) | [0.11] | (0.03) |
| Spouse works in agricultural sector | 0.299** | 0.083** | 0.298** | 0.082** | 0.285** | 0.079** |
| | [0.126] | (0.035) | [0.126] | (0.035) | [0.126] | (0.035) |
| Kitchen located inside dwelling unit | 0.077 | 0.021 | 0.078 | 0.021 | 0.087 | 0.024 |
| | [0.141] | (0.038) | [0.121] | (0.038) | [0.14] | (0.038) |
| Access to ventilation | 0.329 | 0.096 | 0.33 | 0.096 | 0.303 | 0.087 |
| | [0.274] | (0.083) | [0.274] | (0.083) | [0.272] | (0.082) |
| Time to market | 0.01*** | 0.003*** | 0.01** | 0.003*** | 0.01** | 0.003** |
| | [0.004] | (0.001) | [0.004] | (0.001) | [0.004] | (0.001) |
| Expenditure | -0.063*** | -0.018*** | -0.064*** | -0.018*** | 0.065*** | -0.018*** |
| | [0.017] | (0.005) | [0.017] | (0.005) | [0.017] | (0.004) |
| Access to internet | -0.458*** | -0.126*** | -0.458*** | -0.127*** | -0.461*** | -0.127*** |
| | [0.103] | (0.028) | [0.103] | (0.028) | [0.103] | (0.028) |
| Owns television | -0.298* | -0.082* | -0.296* | -0.082* | -0.288* | -0.079* |
| | [0.152] | (0.042) | [0.152] | (0.042) | [0.149] | (0.041) |
| Opportunity to access free cooking fuel | 0.714*** | 0.197*** | 0.714*** | 0.197*** | 0.704*** | 0.194*** |
| | [0.092] | (0.024) | [0.092] | (0.024) | [0.091] | (0.024) |
| *Log-likelihood* | -273.459 | | -273.469 | | -273.979 | |
| *Pseudo $R^2$* | 0.218 | | 0.218 | | 0.216 | |
| $\chi^2$ | 295.6 | | 295.04 | | 292.16 | |

Note: The dependent variable is the fraction of days of dirty fuel usage. AME denotes average marginal effects. The sample size is 557. ***, ** and * indicate statistical significance at the one, five and ten percent levels, respectively. Robust standard errors are in brackets. Standard errors in parentheses are computed by the delta method with robust standard errors for the parameters. The constant terms are not reported for the sake of space.

**Table 5. Robustness check of first stage estimation (linear regression)**

| | (1) | (2) | (3) |
|---|---|---|---|
| $SP(s=1\|c)$ | 0.147 | | |
| | (0.116) | | |
| $SP(s=1\|d)$ | -0.197** | | |
| | (0.096) | | |
| $SP(s=1\|c)-SP(s=1\|d)$ | | 0.176** | |
| | | (0.079) | |
| $SP(s=1\|c)/SP(s=1\|d)$ | | | 0.093** |
| | | | (0.043) |
| Number of cooks | 0.058** | 0.058** | 0.059** |
| | (0.029) | (0.029) | (0.03) |
| Age | -0.002* | -0.002* | -0.002* |
| | (0.001) | (0.001) | (0.001) |
| Years of schooling | -0.016*** | -0.016*** | -0.016*** |
| | (0.004) | (0.004) | (0.004) |
| Hindu | -0.152*** | -0.152*** | -0.16*** |
| | (0.029) | (0.029) | (0.029) |
| Housewife | -0.088 | -0.088 | -0.097 |
| | (0.067) | (0.067) | (0.067) |
| Spouse works in informal sector | 0.022 | 0.023 | 0.015 |
| | (0.036) | (0.036) | (0.036) |
| Spouse works in agricultural sector | 0.08** | 0.079** | 0.076** |
| | (0.038) | (0.038) | (0.038) |
| Kitchen located inside | 0.029 | 0.03 | 0.033 |
| | (0.04) | (0.039) | (0.04) |
| Access to ventilation | 0.108 | 0.108 | 0.1 |
| | (0.083) | (0.083) | (0.083) |
| Time to market | 0.003*** | 0.003*** | 0.003*** |
| | (0.001) | (0.001) | (0.001) |
| Expenditure | -0.013*** | -0.013*** | -0.013*** |
| | (0.005) | (0.005) | (0.005) |
| Access to internet | -0.157*** | -0.158*** | -0.159*** |
| | (0.035) | (0.035) | (0.036) |
| Ownership of television | -0.073** | -0.073** | -0.07** |
| | (0.032) | (0.032) | (0.031) |
| Opportunity to access free cooking fuel | 0.211*** | 0.211*** | 0.209*** |
| | (0.029) | (0.029) | (0.029) |
| $R^2$ | 0.373 | 0.372 | 0.369 |
| *Adjusted* $R^2$ | 0.354 | 0.355 | 0.352 |
| *F(1,557)* | 29.4 | 31.32 | 31.07 |

Note: The dependent variable is the fraction of days of dirty fuel usage. Average marginal effects of the variables are presented. The sample size is 557. ***, ** and * indicate statistical significance at the one, five and ten percent levels, respectively. Robust standard errors are in parentheses. The constant terms are not reported for the sake of space.

## Table 6. Second stage results with the equilibrium SPEs

| | (1a) Coef. | (1b) AME | (2a) Coef. | (2b) AME | (3a) Coef. | (3b) AME |
|---|---|---|---|---|---|---|
| Fraction of days of dirty fuel usage | 4.111*** | 0.614*** | 4.115*** | 0.616*** | 4.172*** | 0.623*** |
| | [0.754] | (0.105) | [0.747] | (0.104) | [0.751] | (0.104) |
| $SP(s = 1\|c)$ | -0.335 | -0.05 | | | | |
| | [0.638] | (0.095) | | | | |
| $SP(s = 1\|d)$ | 0.060 | 0.009 | | | | |
| | [0.693] | (0.104) | | | | |
| $SP(s = 1\|c) - SP(s = 1\|d)$ | | | -0.18 | -0.027 | | |
| | | | [0.439] | (0.065) | | |
| $SP(s = 1\|c)/SP(s = 1\|d)$ | | | | | -0.151 | -0.022 |
| | | | | | [0.337] | (0.05) |
| Number of cooks | 0.032 | 0.005 | 0.027 | 0.004 | 0.023 | 0.003 |
| | [0.237] | [0.035] | [0.232] | (0.035) | [0.231] | (0.034) |
| Age | 0.018* | 0.003* | 0.019* | 0.003* | 0.019* | 0.002* |
| | [0.01] | [0.001] | [0.009] | (0.001) | [0.009] | (0.001) |
| Years of schooling | 0.022 | 0.003 | 0.022 | 0.003 | 0.023 | 0.003 |
| | [0.029] | [0.004] | [0.028] | (0.004) | [0.028] | (0.004) |
| Hindu | 0.163 | 0.024 | 0.166 | 0.025 | 0.178 | 0.027 |
| | [0.269] | (0.04) | [0.27] | (0.041) | [0.267] | (0.04) |
| Housewife | 0.872** | 0.13** | 0.867** | 0.129** | 0.879** | 0.131** |
| | [0.401] | (0.059) | [0.4] | (0.059) | [0.395] | (0.058) |
| Spouse works in informal sector | 0.234 | 0.035 | 0.232 | 0.035 | 0.236 | 0.035 |
| | [0.228] | (0.033) | [0.225] | (0.033) | [0.223] | (0.033) |
| Spouse works in agricultural sector | -0.355 | -0.053 | -0.358 | -0.053 | -0.362 | -0.054 |
| | [0.268] | (0.04) | [0.264] | (0.039) | [0.265] | (0.039) |
| Kitchen located inside dwelling unit | 0.159 | 0.023 | 0.166 | 0.024 | 0.161 | 0.024 |
| | [0.235] | (0.034) | [0.237] | (0.034) | [0.237] | (0.034) |
| Access to ventilation | 0.709 | 0.119 | 0.7 | 0.117 | 0.704 | 0.118 |
| | [0.592] | (0.112) | [0.578] | (0.109) | [0.581] | (0.11) |
| Time to market | 0.003 | 0.000 | 0.003 | 0.000 | 0.002 | 0.000 |
| | [0.009] | (0.001) | [0.009] | (0.001) | [0.008] | (0.001) |
| Expenditure | 0.003 | 0.000 | 0.002 | 0.000 | 0.003 | 0.000 |
| | [0.025] | (0.004) | [0.025] | (0.003) | [0.025] | (0.004) |
| Access to internet | 0.208 | 0.031 | 0.206 | 0.031 | 0.215 | 0.032 |
| | [0.258] | (0.038) | [0.255] | (0.038) | [0.253] | (0.037) |
| Owns television | 0.270 | 0.04 | 0.271 | 0.041 | 0.276 | 0.041 |
| | [0.393] | (0.059) | [0.391] | (0.059) | [0.389] | (0.059) |
| First stage residuals | -0.773 | -0.116 | -0.774 | -0.116 | -0.835 | -0.125 |
| | [0.769] | (0.115) | [0.762] | (0.114) | [0.76] | (0.114) |
| *Log-likelihood* | -150.739 | | -150.739 | | -150.686 | |
| *Pseudo $R^2$* | 0.510 | | 0.509 | | 0.510 | |
| $\chi^2$ | 128.77 | | 128.77 | | 132.47 | |

Note: This table provides estimation results for equation (3), where the dependent variable is the self-reported health status of the individuals (= 1 if the respondent has experienced at least one of the three physical symptoms). AME denotes average marginal effects. The sample size is 557. ***, ** and * indicate statistical significance at the one, five and ten percent levels, respectively. Standard errors in parentheses are computed by the delta method with bootstrap standard errors for the parameters (number of replications 500). The constant terms are not reported for the sake of space.

## Table 7. Policy simulation results

| | Baseline | Information about the health risks of clean fuels | Information about the health risks of dirty fuels | Information about the health risks of both fuels |
|---|---|---|---|---|
| *Panel I: On fraction of days of dirty fuel usage* | | | | |
| Predicted probability | | | | |
| 25% | 0.504 (0.021) | 0.553 (0.018) | 0.481 (0.024) | 0.534 (0.019) |
| 50% | 0.739 (0.015) | 0.775 (0.01) | 0.706 (0.013) | 0.751 (0.014) |
| 75%` | 0.879 (0.011) | 0.893 (0.009) | 0.857 (0.009) | 0.875 (0.009) |
| Mean | 0.678 (0.01) | 0.709 (0.009) | 0.653 (0.01) | 0.686 (0.01) |
| % with reduced dirty fuel usage | | 7.01% | 87.07% | 39.86% |
| % with increased dirty fuel usage | | 92.99% | 12.93% | 60.14% |
| Average change in predicted probability | | 0.031 (0.001) | -0.025 (0.001) | 0.008 (0.001) |
| conditional on reduced dirty fuel usage | | -0.013 (0.003) | -0.030 (0.001) | -0.021 (0.001) |
| conditional on increased dirty fuel usage | | 0.034 (0.001) | 0.013 (0.002) | 0.026 (0.001) |
| *Panel-II: On probability of being sick* | | | | |
| Predicted probability | | | | |
| 25% | 0.745 (0.019) | 0.779 (0.017) | 0.719 (0.019) | 0.749 (0.016) |
| 50% | 0.911 (0.008) | 0.923 (0.006) | 0.897 (0.007) | 0.908 (0.006) |
| 75% | 0.963 (0.003) | 0.964 (0.003) | 0.957 (0.003) | 0.959 (0.003) |
| Mean | 0.823 (0.008) | 0.835 (0.008) | 0.805 (0.009) | 0.818 (0.009) |
| Average change in predicted probability | | 0.013 (0.001) | -0.018 (0.001) | -0.005 (0.001) |

Note: The results in this table are based on the estimated coefficients in Column 1a of Table 4 and Table A1. Bootstrap standard errors are in parentheses (number of replications 500). The values are approximated up to the third decimal place.

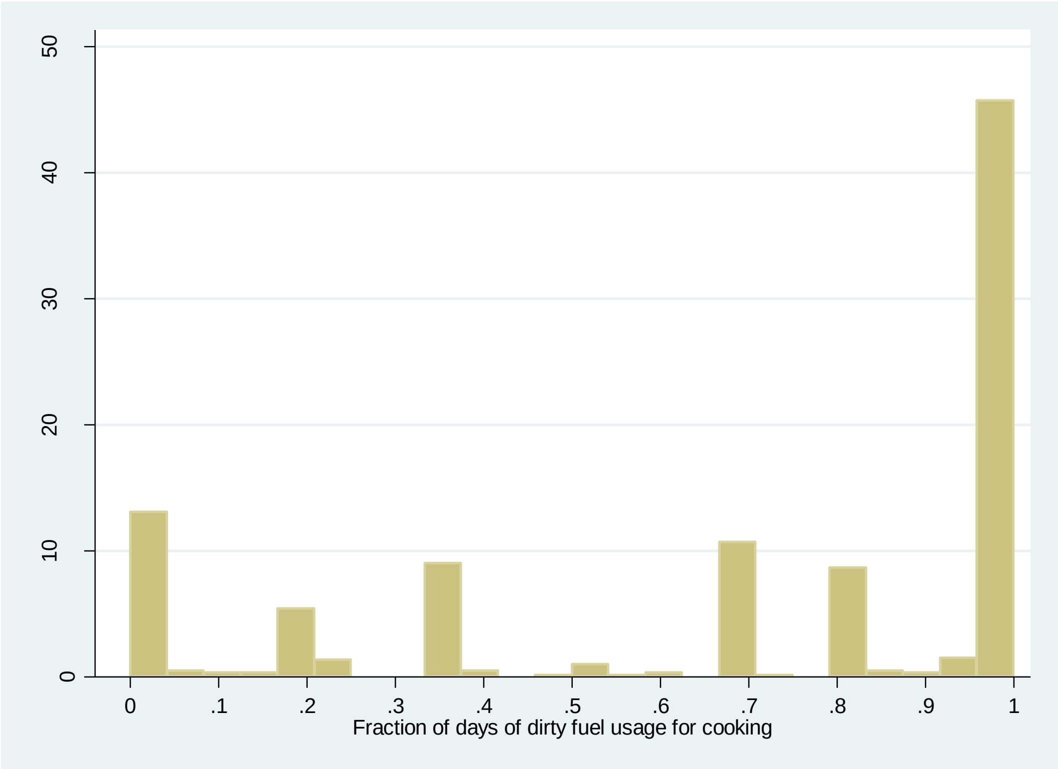


**Figure 1. Distribution of fraction of days of dirty fuel usage for cooking**

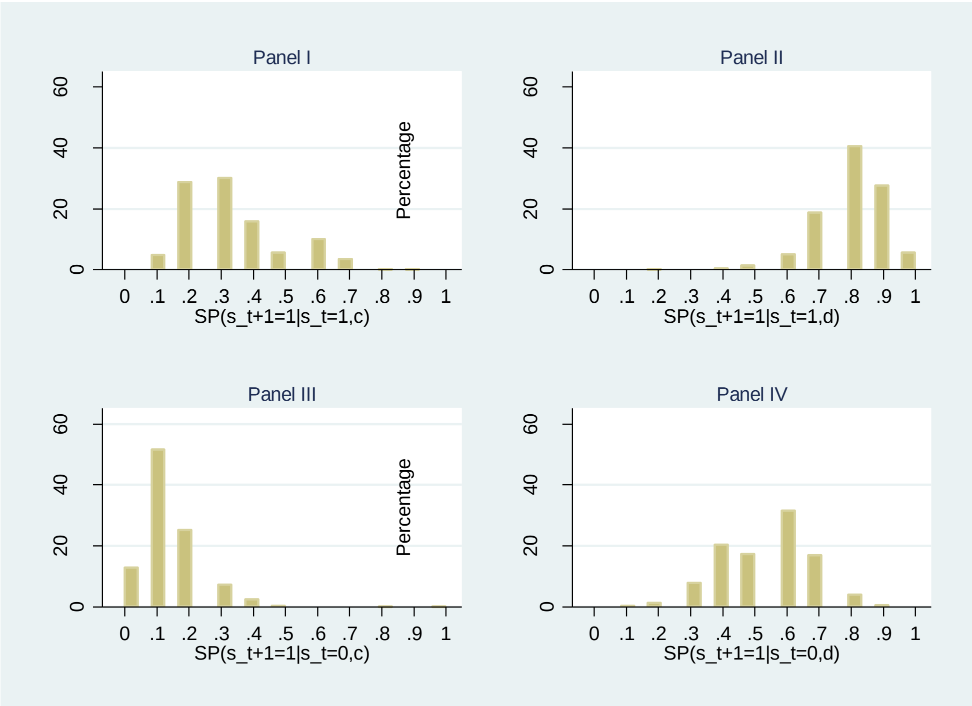


**Figure 2. Distribution of the elicited SPEs**

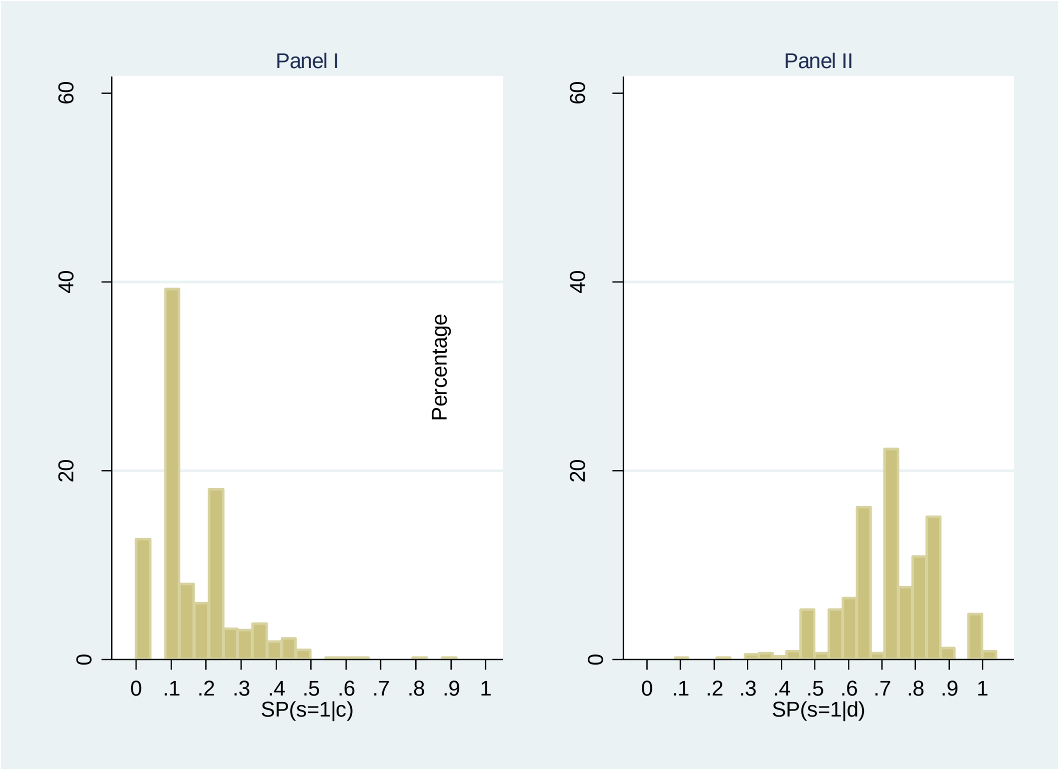


**Figure 3. Distribution of the equilibrium SPEs**

## Appendix

**Table A1. Second stage results with the equilibrium SPE (excluding residuals)**

| | (1a) Coef. | (1b) AME | (2a) Coef. | (2b) AME | (3a) Coef. | (3b) AME |
|---|---|---|---|---|---|---|
| Fraction of days of dirty fuel usage | 3.423*** | 0.515*** | 3.426*** | 0.516*** | 3.424*** | 0.515*** |
| | [.338] | (0.029) | [0.334] | (0.029) | [0.333] | (0.029) |
| $SP(s=1\|c)$ | -0.303 | -0.046 | | | | |
| | [0.635] | (0.095) | | | | |
| $SP(s=1\|d)$ | 0.015 | 0.002 | | | | |
| | [0.687] | (0.103) | | | | |
| $SP(s=1\|c)-SP(s=1\|d)$ | | | -0.141 | -0.021 | | |
| | | | [0.433] | (0.065) | | |
| $SP(s=1\|c)/SP(s=1\|d)$ | | | | | -0.123 | -0.019 |
| | | | | | [0.331] | (0.049) |
| Number of cooks | 0.078 | 0.012 | 0.073 | 0.011 | 0.074 | 0.011 |
| | [0.23] | (0.035) | [0.226] | (0.034) | [0.226] | (0.034) |
| Age | 0.017* | 0.003* | 0.018* | 0.003* | 0.017* | 0.003* |
| | [0.01] | (0.001) | [0.01] | (0.001) | [0.01] | (0.001) |
| Years of schooling | 0.008 | 0.001 | 0.009 | 0.001 | 0.009 | 0.001 |
| | [0.025] | (0.004) | [0.025] | (0.004) | [0.025] | (-0.004) |
| Hindu | 0.071 | 0.011 | 0.073 | 0.011 | 0.075 | 0.011 |
| | [0.242] | (0.037) | [0.243] | (0.037) | [0.238] | (0.036) |
| Housewife | 0.833** | 0.125** | 0.828** | 0.125** | 0.833** | 0.125** |
| | [0.392] | (0.058) | [0.392] | (0.058) | [0.385] | (0.057) |
| Spouse works in informal sector | 0.245 | 0.037 | 0.244 | 0.037 | 0.25 | 0.038 |
| | [0.224] | (0.033) | [0.222] | (0.033) | [0.22] | (0.033) |
| Spouse works in agricultural sector | -0.264 | -0.04 | -0.266 | -0.04 | -0.261 | -0.039 |
| | [0.264] | (0.04) | [0.26] | (0.039) | [0.262] | (0.04) |
| Kitchen located inside dwelling unit | 0.158 | 0.023 | 0.165 | 0.024 | 0.161 | 0.024 |
| | [0.235] | (0.034) | [0.236] | (0.034) | [0.235] | (0.034) |
| Access to ventilation | 0.779 | 0.133 | 0.77 | 0.131 | 0.779 | 0.133 |
| | [0.577] | (0.113) | [0.563] | (0.109) | [0.566] | (0.11) |
| Time to market | 0.005 | 0.001 | 0.005 | 0.001 | 0.005 | 0.001 |
| | [0.009] | (0.001) | [0.009] | (0.001) | [0.009] | (0.001) |
| Expenditure | -0.007 | -0.001 | -0.007 | -0.001 | -0.007 | -0.001 |
| | [0.022] | (0.003) | [0.022] | (0.003) | [0.022] | (0.003) |
| Access to internet | 0.09 | 0.014 | 0.087 | 0.013 | 0.088 | 0.013 |
| | [0.213] | (0.032) | [0.211] | (0.032) | [0.211] | (0.032) |
| Owns television | 0.216 | 0.032 | 0.218 | 0.033 | 0.218 | 0.033 |
| | [0.394] | (0.06) | [0.393] | (0.059) | [0.391] | (0.059) |
| *Log-likelihood* | -151.321 | | -151.37 | | -151.36 | |
| *Pseudo $R^2$* | 0.508 | | 0.508 | | 0.508 | |
| $\chi^2(df)$ | 131.01 | | 134.7 | | 134.63 | |

Note: This table provides estimation results for equation (3), where the dependent variable is the self-reported health status of the individuals (= 1 if the respondent has experienced at least one of the three physical symptoms). The models in this table are restricted versions of those in Table 6 in that the first stage residuals are excluded from a set of regressors. AME denotes average marginal effects. The sample size is 557. ***, ** and * indicate statistical significance at the one, five and ten percent levels, respectively. Standard errors in parentheses are computed by the delta method with bootstrap standard errors for the parameters (number of replications: 500)