Essays on Household Air Pollution in Developing Countries: A Study of Households in an Indian Rural Village

MRIDUCHHANDA CHATTOPADHYAY

(STUDENT ID: 3216A503-1)

Dissertation for Doctor of Philosophy in Economics

Submitted to
Graduate School of Economics
Waseda University
Tokyo

Supervisors:

Prof. Toshi H. Arimura

&

Prof. Hideki Konishi

Contents

List of Figures

List of Tables

Acknowledgement

Since childhood, I had dreamt of being an academician and it was a dream come true for me when I was selected as a recipient of the Japanese Government Monbukagakusha (MEXT) scholarship to do my PhD at the Graduate School of Economics, Waseda University. My journey as a PhD student had been an interesting journey for me and it was almost impossible for me to complete it successfully without the generous support of my supervisors, teachers, and family members. My tenure as a PhD student at the GSE had been academically enriching as I got the opportunity to come in contact with internationally acclaimed professors and their research. I wish to express my earnest gratitude to all the professors and office staff of the Faculty of Political Science and Economics for the support they have extended to me.

I am indebted beyond words to my supervisor Prof. Toshi H. Arimura for all his guidance and support for my academic research. Arimura sensei has encouraged and guided me in every aspect during my tenure as a PhD student under him. I have learnt to be a good academician from Arimura sensei who has always respected my opinions and ideas in research. He has provided me numerous opportunities to participate in workshops and conferences that have enriched me as a researcher. I am short of words to express the amount of generous assistance and support that I have received as his PhD student from him.

I would be failing in my duty if I do not mention the names of Prof. Hajime Katayama (Waseda University), Dr. Hide-Fumi Yokoo (Hitotsubashi University) and Dr. Mari Sakudo (University of Tsukuba) who have guided me in every aspect while conducting this research. I express my respect and gratitude to all of them. I am also highly grateful to my second supervisor, Prof. Hideki Konishi for providing me with all the support that I needed to conduct my research.

It would be wrong on my part if I forget to mention the financial assistance that I have received from JSPS KAKENHI Grant Number 16K13364 and Research Institute of Environmental Economics and Management, Waseda University, Japan for conducting the survey for my doctoral dissertation. I am also thankful to Global Change Program, Jadavpur University, India for providing infrastructural assistance for the survey. I would also take the opportunity to express my gratitude to all my teachers at the Department of Economics, Jadavpur University for nurturing my capabilities as a researcher and guiding me in all aspects since my undergraduate days.

In addition, I would thank all the students in Arimura sensei's research seminar for creating a constructive learning environment in the class. In particular, I would like to thank my friends Guanyu Lu and Naonari Yajima for helping me out in every situation in the last few years.

I am ever grateful to my husband Sumit for providing me with all the necessary emotional support and bearing with me patiently while I am busy with my research in Tokyo. I am indebted for the support that I have received from my mother-in-law. I must mention the contributions of my family members back in India, especially my mother Prof. Madhumita Chattopadhyay, my uncle Prof. Soumyen Sikdar, my aunt Mrs. Paramita Sikdar and my cousins Paramgama and Swarnendu who have always been on my side whenever I have needed them. They deserve all my regards and love.

My father, Prof. Aditya Chattopadhyay is no more with us, but I can feel his presence with me all the time. He would have been the happiest person today, seeing me finally submitting my dissertation thesis. It was his dream to see me getting my doctorate degree from Graduate School of Economics, Waseda University. He was my friend, philosopher, and guide and whatever I am today, it is because of him. I seek his blessings and dedicate my doctoral dissertation to him.

Thesis Abstract

Title of Thesis	Essays on Household Air Pollution in Developing Countries: A Study of Households in an Indian Rural Village

Author's Name	Mriduchhanda CHATTOPADHYAY
Supervisor	Prof. Toshi H. Arimura

Household air pollution (HAP, hereafter) arising from the incomplete combustion of traditional fuels such as firewood, solid biomass fuels and coal, is a salient environmental and health risk particularly in rural areas of developing countries. Despite the alarming health risks from HAP, dirty fuel usage continues unabated particularly in the rural areas of developing countries. In this thesis, we attempt to extend the literature on the reduction of HAP in developing countries in general and rural India in particular. This thesis comprising of three empirical papers, focuses on the economics of HAP with an emphasis on the choice and usage of cooking fuels in rural India. In what follows, we summarize the three chapters of the dissertation thesis along with the publication information for each of them.

Chapter 2: Information dissemination through internet and choice of cooking fuels: A case study of rural Indian households

[Publication information: A preliminary draft of this chapter based on analysis with pilot test data is published as: CHATTOPADHYAY, M., ARIMURA, T. H., KATAYAMA, H., SAKUDO, M., & YOKOO, H. F. (2017). Cooking Fuel Choices—Analysis of Socio-economic and Demographic Factors in Rural India—. *ENVIRONMENTAL SCIENCE*, *30*(2), 131-140. However, the current version of the chapter is a single authored one]

One possible reason behind the unabated use of dirty cooking fuels despite the persistent health hazards from HAP in developing countries may be the knowledge gap. Researchers have identified different sources of information transmission like television, radio and newspaper in developing countries. With the rapid digitalization of services even in rural areas of developing countries, information may disseminate through the access to internet as well. However, the impact of access to information disseminated through internet on cooking fuel choice has not been explored in detail yet. In the second chapter, we try to bridge this gap and try to investigate how the access to information disseminated through internet may affect the likelihood to choose dirty cooking fuels analysing data from 565 rural Indian households.

In estimating individuals' likelihood to choose dirty cooking fuels, one plausible problem may be the endogeneity of 'access to internet'. To address this, we adopt an instrumental variable approach. As an instrument, we have used the 'whether the household is located in an interior region'. Furthermore, to address the feature that the error terms of the fuel choice equation as well as access to internet is jointly distributed, we jointly estimate the two equations using a bivariate probit model. The results of the maximum likelihood estimates suggests that access to internet has a negative and significant association with the likelihood to choose dirty cooking fuels. Testing for the strength of the instrument, we find that our instrument may be a weak one; thus, our results of the maximum likelihood estimation method may suffer from the caveat of weak instrument bias. Therefore, a better and more reliable approach is to use propensity score matching (PSM) approach to test the causal relationship between access to internet and cooking fuel choice and thereby address endogeneity arising from unobservable confounding factors.

Results from the PSM analysis suggests that households with access to internet are approximately 24% less likely to choose dirty cooking fuels in compared to matched control groups (households without access to internet). Furthermore, sensitivity analysis based on bounding approach by Rosenbaum (2002) suggests that, this negative causal relationship between access to internet and likelihood to choose dirty cooking fuels is valid till the effect of hidden bias from the unobservable confounders cause a 60% increase in the odds of assignment to the treatment group compared to the control group.

Chapter 3: Subjective Probabilistic Expectations, Household Air Pollution, and Health: Evidence from cooking fuel use patterns in West Bengal, India

[co-authored with Prof. T.H. Arimura, Prof. H. Katayama, Dr. M. Sakudo, Dr. H.F. Yokoo.; Publication Information: This chapter is published as Chattopadhyay, M., Arimura, T. H., Katayama, H., Sakudo, M., & Yokoo, H. F. (2021). Subjective Probabilistic Expectations, Household Air Pollution, and Health: Evidence from Cooking Fuel Use Patterns in West Bengal, India. *Resource and Energy Economics*, 66, 101262]

An increasing number of empirical studies have investigated the determinants of cooking fuel choice in developing countries, where health risks from household air pollution are one of the most important issues. Despite much evidence in the literature, there remains some unexplored aspect of the choice of household cooking fuel, specifically the role of expectations about health risks from HAP. In the third chapter of this dissertation, we contribute to this stream of literature by exploring the association between of individuals' subjective probabilistic expectations (SPEs) on cooking fuel usage pattern. We also explore how this pattern is associated with individuals' health status. We analyse a unique dataset on individuals' SPE elicited in probabilistic form from 557 survey respondents in rural India. To elicit the individuals' SPEs about the health risks related to HAP, we have adopted an interactive elicitation method using visual aids. A potential problem in estimating the health status of the individuals is the endogeneity of the 'cooking fuel usage pattern' variable. To address this, we adopt an instrumental variable approach. As an instrument, we have used the 'opportunity to access cooking fuel for free' in our analysis. Further, to address the non-linearity associated with binary health status as well as the fact that cooking fuel usage pattern variable lying in the unit interval between 0 and 1, we have adopted the Two Stage Residual Inclusion (2SRI) model. Our results suggest that individuals' SPEs of becoming sick from dirty fuel usage are negatively and significantly associated with the fraction of days of dirty fuel usage in households. Concurrently, dirty fuel usage and self-reported health status of the individual being sick are also significantly correlated. We then conduct a policy simulation of information provision regarding the health risks of dirty 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, these changes may be small in size.

Chapter 4: Economics of clean air: Valuation of reduced health risks from Household Air Pollution - A study of rural Indian households

[Publication information: The version of the chapter submitted for interim reporting of PhD is published as a discussion paper in the WINPEC Discussion Paper series: Chattopadhyay, M. (2021). *Economics of clean air: Valuation of reduced health risks from Household Air Pollution-A study of rural Indian households* (No. 2119). This chapter is a single authored one]

Although the health risks from HAP is salient in developing countries, it can be adequately prevented. To overcome the numerous logistic challenges that may arise during the implementation of interventions to reduce HAP, it is necessary understand the attitude and/or preference of the potential beneficiaries towards such mitigations. The attitude of the individuals towards the reduction of HAP may be studied by understanding the individuals' valuation of the reduced health risks (alternatively, perceived private health benefit) from the same. In the fourth chapter of the dissertation, we attempt to assess the individuals' valuation of reduced health risk from HAP exclusively, derived from a hypothetical improvement in household air quality using a stated preference method. In particular, we contribute to the literature on the economics of HAP by estimating the individuals' willingness to pay for reduction (WTP) in health risks related to HAP using a double bounded dichotomous choice (DBDC) approach. Concurrently, as an extension of estimating the individuals' WTP for reduction in HAP in this study, we attempt to explore the presence and source of starting point bias in our DBDC model. Using a unique contingent survey of 557 respondents in rural India, we estimated the mean annual WTP for the reduction in HAP to be INR 886.59 (~ USD 14.30) from the DBDC model, which accounts for approximately 1.06% of the annual household expenditure. Although we have conducted the study in a different time and with a different sample, the ratio of our estimated WTP to average household expenditure lies in a comparable range with previous literature. In conformation with literature on valuation of other goods and resources, our result shows that the estimated WTP is also lower than the individuals' expenditure on cooking fuels as well as switching cost to switch to cleaner fuels. Furthermore, our analysis suggests the presence of anchoring effect that validates the presence of starting point bias in our DBDC model. We also find that the estimated mean WTP is sensitive to several health and non-health factors. This exercise further enables us to recommend policy prescriptions like generating public awareness about HAP as well as, targeting potential beneficiaries based on observable characteristics to ensure smooth implementation and effectiveness of intervention programs to reduce HAP.

Chapter 1: Introduction

1.1.Background

Household air pollution (HAP, hereafter) arising from the incomplete combustion of dirty cooking fuels[1] such as firewood, solid biomass fuels (agricultural crop residue and animal dung cakes) and coal continues to be a global health threat (WHO, 2018). Incomplete combustion of such fuels in traditional or high efficiency stoves yields high levels of pollutants such as benzene, formaldehyde, and polycyclic aromatic carbons (American Lung Association, 2011). Alarmingly, the particulate concentrations in kitchens in developing countries often exceeds the prescribed levels mentioned in the guidelines (Duflo et al., 2008). For example, the mean 24-hour $PM_{2.5}$ concentration in kitchen area of households using solid fuel in India is around 609 mg/m3 (Balakrishnan et al., 2013). Analogous evidences are also observed by Dasgupta (2004) and Zhang and Smith (2007) in other developing countries as well. The exposure to HAP is particularly high among women and young children who spend majority of their time near the domestic hearth (e.g., Jeuland et al., 2015).

A causal relationship between various health risks including cardiovascular, respiratory, vision-related diseases is well-established in literature (Smith and Pillarisetti, 2017). Estimates from WHO (2018) demonstrates that globally, 3.8 million annual premature deaths are caused due to HAP of which 6% are in developing countries (Ritchie and Roser, 2020). In addition, HAP may have a long-term effect on general well-being as, early exposure to HAP during early childhood may have a negative impact on lung development (Duflo et al., 2008). Evidence of the association of HAP with infant mortality as well as, birth outcomes particularly in developing countries is also available in literature (e.g., Franklin et al., 2019).

[1] Based on the emission of smoke during combustion, cooking fuels are classified as clean and dirty. Fuels such as electricity and LPG are classified as clean cooking fuels as they emit low or no smoke during combustion. On the contrary, fuels such as firewood, solid biomass fuels and coals are often referred to as dirty cooking fuels due to their high smoke emission during combustion. Although WHO (2018) has classified kerosene to be a dirty cooking fuels, we have addressed it as clean cooking fuels following Duflo et al (2008) among others, for reasons mentioned in the next chapters.

Despite these alarming figures, approximately 3 billion people globally continue depending on dirty cooking fuels to meet their regular household energy demand (WHO 2018), majority of them residing in India, China, and sub-Saharan African countries (Bonjour et al., 2013). Furthermore, around 2.8 billion households (among which 0.5 billion resides in urban areas) have reported to find commercial clean fuels to be expensive or irregularly supplied, thus making them less attractive to use (e.g., Grieshop et al., 2011). Without dramatic changes through policy interventions, the number of individuals relying on dirty cooking fuels is expected to remain roughly same through 2030 in developing countries (IEA 2017).

1.2.Research objective and contributions of the study

In the above background, our thesis focuses on the economics of HAP with an emphasis on the choice of cooking fuels in rural India. The households can make a choice regarding the cooking fuels as well as, can have a perception of the externality caused from HAP. Hence, the economics of this environmental problem can be perceived from the idea that individuals are expected to derive a negative utility from HAP, and they make choices about reducing it. In this thesis, we attempt to extend the literature on the reduction of HAP in developing countries in general and rural India in particular.

Apart from the different economic, demographic, accessibility and supply-side factors, there is a growing consensus among researchers that information provision and/or access to information can be a key determinant of cooking fuel choice (e.g., Zahno et al., 2020). With the rapid digitalisation of services and easy access of internet through smartphones in developing countries, information may disseminate through the access to internet as well. To the best of our knowledge, there is no evidence in literature that such access to information through internet may have some impact on individuals' cooking fuel choice. We try to contribute to the literature by examining the following hypothesis in the second chapter of our dissertation thesis: *Access to information disseminated through internet is likely to reduce the likelihood to choose dirty cooking fuels*.

In the literature of the determinants of cooking fuel choice, there remains an unexplored aspect, specifically, the role of expectations about health risks from HAP. Without precise estimates of and/or information about the health risks related to HAP, it becomes difficult for policy makers to infer what actually motivates the choice – preference or expectation as different combinations of the two may lead to same observed choice. In the third chapter of the thesis, we attempt to fill this gap in the literature by analysing the role of individuals' subjective probabilistic expectations (SPEs) of becoming sick with HAP-related physical symptoms on individuals' cooking fuel usage pattern and health. Specifically, we address whether individuals' subjective probabilistic expectations have any role in individuals' cooking fuel usage patterns and in turn whether usage patterns are associated with individuals' health. If so, we may conclude that, SPEs have a direct association with cooking fuel usage patterns and an indirect association with individuals' health.

In the fourth chapter, we attempt to assess the individuals' valuation of reduced health risk from HAP, derived from a hypothetical improvement in household air quality using stated preference method. Compared to the literature on environmental valuations, research on the valuation of economic cost of HAP is relatively small (Jeuland et al., 2015). We contribute to the literature on the economics of HAP by estimating the individuals' willingness to pay for reduction (WTP) in health risks related to HAP using a double bounded dichotomous choice (DBDC) approach. Concurrently, as an extension of estimating the individuals' WTP for reduction in HAP in this study, we attempt to explore the presence and source of starting point bias in our DBDC model.

1.3. Survey site

India continues to be one of the major hotspots of HAP accounting for approximately 28% annual deaths among developing countries due to HAP (Rohra and Taneja, 2016). Among the different states of India, the problem of HAP is quite acute in West Bengal causing approximately 39000 HAP-related deaths in 2019 (Balakrishnan et al., 2019). Although the penetration of LPG in West Bengal is around 46%, it is mainly restricted in the urban areas (MPNG, 2016). Incidentally, in comparison to over-all India (83.8%), rural areas of West Bengal show a high incidence of dirty fuel usage (92.8%) (NSSO, 2015)[2]. This indicates that usage of dirty cooking fuels is quite prevalent in rural regions of West Bengal, thus, motivating our research problem and design. For our analysis, we have selected the villages under the Dhapdhapi II village council in the state of West Bengal to be our research site. Our survey site is located in the south-eastern part of the state of West Bengal and has a humid tropical climate.

Dhapdhapi II-village council comprises seventeen villages with majority of the respondents continuing to use traditional cooking fuels although the networks for LPG distribution is active. Located at a distance of 35 – 40 kilometres from the state capital, Kolkata, the survey site has a total adult population of 13024 individuals as of January 01, 2016. Due to its proximity to the metropolis, these villages have easy access to modern amenities but simultaneously retain the typical traits of a rural area in any developing country. The location of the household in the map of India is shown in Figure 1.1.

<Figure 1.1 approximately here>

The households in the survey region are electrically grid connected however, frequent power cuts and low voltage hinder using electricity as cooking fuels. In the survey site, one can switch to clean cooking fuels by paying a fixed cost but having the option to collect cooking

[2] West Bengal is among the top five states in India in terms of dirty cooking fuel usage in rural areas (NSSO, 2015)

fuels for free (or at a nominal cost), makes switching to clean cooking fuels less attractive (Yokoo et al, 2020)[3]. In addition, the flagship program of Government of India – Pradhan Mantri Ujjwala Yojana (PMUY), to enable low-cost access to LPG for all, has been implemented in the survey region from May 2016 (MPNG, 2016) thus ensuring access to modern cooking fuels. Another option to reduce the health hazards from HAP is exposure reduction, for example by cooking outdoors. The rate of outdoor cooking varies tremendously between countries as well as between rural and urban areas (Langbein et al., 2017). Although not absent, the practice of outdoor cooking is extremely nominal in our survey region[4].

1.4. Survey design

We have collected the data from the survey site in two rounds with a gap of one year in between during December 2016 to January 2017 and December 2017 to January 2018. This survey was financially supported by JSPS KAKENHI Grant Number 16K13364 and Research Institute of Environmental Economics and Management, Waseda University, Japan. Global Climate Change, Jadavpur University, India provided infrastructural assistance for the survey.

We have selected the individuals who are primarily responsible for cooking in the households as our respondent; consequently, all of them were female based on traditional norms and practice in rural areas of developing countries. Due to the cultural norms and beliefs, it was less likely that the respondents would share their responses freely (or would be allowed to respond by household members) with an unknown researcher. Hence, we recruited a team of six enumerators with undergraduate degree as survey assistants, of whom three were local

---

[3] Yokoo et al. (2020) in our concurrent work has observed that in the survey site, during the time of the survey, one needed to pay approximately INR 5000 (~ USD 80) as fixed cost to switch to clean cooking fuels such as LPG. Evidence from Kar et al. (2019) further substantiates similar amount (INR 5026) for fuel switching using data from other Indian states.

[4] In our sample only about 2% of the households have reported to practice of cooking outdoors and that too only in winter months of the year.

residents. Since we had local individuals in the enumerator team, getting access to the local community where our respondents resided, was relatively easy. To reach the respondents more efficiently as well as ensure a high response rate, the survey was conducted in a door-to-door interview method and the response was recorded in pen and paper. The data collected was later coded and fed into the Excel sheet for analysis.

We have used a stratified random sampling method to choose 600 respondents among the 2286 households listed on the electoral roll of Dhapdhapi-II village council, which is publicly available online on the website of the Election Commission of India. We have used stratified random sampling method for its relative efficiency over other probability and non-probability sampling methods (Buddhakulsomsiri and Parthanadee, 2008). A 'part' served as our stratification unit. A 'part' is a stratification unit within each electoral constituency and in rural areas, each part roughly corresponds to one village. The seventeen villages under the Dhapdhapi II village council are classified into 16 parts. As the population size under each part was not uniform, we have selected the sample size in each part proportional to the corresponding population size.

The survey questionnaire included questions on individuals' cooking fuel usage behaviour, health status, subjective probabilistic expectations regarding the health risks related to HAP, valuation of risk related to HAP apart from the common socio-demographic and economic variables[5]. The survey questionnaire was developed based on previous studies (e.g.: Hanna et al., 2016; Delavande and Kohler, 2016; NSSO, 2015). The survey questionnaire was composed in English and later translated to Bengali, a native language of the respondents before the implementation of the survey. Before the pilot survey, the enumerators attended training sessions for a week to be familiar with the purpose of the survey as well as how to conduct the survey.

[5] Relevant parts of the questionnaires are attached as appendix at the end of the chapters.

The questionnaire was tested through a pilot survey among 70 households in the survey site chosen separately using the aforementioned sampling method in August 2016. To account for the fact that some of the respondents may not be literate, the enumerators had read out the information sheet and questions in Bengali, responding to any queries or questions. The respondents were informed clearly about the affiliation and the purpose of our survey before interviewing and we received signature or thumbprint (if the respondent was illiterate) showing consent from the respondents who participated in the survey. We also informed the respondents that they were allowed to terminate the survey at any point of time during the interview for any reason. No compensation of any kind was provided to any participants during or after the survey. Anonymity of the respondents were maintained during the survey as well as analysis and we did not collect any personal identifying information during the survey. We strictly abided by these guidelines and followed the same procedures while conducting both the rounds of the main survey as well.

During the pilot test, the respondents took around 30 – 40 minutes to answer to all the questions. Following the preliminary survey, the questionnaire was revised accordingly, and the enumerator team was trained with the revised questionnaire for three days. The first round of survey was implemented in December 2016 – January 2017.

Our team of enumerators visited the selected 600 sample households and elicited the information, thus ensuring a high response rate (99%). The respondents took approximately 20 – 25 minutes to respond to the survey. Out of the total 600 households, only four households did not cooperate and refused to complete the survey; hence their responses were excluded. Thus, our sample size at the end of first round of survey was 596.

With a gap of one year, the same enumerator team made a repeat visit to the same sampled households interviewed in the first round. They were able to elicit the responses from 588 out of the 596 original respondents (representing an attrition rate of 1.34%). Thus, our sample size at the end of the second round is 588. Figures 1.2 to 1.4[6] present how the survey was conducted.

<Figure 1.2 approximately here>

<Figure 1.3 approximately here>

<Figure 1.4 approximately here>

1.5. Fuel market and institution in rural India

According to micro-economic theories, under normal functioning of domestic fuel markets, price adjustments allow to equate demand with supply, leading to market clearance (Gupta and Köhlin, 2006). This ultimately determines the choices among the options for domestic fuels in the households. However, this theoretical approach does not hold true in reality particularly in rural areas of developing countries. This may be due to the following two reasons- first, presence and influence of public sector as seller of domestic energy and second, barriers to access of cooking fuels.

Presence and intervention of public sector in the fuel market implies that the supply and price of a certain type of cooking fuel is controlled by the relevant department of the Government of India. For example, the provision and price of LPG is controlled by the Ministry of Petroleum and Natural Gas. For domestic purposes, LPG is usually marketed in cylinders of 14.2Kg by government-controlled companies like Indian Oil Corporations or Bharat Petroleum[7]. LPG is available at a subsidised price while the subsidy varies across months depending on the international market price of LPG.

[6] To maintain the anonymity of the respondents, we have blurred their faces in Figure 1.2 to 1.4..
[7] During the time of the survey, the price of the LPG cylinders ranged from INR 605-766 (USD 9-11 (Indian Oil Corporation, 2021).

Since 2015, the Government of India has adopted three major policies targeted towards promoting LPG usage particularly among poor and rural households where, switching to clean cooking fuels is financially challenging. These policies include (a) the Pahal that directly transfers the subsidies directly to individuals' bank account, to have transparency in the operation; (b) Give it Up that enables the households with annual income more than INR 1 million to voluntarily surrender the LPG subsidy and transfer it to poor households; (c) Pradhan Mantri Ujjwala Yojana (PMUY) that aims to provide free LPG connections to poor households (Khan, 2017). As of March 2019, around 10 million households have voluntarily surrendered the LPG subsidy under Give it Up scheme while, around 270 million households have benefited from the Pahal scheme with around INR 1355 billion (USD 19 billion) transferred to their bank accounts (PIB, 2019).

Apart from being affected by price regulations, the market for LPG may suffer from uncertainty in availability and supply side shocks particularly in rural areas mainly due to storage and uncertainty in transportation. Although one can avail the LPG through online or over-phone booking, the time from booking to delivery of the cylinder may take one or two weeks. This often results in using LPG as the secondary fuel in the rural areas with the primary continuing being the firewood. For instance, Gould and Urpelainen (2018) have observed that fewer than 60% of the LPG users consider it as their primary source and only 4% of them use it exclusively in their study of rural India.

In rural areas, one can access cooking fuels without incurring any monetary cost in spite of the presence of open markets. 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 (Das and Srinivasan, 2012). This easy availability of firewood is one of the predominant reasons for using it as primary cooking fuel in the rural areas.

Apart from the two major cooking fuels namely, LPG and firewood, other available cooking fuel options in rural India include kerosene, coal and solid bio-mass fuels like agricultural crop residue and cow dung cakes. Of late, the percentage of households using these fuels as their primary sources of cooking has considerably diminished. These fuels are either used as supplementary source of fuels, to keep the fire burning or to initiate a higher temperature (Gupta and Köhlin, 2006).

Similar to LPG, the pricing of kerosene and coal are also determined by the relevant ministries. While the supply of kerosene is controlled by Ministry of Petroleum and Natural Gas, coal for domestic purpose is available from the coal fields owned by central public sector undertakings under the direct control of Ministry of Coal and Mines. Both kerosene and coal are available in the rural areas through the Public Distribution System (PDS). Each household is entitled to purchase a fixed amount of kerosene (coal) per week from the PDS by showing their 'ration card' that stands as an identification for their residence and nationality[8]. Recently, in rural areas of India, there is an open market as well as a black market for kerosene[9]. Unlike kerosene or coal, solid bio-mass fuels are quite popular among the rural households due to ease of availability as well as low or no cost from farms and domestic animals.

[8] Initially, households that lie below the poverty line are provided with a 'special ration card' by the government that enables them to purchase kerosene and coal from the PDS at a subsidized rate. Since 2017, these households have to purchase kerosene from PDS at a non-subsidised rate and the applicable subsidy is directly transferred to the bank account of the beneficiaries to ensure better subsidy management (MPNG, 2016)

[9] The kerosene that is sold in PDS is a blue liquid while that sold in open market is a colourless liquid.

Figure 1.1. Location of the survey site in the map of India

[source: https://geology.com/world/india-satellite-image.shtml (accessed on March 21, 2021)]

Figure 1.2. Conducting the field survey

Figure 1.3. Conducting the field survey

Figure 1.4. Conducting the field survey

Chapter 2: Information dissemination through internet and choice of cooking fuels: A case study of rural Indian households

2.1 Introduction

Household Air Pollution (HAP, hereafter) caused primarily by the incomplete combustion of dirty cooking fuels, is a global health threat; accounting for approximately 3.8 million annual deaths (WHO, 2018). In particular, the developing countries suffer from exposure to HAP, with approximately 3.7% of the loss of disability-adjusted life years (Duflo et al., 2008). Despite this alarming health risks, dirty fuel usage continues unabated in low-income households, particularly in rural areas.

It is therefore necessary to understand why a large population of developing countries continue to use such dirty cooking fuels. Empirical evidences suggest that individual-specific factors such as income (e.g., Heltberg, 2005) and education (Farsi et al., 2007), are likely to influence individuals' cooking fuel choice. Jeuland et al (2015) also point that supply side factors such as relative cost advantage and ease of availability may also influence individuals' cooking fuel choice in developing countries.

This study tries to extend the above line of literature on cooking fuel choice in developing countries. In particular, this study attempts to investigate the key determinants of cooking fuel choice focussing on the access to information disseminated through internet

Evidence from literature including both observational and intervention studies, show ambiguous impact of information access and/or provision on the choice of clean cooking fuels in developing countries, particularly in rural areas. For example, Dendup and Arimura (2019)

Publication information: A preliminary draft of this chapter based on analysis with pilot test data is published as: CHATTOPADHYAY, M., ARIMURA, T. H., KATAYAMA, H., SAKUDO, M., & YOKOO, H. F. (2017). Cooking Fuel Choices—Analysis of Socio-economic and Demographic Factors in Rural India—. *ENVIRONMENTAL SCIENCE*, *30*(2), 131-140. However, the current version of the chapter is a single authored one.

in their observational studies, have established that better access to information is likely to enhance the chance of adoption of clean cooking fuels in rural Bhutan. Similarly, Zahno et al. (2020) through their randomised control study, suggest that information provision in form of health messaging is effective to promote clean cooking fuels in rural India. Contrarily, Jeuland et al (2014b) in their discrete choice experiment, have observed limited impact of information on adoption of clean cooking fuels/technology in rural India.

Dendup and Arimura (2019) in their study have considered the access to information disseminated through television. However, with the rapid digitalization of services even in rural areas of developing countries, information may disseminate through the access to internet as well. To the best of our knowledge, there is no evidence in literature that such access to information may have some impact on individuals' cooking fuel choice. This study attempts to bridge this gap in the literature by hypothesizing that better access to information disseminated through internet may reduce the likelihood to choose dirty cooking fuels. In particular, this chapter extends the study of cooking fuel choice by Chattopadhyay et al (2017) in rural India[1] but here we address the issue of potential endogeneity that may arise in estimating individuals' fuel choice.

We analyse a unique dataset of 565 individuals from rural Indian households. In estimating individuals' likelihood to choose dirty cooking fuels, one plausible problem may be the endogeneity of 'access to internet'. To address this, we adopt an instrumental variable approach. As an instrument, we have used the 'whether the household is located in an interior region'. Furthermore, to address the error terms of the fuel choice equation as well as access to internet is jointly distributed, we jointly estimate the two equations using a bivariate probit

[1] Chattopadhyay et al. (2017) have examined the key determinants of cooking fuel choice using 68 sample households in rural West Bengal, India. They find that in conformation to literature, several individual-specific covariates such as income, years of education affect the likelihood to choose clean cooking fuels.

model. The results suggests that access to internet has a negative and significant association with the likelihood to choose dirty cooking fuels.

Testing for the strength of the instrument, we find that our instrument may be a weak one; thus, our results of the maximum likelihood estimation method may suffer from the caveat of weak instrument bias. Therefore, a better and more reliable approach is to use propensity score matching (PSM, hereafter) approach to test the causal relationship between access to internet and cooking fuel choice and thereby address endogeneity in the model arising from unobservable confounders. Results from the PSM suggest that household with access to internet tend to have around 24 percentage points lower likelihood to choose dirty cooking fuels compared to matched control group. Sensitivity analysis further indicates that this significant negative effect remains valid even if the unobserved confounders cause an additional 60 percent increase in the odds of assignment of individuals to the treatment group compared to the control group.

The remainder of the chapter is organized as follows. Section 2 describes the survey methodology and variables considered for the study along with their summary statistics. The next section presents our empirical model and results. Results of the sensitivity study is also presented here. Section 4 concludes by discussing the directions of future research.

2.2. Introduction to dataset

2.2.1. Survey design

This study uses the data collected during December 2016 to January 2017. Out of the 600 households randomly chosen in our survey area four households refused to cooperate and hence are not surveyed. For this analysis, we exclude from the sample the respondents who have no spouse or provide no information of spouse. Thus, our effective sample size reduces to 565.

Table 2.1 presents the descriptive statistics of the variables used in the study which is explained in detail in the following section.

<Table 2.1 approximately here>

2.2.2. Description of the variables and their summary statistics

Cooking fuel choice

Following previous studies (e.g.: Duflo et al., 2008), we have considered two broad categories of cooking fuels – clean and dirty, as our variable of interest. For this purpose, we create an indicator variable that takes the value unity if the households are primary dirty cooking fuel users[2]. Indeed, the proportion of dirty cooking fuel users are found to be quite high; around 77 percent of the respondents have reported to use dirty cooking fuels as their primary cooking fuels (see Table 2.1).

In order to understand how the choice of cooking fuels correspond to household expenditure (implicitly income), the proportion of primary clean (dirty) fuel users is plotted against expenditure deciles in Figure 2.1. As seen in Figure 2.1, the proportion of clean (dirty) fuel users increases (decreases) steadily after the income passes a certain threshold level (beyond income decile 7). Intuitively, this suggests that with higher income, individuals tend to switch to clean cooking fuels as their primary fuels.

[2] Although WHO (2018) classifies kerosene to be dirty cooking fuel, we have implemented our survey before this classification was published. In our study conducted prior to 2018, we have consistently followed the nomenclature referred by Duflo et al. (2008) among others, where kerosene is classified as clean cooking fuel. However, we have tried to emphasize that such constraint would lead to a negligible bias in this context, as evident from the following discussion .In our effective sample, there is only one sample unit that uses kerosene as the primary cooking fuel. Even if, we include kerosene to be dirty cooking fuel, the proportion of individual who are primarily dirty fuel user does not change much (mean changes from 0.77 to 0.78 and standard deviation changes from 0.416 to 0.417). Furthermore, we have found that the association between the variable "primarily dirty cooking fuel user" once including kerosene and once, excluding kerosene, using Pearson's correlation coefficient which is obtained to be 0.995. The large sample test using z-transformation (Fisher, 1921) suggests that an extremely significant and almost unity correlation coefficient between the two versions of the two variables (test statistics: 8.246 ($p < 0.001$)). In other words, these two series may be considered to be almost same, and this further indicates that even if there is any bias arising from not including kerosene in dirty cooking fuels, this bias is expected to be statistically insignificant.

However, the relationship seems to be fluctuating in the lower six income deciles. For example, the proportion of primary clean (dirty) users decreases (increases) in the first two income deciles followed by an increasing (decreasing) trend in next two deciles, again showing a declining (ascending) trend in the fifth and sixth decile. This non-monotonic trend in both primary clean fuel users as well as dirty fuel users across different income deciles may suggest the relevance of other actors such as preferences that may influence the individuals' cooking fuel choice decision[3].

<Figure 2.1 approximately here>

Access to information

There may be various channels through which the households may have access to information about health risks related to HAP like newspaper, television, radio, and internet. In order to promote digital literacy and faster access to information in India particularly in rural areas, the Government of India has launched the flagship programme of Digital India in 2015 and another program in association with Google India, Tata Trust and Intel India targeted for women (Paul et al., 2017). As on December 2015, approximately 87 million rural India households have access to internet (IMAI, 2015). Of late the Government of India has initiated an internet-based application programme known as e-Gram Swaraj portal and a mobile application to conduct and coordinate administration at the village level (MoPR, 2020). This provides us with the rationale to choose the access to internet to be the route of dissemination of information in this study.

[3] Another reason behind this fluctuating behaviour in the lower income deciles may the relatively small sample size used in the study in comparison to other literatures using national level sample (e.g., Farsi et al., 2007).

We therefore asked the respondents the following binary question: "*Do you have access to internet*?". Seventeen percent of the respondents are found to have access to internet in the sample (see Table 2.1).

Other covariates

Individuals' cooking fuel choice may depend on several other individual- and household-specific covariates apart from access to information disseminated through internet. Therefore, in our analysis, we control for a set of factors including number of cooks (surrogate for household size), respondents' age and years of schooling[4], dummy for the occupation of the respondent (respondent is housewife), dummy for the household decision-maker (respondent is household decision-maker), dummies for the occupation of the spouse (spouse works in informal sector and that in agricultural sector), dummy for religion (respondent is Hindu), time needed to reach the nearest motorable road on foot (in minutes) (surrogate for accessibility to cooking fuels), dummy for access to electricity (household has access to electricity), total monthly household expenditure (surrogate of household income, as already mentioned), dummy for participation in microfinance (respondent has participated in any micro financial scheme), dummy for ownership of land (respondent household owns land), dummy for ownership of ration card (surrogate of access to public distribution system; respondent owns ration card), dummy for owning cards identifying them to lie below poverty line (BPL)[5]

[4] Literacy have been considered to be an indicator of awareness (Malla and Timilsina, 2014) and hence, it is likely to influence individuals' cooking fuel choice. In developing economies particularly in rural areas, females are exposed to formal education in their early years of life. However, owing to several compelling factors, majority of them are forced to withdraw before completing the primary level (McConnell and Mupuwaliywa, 2016). To accommodate this feature in our study, we have elicited individuals' year(s) of education directly.

[5]BPL refers to Below Poverty Line households. The poverty line for rural West Bengal based on capita consumption expenditure per month is fixed at INR 783 (Reserve Bank of India, 2012) where USD 1= INR 68 (average of the average exchange rate in December 2016 and January 2017).

(respondent household owns card that identifies them to belong to below poverty line) and dummy for the ownership of television (household owns television).

2.3. Estimation

2.3.1. Estimation model and results

Estimation model

To quantitatively identify the role of internet on individuals' cooking fuel choice decision, we estimate the fuel choice on other covariates including internet access using the data collected in the survey. The theoretical underpinning of the problem under consideration essentially follows from the random utility theory (McFadden, 1975).

As a benchmark, we assume that the individual i's underlying utility (U_i^*) from choosing a particular cooking fuel is unobservable and depends on access to internet ($internet_i$), household expenditure ($expen_i$), years of schooling ($school_i$) and a vector of other individual and household characteristics ($\boldsymbol{X}_i$):

$$U_i^* = \beta_0 + \beta_1 internet_i + \beta_2 expen_i + \beta_3 school_i + \boldsymbol{X_i'\gamma} + u_i \qquad (2.1)$$

where, u_i is an idiosyncratic term uncorrelated with $expen_i, school_i, \boldsymbol{X_i}$ and is assumed to be normally distributed with mean 0 and variance σ_u^2.The observed cooking fuel choice of the individual i ($fuel_i$) is an indicator variable that takes the value unity if he/she chooses dirty cooking fuel. We assume that $fuel_i$ and U_i^* are associated in the following manner: $fuel_i = 1\ if\ U_i^* \geq 0$ and $fuel_i = 0\ if\ U_i^* < 0$.

One concern regarding the above model is the possibility of endogeneity problems. In our sample, whether the households have access to information is not randomly assigned, hence the exogeneity assumption of the variable 'access to internet' may be questioned. For example, the presence of unobservable confounders such as preference for technology may be associated with access to internet leading to the endogeneity issue.

To address the possible endogeneity of ‘access to internet’, we adopt IV method. We propose that individuals’ underlying decision to adopt internet ($internet_i^*$) is latent:

$$internet_i^* = \delta_0 + \theta_1 z_i + \delta_1 expen_i + \delta_2 school_i + \boldsymbol{X_i'\tau} + v_i \tag{2.2}$$

and $internet_i$ and $internet_i^*$ being associated in the following manner: $internet_i = 1\ if\ internet_i^* \geq 0\ and\ internet_i = 0\ if\ internet_i^* < 0,$ and z_i is a suitable instrumental variable. Furthermore, unobservable (e.g.: preference for technology) may affect both the individuals cooking fuel choice as well as access to internet. This leads us to assume that the error components of equations (2.1) and (2.2) are jointly distributed as $\begin{bmatrix} u_i \\ v_i \end{bmatrix} \sim N_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \rho\sigma_u\sigma_v \\ \rho\sigma_u\sigma_v & \sigma_v^2 \end{bmatrix}\right)$. This formulation calls for the joint estimation of parameters through the maximum likelihood estimation method.

As our instrument, we construct a variable based on the following question: “How much time in minutes you need to reach the market from your house by walking?”. Based on this information, we create an indicator variable $interior_i$ that takes the value one if the time taken to reach the market is more than the average time taken by all the samples[6]. We interpret this variable as representing geographical dispersion of the location of the households or, precisely, whether households are located in an interior region. It is expected that higher geographical dispersion of the location of the households may result in positioning the household beyond

[6] Ideally, whether or not the household is located in an interior region should be constructed based on the information from the distance to the mobile tower in the context of access to internet. However, based on the information collected from survey (verified from the satellite image from Google maps), there is no mobile towers in the periphery of our survey site. Therefore, we chose market which is the most centrally located area to construct our instrument.
In our survey site (Dhapdhapi II village council), there is only one market (Dhapdhapi Bazar) which is same for all the residents. It is located in the most central area in the village where all the individuals come to purchase their daily consumption goods and services. The average time required to reach the market is approximately 18 minutes.

the network coverage area in developing countries, thus reducing internet accessibility[7]. Therefore, our instrument is expected to be relevant in that it is negatively and significantly associated with $internet_i$.

The exclusion restriction requires that $interior_i$ is uncorrelated with the unobserved individual-specific component of $fuel_i$, while affecting it only through $internet_i$. A potential threat to this exogeneity assumption may arise as households located in an interior region may have some association with the degree of health risk tolerance, which in turn may influence the cooking fuel choice. This factor may invalidate our instrument exogeneity assumptions by creating a correlation between the error term and the instrument. In particular, individuals with high risk tolerance, tend to undermine the health risk from HAP and prefer to use dirty fuels more than individuals with lower risk tolerance; thus, choosing a more interior region to live where dirty cooking fuels is easily available.[8]

Nonetheless, this threat does not appear to be plausible given that our research site is located in rural India. The patriarchal social structure and the patrilocal residence system prevalent in India offer women having a limited (if not, no) decision making capacity on marriage decisions including the choice of spouses and in-laws. Hence, they have a limited or no-say over the choice of the location of the household. Furthermore, male dominance in household money management as well as, purchase decisions in non-metropolis households further offers no chance to married women to choose location of the house even if it is purchased after marriage (Singh & Bhandari, 2012). Since the respondents in our sample are essentially married women, they are unlikely to have a say or choice about the location of households, regardless of their preference.

[7] Correa et al. (2017) provide empirical evidence for this finding in their study in Chile.

[8] In order to control for the accessibility of the cooking fuel, we have included the variable 'time to reach the nearest motorable road on foot'. There is a possibility that the selected instrument, $interior_i$ and access to cooking fuel together when included in the model may result in multicollinearity particularly in the estimation of equation (2.2). In this regard, it may be noted that all the values in the VIF matrix on the basis of our data lies in the range of 1.0 to 1.5 which indicates that such possibility may be ruled out

Estimation results

Table 2.2 presents the estimation result of where columns (1a) and (2a) present the results of equation (2.2). Our instrument is significantly and negatively associated with the access to internet ($p < 0.05$); the higher is the likelihood that a household is located in an interior region, the lower is the probability of having access to internet. Based on this result, we argue that our instrument seems to be relevant[9].

We present the estimation results of equation (2.1) in columns 1b and 2b in Table 2.2. We find that number of cooks, years of education, household expenditure, occupation of the spouse, time to road, ownership of land, ownership of BPL card, access to PDS, as well as ownership of television significantly affect the household cooking fuel choice. Furthermore, we find that access to internet deters the likelihood of choosing dirty cooking fuels ($p < 0.01$). In order to assess the joint impact of access to internet and years of schooling, we include an interaction term between the two in our model. We find that the interaction term is negative and significant ($p < 0.1$).

Our results further indicate that the correlation between the error terms of equation (2.1) and equation (2.2) is not significant at the ten percent level in both the models. In other words, there is little evidence that the error terms of the two equations are jointly distributed and equation (2.1) can be estimated separately using a naïve probit model[10].

[9] To examine the strength of our instrument, we conduct the test for weak instrument proposed by Olea and Pflueger (2013) which is designed for linear models. Under the assumption that our model is linear, we conduct an IV regression and obtain a significant negative association between $interior_i$ and $internet_i$. However, the effective first-stage F-statistic is 3.33, which is lower than the critical value of 12 (significance level of 10%). Although this test may not be valid in binary model, this result however seems to indicate that our instrument may be weak, thereby raising the issue of weak instrument bias. Therefore, in order to address the possible endogeneity issue, the propensity score matching may be a better approach as we shall discuss later. Further, we would like to add that, the null hypothesis that the model is weakly identified can be rejected at 10% (Anderson canon. corr. LM statistic: 3.418; p-value: 0.064).

[10] Under the assumption of linearity of the model, we test the null hypothesis that 'access to internet' is exogenous. We obtain the Wu-Hausman F test statistics to be 2.416 (p-value 0.12) and the Durbin-Wu-Hausman χ^2 test statistics is 2.484 (p-value 0.11) suggesting that the null hypothesis may not be rejected. This further suggests that equation (2.1) can be estimated using a naïve probit model.

<Table 2.2 approximately here>

Table 2.3 presents the estimation results of equation (2.1) using a naïve probit model. In conformation with previous literature, we find that individuals with higher years of schooling, belonging to households with higher income, owning television and having access to the public distribution system have a lower likelihood of choosing dirty cooking fuels. On the other hand, households located further from motorable roads and owning land have a higher tendency to choose dirty cooking fuels. Analogously, households having higher household size (proxied by number of cooks), spouse working in agricultural sector and households owning BPL cards have a higher likelihood to choose dirty cooking fuels. Furthermore, having access to internet has a negative and significant association with the likelihood to choose dirty cooking fuels.

It will be interesting to compare the size of the association of access to internet with cooking fuel choice to that of households' expenditure and individuals' years of schooling. However, as access to internet is discrete variable, while the latter two are continuous ones, it will be difficult to comment directly based on the estimated coefficients. One plausible approach can be to compare the average marginal effects (AME) of the three selected covariates. The AMEs are presented in column (1b) and (2b) of Table 2.3.

We find that households with access to internet have approximately 12 percentage point less likelihood to choose dirty cooking fuels than households without access to internet (p < 0.01). On the other hand, the AME of years of schooling is obtained to be -0.016 (p <0.01); a ten percentage point increase in individuals' years schooling is likely to increase individuals' likelihood to choose dirty cooking fuels by approximately 16 percentage points. The magnitude of the impact of household expenditure is similar to that of years of schooling; the AME of household expenditure is estimated as -0.014 (p <0.01). In other words, if the household expenditure increases by ten percentage points, her likelihood to choose dirty cooking fuels reduces by 14 percentage points. Therefore, we can conclude that the magnitude of the impact of 'access to internet' on likelihood to choose dirty cooking fuels lie in a comparable range if there is a ten-percentage points increase in 'years of schooling' and 'household expenditure'.

<Table 2.3 approximately here>

To look into the joint impact of 'access to internet' and 'years of schooling' on cooking fuel choice decision, we include an interaction term of the two variables in the model. The interaction term has a significant and negative effect on the likelihood to choose dirty cooking fuels ($p < 0.1$) (see column (2b) in Table 2.3). To be specific, for individuals with access to internet, increase in education by one additional year is likely to reduce the likelihood to choose dirty fuels by 8.1 percentage points, in comparison to those without access to internet. This is also evident from the predictive margins and contrasts in predictive margin as presented in Figure 2.2.

<Figure 2.2 approximately here>

Panel I of Figure 2.2 presents the predictive margin of access to internet on choice of dirty cooking fuels across levels of years of schooling. Although the confidence intervals are overlapping, it is evident that the effect in years of education is higher for households with access to internet than their counterparts without access to internet. This is also reflected in the contrasts in predictive margin presented in Panel II of Figure 2.2. The contrast is downward sloping: the difference in the predicted likelihood to choose dirty cooking fuels between households with access to internet and that without internet, increases with the increase in years of education. This result further demonstrates that literacy coupled with access to internet is effective in reducing the likelihood to reduce dirty cooking fuel choice.

2.3.2. Average treatment effect and hidden bias

Our analysis provide evidence in favour of our hypothesis that access to internet is negatively and significantly associated with the likelihood to choose dirty cooking fuels. However, there may be several confounding factors such as, individuals' preference for technologies that may influence both individuals' access to internet as well as the cooking fuel choice, thereby raising the endogeneity issue. Although we have tried to address this issue using the instrumental variable approach (results presented in Table 2.2), there may be some caveats in the analysis as our instrument is a weak one as discussed in previous sub-section. This may give rise to the weak instrument bias thereby raising question on the inference given that endogeneity issue may not be properly addressed. In this regard, one plausible and reliable solution to address the problem of endogeneity is to use the PSM approach (Rosenbaum and Rubin, 1983). Guo (2015) and Guo and Fraser (2014) provide a detailed review on how PSM has been applied in variety of disciplines including economics to address the problem of endogeneity in recent times.Therefore, we take the PSM approach to estimate the causal relationship between 'access to internet' and 'cooking fuel choice', where we attempt to isolate the specific effect of the treatment (access to internet) on outcome (cooking fuel choice) and compare the likelihood to choose dirty cooking fuels of the individuals with and without access to internet who are otherwise similar to each other in terms of other covariates.

The PSM approach in observational study resembles an experimental design where we balance for the covariates of the treated individuals (individuals with access to internet) and individuals in the control group (those without the internet access) (Kirk and Sampson, 2013). In our sample, 96 individuals belong to the treatment group while 469 individuals belong to the control group.

Table 2.4 presents the result of the PSM analysis. We have obtained a significant negative average treatment effect on the treated of the magnitude -0.239 (standard error: 0.068). In other words, the probability to choose dirty cooking fuels for an individual with access to internet is 23.9 percentage points lower than that of the matched control group member.

<Table 2.4 approximately here>

As mentioned before, there still remains possibility of the presence of unobserved confounders which when included in the analysis may alter the estimation results. To investigate the extent of the validity of our causal inference in presence of the hidden bias due to the unobserved confounders, we conduct a sensitivity analysis based on the bounding approach proposed by Rosenbaum (2002). In contrast to the model estimated by Rosenbaum (2002) that involves a continuous response variable, the dependent variable in our estimation involves a binary variable. In such a setup, bounding approach based on Mantel-Haenszel test statistics (Mantel and Haenszel, 1959) that borrows the idea of Rosenbaum (2002), is relevant[11].

Table 2.5 presents the results of the sensitivity analysis. Γ in Table 5 refers to the percentage increase in the odds of assignment to the treatment group compared to the control group due to unobservable confounders. The PSM analysis suggests that individuals with access to internet is less likely to choose dirty cooking fuels indicating a negative selection bias. For this purpose, we focus on the Q_{MH}^{-} statistic and its corresponding level of significance presented in Table 2.5.

In Table 2.5, $\Gamma = 1$ corresponds to the situation where there is no hidden bias and therefore, we may conclude that access to internet has a significant negative effect on the likelihood to choose dirty cooking fuels (Q_{MH}^{-} = 3.246; $p_{MH}^{-} = 0.001$). At $\Gamma = 1.1$, we examine the effect of hidden bias from the unobservable confounders that will cause a 10 percent increase in the odds

[11] The two bounds of the Mantel Haenszel test statistics, Q_{MH}^{+} and Q_{MH}^{-} corresponding to the positive selection bias and negative selection bias along with their levels of significance p_{MH}^{+} and p_{MH}^{-}, respectively may be computed using the user-written command *mhbound* in statistical software STATA (Becker and Caliendo, 2007).

of assignment to the treatment group compared to the control group. In this scenario, we continue to observe a significant negative effect of the access to internet on the likelihood to choose dirty cooking fuels ($Q_{MH}^{-} = 2.933$; $p_{MH}^{-} = 0.002$). This significant negative causal inference continues to hold true until $\Gamma = 1.6$. Beyond $\Gamma = 1.6$, the hidden bias arising from unobserved confounders is severe enough to cause the treatment effect of access to internet to be no longer significant at 5% level of significance. Therefore, we may conclude that our causal inference that internet access is likely to deter the dirty cooking fuel choice remains valid even if the hidden bias from the unobserved confounders causes an additional 60 percent increase in the odds of assignment to the treatment group compared to the control group.

<Table 2.5 approximately here>

2.4. Conclusion

Health risks related to HAP is salient in developing countries. Despite the alarming effects of the HAP on individuals' health, a significantly high proportion of individuals continue to use dirty cooking fuels particularly in rural areas of developing economies. We use a unique survey data from 565 rural India households to explore the role of information disseminated through access to internet on individuals' likelihood to choose dirty cooking fuels

Our results provide evidence in support of our hypothesis access to information disseminated through internet may significantly reduce the individuals' likelihood to choose dirty cooking fuels. Results from the PSM suggests that households with access to internet tend to have approximately 24 percentage points lower likelihood to choose dirty cooking fuels than similar individuals without access to internet. Sensitivity analysis further indicates that this causal inference is likely to hold true even if the unobserved confounding covariates increases the odds of assignment to access to internet by an additional 60 percent relative to individuals without access to internet.

Above evidence leads to at least two important directions about policy design. First, since access to information disseminated through internet can be an important determinant to deter dirty fuel choice, enabling better access to internet to the households may lead to better awareness about health hazards related to HAP specifically in rural areas. Although the Government of India is currently undertaking two major programs, one for the provision of free LPG connection as well as low-cost LPG usage and other, for access to internet, these two programs should complement each other particularly in rural areas. In this way, the problem of HAP may be gradually diminished in developing countries in general and India in particular. Second, the results suggest that effect of information provision may vary across different levels of education. This may indicate that individuals with lower levels of education may not utilize the information provision effectively even if, they have access to internet. Therefore, provision of education is likely to be another long-term policy prescription that will help to reduce the issue of HAP. Alternatively, the households having access to information, but with lower levels of education may be considered to be the target households for information provision.

Although we have tried to address the issue in a meaningful manner, focussing only on the influence of individual and household-specific covariates on cooking fuel choice and usage may be an over-simplification. There is a possibility that individuals' perceptions about health risks related to different cooking fuel categories may affect individuals' choice of cooking fuels. Consequently, one possible future research avenue is the extension of our analysis to assess the impacts of individuals perception of health risks on cooking fuel choice and usage. Furthermore, individuals' choice of cooking fuel may be motivated by individuals' valuation of the health risk related to HAP and not exclusively by individual and household-specific covariates. . For a holistic analysis of the individuals' choice of cooking fuel as well as cooking fuel usage pattern, we need to explore such valuation of health risks related to HAP. These will be taken up later and we plan to extend our research in this direction

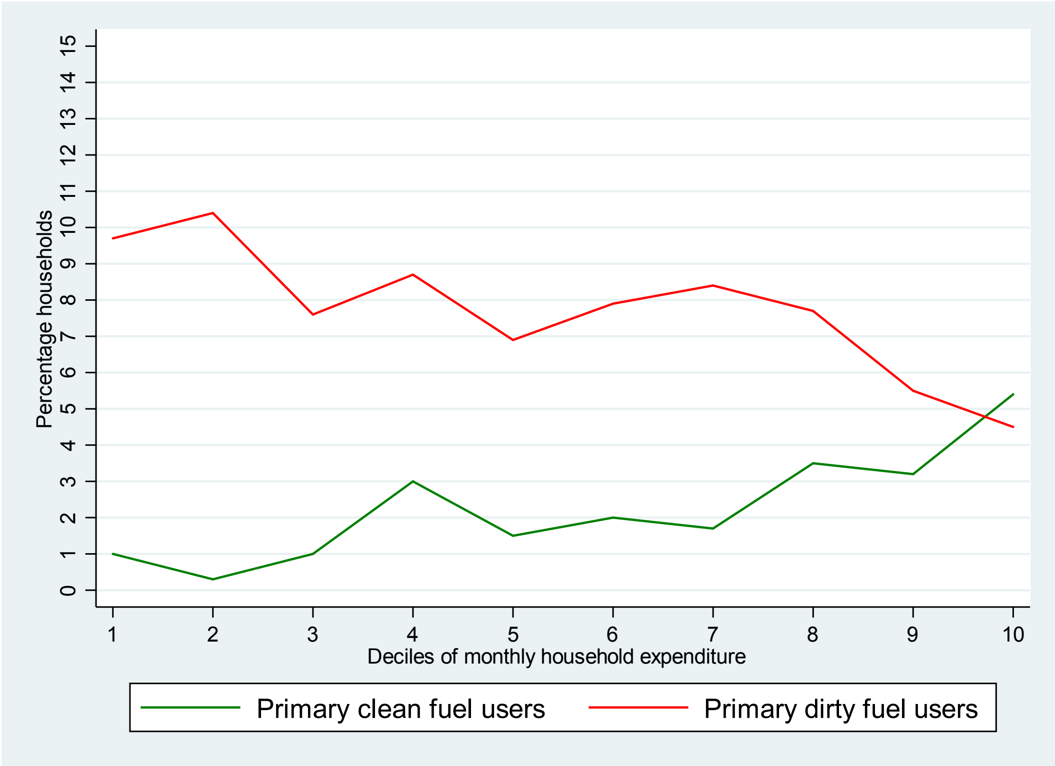

Figure 2.1. Distribution of proportion of primary cooking fuel users across income deciles

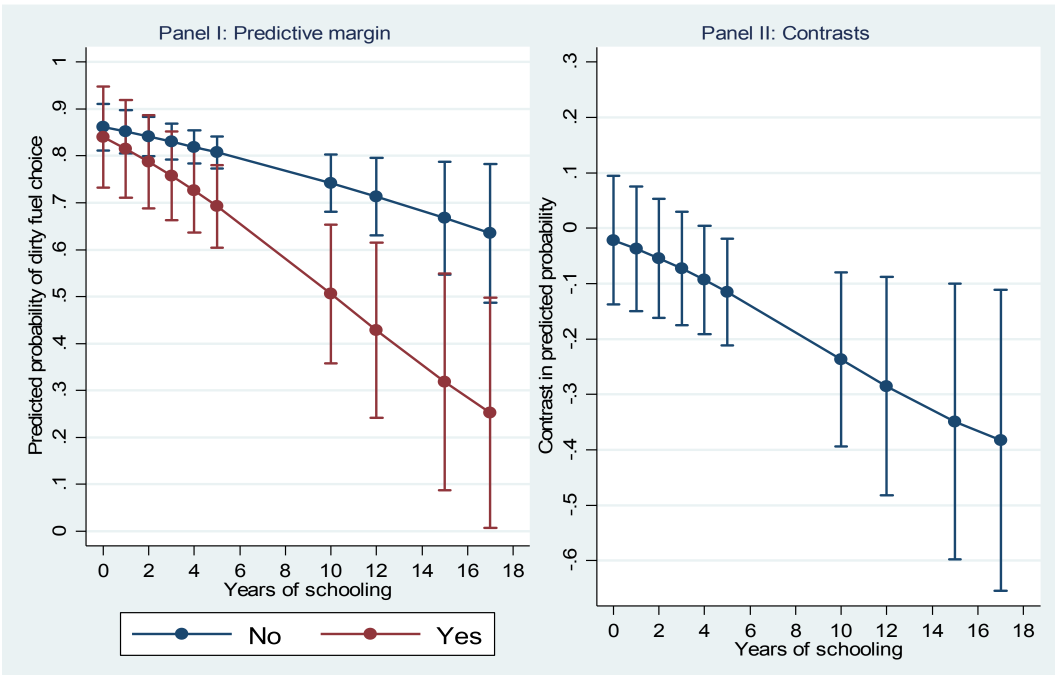

Figure 2.2: Effect of internet access on dirty cooking fuel choice across different years of schooling

Table 2.1. Descriptive statistics

	Mean	SD	Min	Max
Cooking fuel related variables				
Primary dirty cooking fuel user (binary)	0.77	0.42	0.00	1.00
Access to information-related variables				
Access to internet (binary)	0.17	0.38	0.00	1.00
Other covariates				
Number of cooks	1.13	0.41	1.00	4.00
Age	36.81	10.81	16.00	75.00
Years of schooling	4.87	4.16	0.00	17.00
Housewife (binary)	0.97	0.18	0.00	1.00
Household decision-maker (binary)	0.06	0.23	0.00	1.00
Spouse works in informal sector (binary)	0.43	0.50	0.00	1.00
Spouse works in agricultural sector (binary)	0.30	0.46	0.00	1.00
Hindu (binary)	0.68	0.47	0.00	1.00
Time to motorable road (in minutes)	12.97	12.82	1.00	70.00
Electrification (binary)	0.98	0.13	0.00	1.00
Expenditure (in 1000 INR)	7.10	4.78	1.50	55.00
Participates in micro-finance (binary)	0.27	0.44	0.00	1.00
Ownership of land (binary)	0.50	0.50	0.00	1.00
Access to PDS (binary)	0.96	0.19	0.00	1.00
Ownership of BPL cards (binary)	0.27	0.44	0.00	1.00
Ownership of television (binary)	0.75	0.44	0.00	1.00
Households located in an interior region (binary)	0.34	0.47	0.00	1.00

Note: The sample size is 565. PDS refers to the Public Distribution System. BPL refers to Below Poverty Line households. The poverty line for rural West Bengal based on capita consumption expenditure per month is fixed at INR 783 (Reserve Bank of India, 2012) where INR 68= USD 1 (average of the average exchange rate in December, 2016 and January, 2017).

Table 2.2. Factors affecting cooking fuel choice (bivariate probit)

	(1a) internet	(1b) dirty fuel	(2a) internet	(2b) dirty fuel
Household located in interior	-0.401**		-0.410**	
	(0.203)		(0.204)	
Years of schooling	0.038**	-0.062***	0.036*	-0.048**
	(0.019)	(0.019)	(0.019)	(0.021)
Access to internet		-1.513***		-0.935
		(0.471)		(0.651)
Internet* Years of schooling				-0.072*
				(0.041)
Number of cooks	-0.144	0.447**	-0.153	0.453**
	(0.174)	(0.216)	(0.175)	(0.22)
Age	0.025***	-0.001	0.025***	-0.002
	(0.007)	(0.008)	(0.007)	(0.008)
Housewife	-0.272	-0.451	-0.271	-0.479
	(0.370)	(0.402)	(0.372)	(0.421)
Household decision-maker	0.614**	-0.147	0.610**	-0.219
	(0.266)	(0.330)	(0.266)	(0.34)
Spouse works in informal sector	0.123	0.249	0.119	0.234
	(0.172)	(0.158)	(0.173)	(0.160)
Spouse works in agricultural sector	-0.192	0.402**	-0.196	0.412**
	(0.205)	(0.203)	(0.206)	(0.206)
Hindu	-0.068	-0.269	-0.075	-0.268
	(0.171)	(0.170)	(0.171)	(0.171)
Time to road	0.003	0.021***	0.003	0.022***
	(0.009)	(0.008)	(0.009)	(0.008)
Expenditure	0.073***	-0.037	0.074***	-0.046
	(0.019)	(0.0263)	(0.019)	(0.0280)
Participates in microfinance	0.304**	-0.027	0.305**	-0.041
	(0.15)	(0.155)	(0.150)	(0.158)
Owns land	0.409***	0.379**	0.408***	0.378**
	(0.155)	(0.147)	(0.156)	(0.150)
Access to PDS	0.420	-0.850*	0.436	-0.884*
	(0.399)	(0.505)	(0.401)	(0.516)
Owns BPL card	-0.291	0.315*	-0.290	0.346*
	(0.179)	(0.181)	(0.179)	(0.184)
Owns television	0.243	-0.550**	0.258	-0.557***
	(0.200)	(0.214)	(0.200)	(0.215)
ρ	0.633		0.528	
	(0.394)		(0.404)	
Log likelihood	-421.5		-419.9	
χ^2	266.3		246.6	

Note: This table provides the joint estimation results for equation (2.1) and (2.2) under the assumption of endogeneity. The dependent variable is access to internet (=1 if the household has access to internet) in columns 1a and 2a and, choice of cooking fuel (=1 if dirty cooking fuel is chosen) in columns 1b and 2b. The sample size is 565. ***, ** and * indicate statistical significance at the one, five and ten per cent levels respectively. Robust standard errors are presented in parenthesis. The constant terms are not reported for the sake of space

Table 2.3. Factors affecting cooking fuel choice (naïve probit model)

	(1) coef.	(2) AME	(3) coef.	(4) AME
Years of schooling	-0.074***	-0.016***	-0.056***	-0.015***
	(0.019)	(0.004)	(0.021)	(0.004)
Access to internet	-0.570***	-0.121***	-0.06	-0.122***
	(0.174)	(0.036)	(0.317)	(0.043)
Internet* Years of schooling			-0.081*	
			(0.046)	
Number of cooks	0.547***	0.116***	0.529**	0.111**
	(0.199)	(0.042)	(0.206)	(0.043)
Age	-0.007	-0.001	-0.007	-0.001
	(0.007)	(0.001)	(0.007)	(0.001)
Housewife	-0.437	-0.082	-0.467	-0.086
	(0.365)	(0.059)	(0.374)	(0.06)
Household decision-maker	-0.333	-0.076	-0.387	-0.089
	(0.258)	(0.063)	(0.267)	(0.066)
Spouse works in informal sector	0.257	0.055	0.235	0.049
	(0.16)	(0.034)	(0.161)	(0.033)
Spouse works in agricultural sector	0.472**	0.100**	0.469**	0.099**
	(0.201)	(0.0423)	(0.199)	(0.041)
Hindu	-0.276	-0.059	-0.27	-0.057
	(0.179)	(0.038)	(0.177)	(0.037)
Time to road	0.022**	0.005**	0.023**	0.005**
	(0.01)	(0.002)	(0.01)	(0.002)
Expenditure	-0.065***	-0.014***	-0.069***	-0.014***
	(0.021)	(0.004)	(0.022)	(0.005)
Participates in microfinance	-0.111	-0.024	-0.114	-0.024
	(0.155)	(0.033)	(0.156)	(0.033)
Owns land	0.323**	0.069**	0.328**	0.069**
	(0.159)	(0.033)	(0.159)	(0.033)
Access to PDS	-1.053**	-0.223**	-1.048***	-0.22***
	(0.436)	(0.092)	(0.397)	(0.083)
Owns BPL card	0.396**	0.081**	0.414**	0.083**
	(0.177)	(0.034)	(0.175)	(0.033)
Owns television	-0.615***	-0.12***	-0.606***	-0.116***
	(0.197)	(0.035)	(0.198)	(0.034)
Log likelihood	-215.4		-213.4	
Pseudo R^2	0.293		0.300	
χ^2	144.1		134.9	

Note: This table provides the estimation results for equation (2.1) under the assumption that the model does not suffer from endogeneity. The dependent variable is choice of cooking fuel (=1 if dirty cooking fuel is chosen) AME refers to the Average Marginal Effect. The sample size is 565. ***, ** and * indicate statistical significance at the one, five and ten per cent levels respectively. Robust standard errors are presented in parenthesis and delta standard errors are presented in brackets. The constant terms are not reported for the sake of space.

Table 2.4. Estimated propensity score matching

Sample	Treated	Controls	Difference	t-stats
Unmatched	0.489	0.827	-0.338	7.5***
			(0.045)	
Average Treatment Effect on the treated	0.489	0.729	-0.239	3.49***
			(0.068)	

Note: This table provides the estimated PSM in our sample. The treatment group is the households that have access to internet and the control group is the households without access to internet. The sample size in the treatment group is 96 while that in the control group is 469. The outcome variable is whether the household chooses dirty cooking fuel. Values in parenthesis represents standard errors. ***, ** and * indicate statistical significance at the one, five and ten per cent levels respectively.

Table 2.5. Mantel-Haenszel bounds for hidden bias

Γ	Q_{MH}^{+}	p_{MH}^{+}	Q_{MH}^{-}	p_{MH}^{-}
1	3.246	0.001	3.246	0.001
1.1	3.578	0.000	2.933	0.002
1.2	3.875	0.000	2.642	0.004
1.3	4.15	0.000	2.375	0.009
1.4	4.406	0.000	2.129	0.017
1.5	4.647	0.000	1.9	0.029
1.6	4.873	0.000	1.687	0.046
1.7	5.087	0.000	1.487	0.068
1.8	5.29	0.000	1.3	0.097
1.9	5.483	0.000	1.122	0.131
2	5.667	0.000	0.954	0.17

Note: This table presents the Mantel-Haenszel bounds for the variable choice of dirty cooking fuels. Γ represents the odds of differential assignment due to unobserved factors. Q_{MH}^{+} represents the Mantel-Haenszel statistic under the assumption of overestimation of treatment effect. Q_{MH}^{-} stands for the Mantel-Haenszel statistic under the assumption of the underestimation of treatment effect. The corresponding level of significance is presented in the parenthesis. p_{MH}^{+} and p_{MH}^{-} represents the corresponding levels of significance for Q_{MH}^{+} and Q_{MH}^{-} respectively.

Chapter 3. Subjective Probabilistic Expectations, Household Air Pollution, and Health: Evidence from Cooking Fuel Use Patterns in West Bengal, India

3.1. Introduction

Household air pollution (hereafter, HAP), caused mainly by the incomplete combustion of dirty cooking fuels coupled with inefficient cooking practices, is a global health threat accounting for more than 3.8 million annual deaths worldwide (WHO, 2018). In particular, developing countries suffer from exposure to HAP, accounting for approximately 3.7% of the loss of disability-adjusted life years (WHO, 2007).

In developing countries and particularly in rural areas, health risks related to HAP are preventable through reduction in exposure by the use of clean cooking fuels. Several studies are being conducted to understand why large proportions of the populations of developing countries continue to use dirty cooking fuels despite the alarming health risks of HAP. Previous studies revealed that several other factors apart from the usual economic and demographic factors, such as income (e.g., Heltberg, 2005) and education (e.g., Farsi et al., 2007), may significantly affect the cooking fuel choices of households in developing countries. Notable among these are energy access (Pachauri and Jiang, 2008) and supply side factors such as availability and relative cost advantages (Jeuland et al., 2015). 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. Recently, Zahno et al. (2020) identified that information provision may be an effective way to encourage

---

Publication Information: This chapter is published as Chattopadhyay, M., Arimura, T. H., Katayama, H., Sakudo, M., & Yokoo, H. F. (2021). Subjective Probabilistic Expectations, Household Air Pollution, and Health: Evidence from Cooking Fuel Use Patterns in West Bengal, India. *Resource and Energy Economics*, 66, 101262. This chapter is co-authored with Prof. T.H. Arimura, Prof. H. Katayama, Dr. M. Sakudo and Dr. H.F. Yokoo.

individuals in developing countries to increase their clean cooking fuel usage.

Alternatively, several studies have explored the avenue towards HAP mitigation through the adoption of clean cooking technology. Regardless of the expected health benefits of such technologies, liquidity constraints (Bensch et al., 2015) and a failure to perceive the seriousness of such health risks (Mobarak et al., 2012) may pose significant barriers, particularly to the adoption of such technology. Apart from factors related to fuel and technology, several individual-specific factors may significantly influence adoption decisions (e.g., Lewis and Pattanayak, 2012; Jeuland et al., 2014a). Researchers observe that the upgradation of supply chain and demand promotion (e.g., Pattanayak et al., 2019; Levine et al., 2018) and the subsidization of the free distribution of clean cooking technology (Bensch and Peters, 2020) may help overcome the liquidity constraint.

This study extends former lines of research related to the choice of household cooking fuels in developing economies. Despite much evidence in the literature, there remains some unexplored aspect of the choice of household cooking fuel, specifically the role of expectations about health risks from HAP. In the absence of precise estimates of and/or information about the health risks related to HAP, individuals may have different expectations about these health risks and therefore make different decisions regarding the use of dirty fuel, regardless of personal preferences. Furthermore, researchers typically have no information about individuals' expectations for HAP'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, 2016).

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, 2016) and mental health and labor supply (Baranov et al., 2015) among others. These studies confirm that expectations play a certain role in various choice situations.

The present study attempts to bridge the existing gap in the literature on the choice of household cooking fuel vis-à-vis subjective probabilistic expectations (hereafter, SPEs). In particular, we investigate the role of individuals' expectations of becoming sick with diseases typically observed from HAP exposure (specifically, dry cough, sore or runny eyes, and difficulty breathing) on their cooking fuel usage pattern and health status. By the cooking fuel usage pattern, we refer to the fraction of days in which dirty fuel is used over a 30-day period. We explore the association between individuals' SPEs of becoming sick with such diseases and their cooking fuel usage patterns. Concurrently, we investigate how the individual's cooking fuel usage pattern, in turn, has some role in 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 HAP are associated with their health status indirectly by influencing their cooking fuel usage patterns.

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 HAP, we have adopted an interactive elicitation method using visual aids, as developed by Delavande and Kohler (2009). A potential problem in estimating the health status of the individual is the endogeneity of the 'fraction of days of dirty fuel usage for cooking'. To address this issue, we have adopted instrumental variable approach. As an instrument, we have used the 'opportunity to access cooking fuel for free' in our analysis. Further, to address nonlinearity associated with binary health status, we have adopted the Two Stage Residual Inclusion (2SRI) model developed by Terza et al. (2008).

Our results show that SPEs of becoming sick from dirty fuel usage are negatively and significantly associated with the dirty fuel usage in the household. Moreover, dirty fuel usage and self-reported health status of becoming sick are statistically significantly correlated.

Given this role played by SPEs, we conduct a policy simulation where we assume information on health risks of dirty fuel usage is hypothetically provided. Following Delavande and Kohler (2016), we assume that hypothetical information provision is successful at educating people, who fully revise their SPEs to the provided level. Using estimated coefficients, we calculate predicted probabilities of fuel usage and health status for a baseline and the policy scenario. Our simulation demonstrates that although the provision of information may lead to statistically significant changes in fuel usage pattern and health status, this change may be small in size.

The remainder of the chapter 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 results of the hypothetical information provision policy simulation. Possible biases in the findings arising from the endogeneity of the SPEs and omitted variables with regard to unobserved factors such as innate ability and female empowerment are acknowledged and discussed in Section 5. Section 6 concludes by discussing the directions of future research.

3.2. Introduction to dataset

3.2.1. Survey design

This study uses the data collected from both the rounds (December 2016- January 2017 and December 2017- January 2018). Out of the 600 households randomly selected and interviewed during the first round (December 2016 and January 2017), 596 respondents completed the survey. From December 2017 to January 2018, the enumerators 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 3.1 presents descriptive statistics of the variables used in this study; these will be explained in the next subsection.

<Table 3.1 approximately here>

3.2.2. Descriptions of the variables and their summary statistics

Fuel usage pattern

Empirical evidence in developing countries, particularly in rural areas, indicates the simultaneous use of multiple cooking fuels, often referred to as fuel stacking (Gould and Urpelainen, 2018). To account for this in this study, the fuel usage pattern (represented by the proportion of days dirty fuel is used for cooking) is set as the variable of interest instead of the primary cooking fuel type as has been done in previous literature (e.g., Dendup and Arimura, 2019). 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[1] in the 30 days prior

[1] Although WHO (2018) classifies kerosene to be dirty cooking fuel, we have implemented our survey before this classification was published. In our study, we have consistently followed the nomenclature referred by Heltberg (2005), among others, where kerosene is classified as clean cooking fuel. We have elicited the SPE of becoming sick from clean cooking fuels by mentioning about LPG/kerosene in 2016-17, i.e., prior to the publication of the WHO categorization. This has constrained us to revert to referring kerosene as dirty cooking fuel., However, we have tried to emphasize that such constraint would lead to a negligible bias in this context, as evident from the following discussion.

In our effective sample, only 4% of the households use kerosene at least once in a 30-day period, thereby, having a non-zero value in the share of days of kerosene usage. Even if we include kerosene to be a dirty cooking fuel, the variable 'fraction of days of dirty fuel usage in a 30-day period' is not changed much (mean changes from 0.678 (sd:0.378) to 0.686 (sd:0.379)). Furthermore, we have also found that there is high association (Pearson's correlation coefficient is 0.9934) between the two versions of the variable 'fraction of days of dirty fuel usage'. We perform the large sample test of the hypothesis that the population correlation coefficient is 0.99 against the alternative that it is more than 0.99 using the z-transformation (Fisher, 1921). The value of the test statistic is obtained as 4.946 ($p < 0.001$) indicating an extremely significant and almost unity correlation coefficient between the two versions of the two variables. In other words, these two series may be considered to to the previous month. 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 3.1), suggesting the prevalence of dirty fuel usage in rural areas of India. To better understand this variable, we also draw its distribution in Figure 3.1. Although a certain portion of households are observed at the endpoint 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.

<Figure 3.1 approximately here>

Self-reported health status

Evidence of the association of HAP with various health risks, including cardiovascular, respiratory, and vision-related diseases, is available in the literature (Smith and Pillarisetti, 2017). Therefore, ideally, we should have incorporated the health information of the individuals for the abovementioned symptoms using a professional medical team. However, the process of conducting clinical tests for each respondent in the sample would have been costly. As an alternative, we have followed Hanna et al. (2016) and incorporated information on physical symptoms using individuals' past recollection of the experience of suffering from such symptoms in the last 30 days. From the questionnaire of Hanna et al. (2016), we included ten cardio-vascular, respiratory, and vision-related symptoms in our questionnaire during the pilot study in August 2016 among 70 households in the same locality. Of them, three symptoms, namely, dry cough, sore and runny eyes, and difficulty breathing, were found to be the most prominent features. Consequently, we selected them in the final questionnaire to conduct the detailed survey.

---

be almost same and this further indicates that even if there is any bias arising from not including kerosene in dirty cooking fuels, this bias is expected to be statistically insignificant.

The self-reported health status of the respondents refers to whether the respondent has experienced at least one of three minor, yet common physical symptoms caused by HAP—dry cough, sore or runny eyes, and difficulty breathing—in the last 30 days. 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 3.1). Along 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 (2016). 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 (out of ten candies) to her perceived likelihood of the occurrence of an event. 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[2] in the next three months given that she is currently sick (not sick) and uses dirty fuels only. 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 3.2 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[3], to state how likely she thinks that the hypothetical individual is likely to remain (become) sick in next 3 months from using a)LPG/kerosene and b)coal/solid biomass fuels/firewood given she is currently sick (not sick). 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.

[2] 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.
[3] The instructions and survey instruments used to elicit the SPEs are presented in the appendix (3.A2).

<Table 3.2 approximately here>

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 3.2, 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 3.1, the mean of the difference is found to be -0.47: the respondents on average think that clean fuel usage is 47 percentage points 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 3.1; the respondents on average think that dirty fuel usage is 40 percentage points more likely to degrade their health than clean fuel usage.

<Figure 3.2 approximately here>

Using the elicited SPEs conditional on dirty fuel usage, we calculate the equilibrium distribution of the Markov process, denoted as $SP(s = 1|\ d)$[4]. 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 3.3 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 3.1; the respondents, on average, 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.

<Figure 3.3 approximately here>

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

[4] $SP(s = 1|d) = \frac{SP(s_{t+1} = 1|s_t = 0, d)}{1+[SP(s_{t+1} = 1|s_t = 0, d)-SP(s_{t+1} = 1|s_t = 1, d)]}$. 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.

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[5]), 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 from forests, common lands, farms, and domestic animals (IEA, 2017). Such access may influence the respondents' cooking fuel usage pattern. We therefore asked the respondents whether they have any opportunities to collect/obtain cooking fuels for free. Fifty-nine percent of the respondents were found to have access to free cooking fuels (see Table 3.1).

3.3. Estimation models and results

We address whether SPEs play any role in individuals' cooking fuel usage patterns, and in turn whether usage patterns are associated with individuals' health. If so, we may conclude that SPEs have a direct association with cooking fuel usage patterns and an indirect association with individuals' health. For this purpose, we take a two-step approach: First, we estimate the role of SPEs on individuals' cooking fuel usage patterns. Second, we estimate the relationship of cooking fuel usage patterns on individuals' health. This allows us to infer the extent to which SPEs are associated with individuals' health conditions indirectly through cooking fuel usage.

[5] By ventilation, we refer to any facility that reduces the incidence of smoke in the cooking area. Although approximately 97% of the respondents reported having ventilation, this mainly refers to the presence of windows or small holes in the walls of the cooking area.

For this estimation, we need to address several econometric issues. First, the dependent variable in the model of the cooking fuel usage pattern has a particular feature; specifically, it is restricted to values in the unit interval. In addition, a non-negligible fraction of observations is 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 health status equation that includes the 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 aforementioned 2SRI method.

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 modelling 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 subsections, we describe the estimation model and results in detail.

3.3.1. Association between SPEs and cooking fuel usage patterns

Estimation model

We assume that individual i chooses fuel usage pattern (w_i) based upon her SPEs of becoming 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:

$$E(w_i|\boldsymbol{spe_i}, \boldsymbol{z_i}, free_i) = \Phi(\beta_0 + \boldsymbol{\beta_1 spe_i} + \boldsymbol{\beta_2 z_i} + \beta_3 free_i), \tag{3.1}$$

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 (3.1) in the following form:

$$w_i = \Phi(\beta_0 + \boldsymbol{\beta_1 spe_i} + \boldsymbol{\beta_2 z_i} + \beta_3 free_i) + r_i, \tag{3.2}$$

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

We estimate equation (3.1) (or, equivalently, equation (3.2)) 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 (3.1) 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").

<Table 3.3 approximately here>

Column 1 of Table 3.3 presents the estimation results. In line with the previous literature, the results may 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[6]. It may also be found

[6] In India, religious practices may influence cooking fuel usage due to cultural beliefs, cooking practices, and preferences for certain varieties of food, among other factors. Studies such as Gupta and Kohlin (2006) suggest that Islamic households show a higher tendency of dirty fuel usage compared to their Hindu counterparts due to their unique cooking practices and food preferences.

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" may 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 (3.1). As presented in Column 1a of Table 3.4, 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 may lower the fraction of days of dirty fuel usage by approximately 1.9 percentage point.

<Table 3.4 approximately here>

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 may 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 are associated with individuals' 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 (3.1) by fitting a linear regression[7]. We observe that, 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.

3.3.2. Associations between cooking fuel usage patterns and 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 ($\boldsymbol{x_i} = [\boldsymbol{z_i}\ \boldsymbol{spe_i}]$):

$$s_i^* = \gamma_0 + \gamma_1 w_i + \boldsymbol{\gamma_2 x_i} + \gamma_3 r_i + u_i, \tag{3.3}$$

where u_i is an idiosyncratic error term that is uncorrelated with w_i and $\boldsymbol{x_i}$. In equation (3.3), the presence of r_i defined in equation (3.2) 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

[7] The estimation result of the robustness check is presented in appendix (3.A3)

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:

$$\Pr(s_i = 1 | w_i, \boldsymbol{\gamma_2 x_i}, r_i) = \Phi(\gamma_0 + \gamma_1 w_i + \boldsymbol{\gamma_2 x_i} + \gamma_3 r_i). \tag{3.4}$$

Equation (3.4) suggests a two-step estimation procedure, which is an application of the 2SRI. The first step is to estimate equation (3.2) (which we have already done) and compute a residual $\hat{r}_i$ for each i. The next step involves replacing r_i with $\hat{r}_i$ in equation (3.4) 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 (3.2) should not be included in equation (3.4) 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.

For our instrument to be valid, it must be sufficiently correlated with the endogenous regressor (i.e., instrument relevance), while it must not be correlated with the error term in equation (3.3) or must not directly influence the dependent variable (i.e., instrument exogeneity). The instrument seems to be relevant in that it is strongly correlated with the fuel use pattern (captured by the fraction of days of dirty fuel use), as presented in Tables 3.3 and 3.4[8]. On the other hand, the exogeneity of the instrument cannot be tested formally because the model is just identified. We therefore discuss the plausibility of the exogeneity assumption.

[8] To examine the strength of our instrument, we conduct the test for weak instrument proposed by Olea and Pflueger (2013) under the assumption of linearity of the model. The effective first-stage F-statistic is 59.598, which is higher than the critical value of 12 (significance level of 1%). Although this test may not be valid in binary model, this result however seems to indicate that our is a strong one. Further, we would like to add that,

First, it is not unreasonable to assume that the opportunity to access free cooking fuels does not directly influence the health status of the individual. If the opportunity to access free cooking fuels is associated with the health status of the individuals, the fuel use pattern should be a mediator, representing the only way through which free cooking fuel is indirectly associated with health status.

The most likely threat to the exogeneity assumption is unobserved individual-specific factors. A usual suspect is the degree of health unconsciousness that is contained in the error term of equation (3.3) and reflects that health-conscious individuals tend to have better health than their health-unconscious counterparts. This factor may invalidate the exogeneity assumption of our instrument by inducing a correlation between the error term and our instrument. The correlation may occur because health-unconscious individuals prefer to use firewood more than their health-conscious counterparts; the former may be more likely than the latter to choose to live where they can collect free firewood.

The threat, however, may not be so plausible when considering that the sample has been collected in India. Indian society is essentially patriarchal in nature (Perianayagam and Srinivas, 2012), which is clearly reflected in the structure of marriage. Ethnographic studies suggest that women have a low or no say over marriage decisions (including the choice of spouse or in-laws) in India (Allendorf, 2013). The patrilocal residence system prevalent in India furthermore offers married women limited or no chance to choose the location of the household (Khalil and Mookerjee, 2019). Since the respondents in our sample are essentially married women, they are unlikely to have a choice about the location of their households, regardless of whether they are health-(un)conscious. As a result, the opportunity to access cooking fuel for free is unlikely to be correlated with the degree of health-(un)consciousness.

---

the null hypothesis that the model is weakly identified can be rejected at 1% (Anderson canon. corr. LM statistic: 16.38; *p*-value: 0.000).

Despite the argument, we admit that the exogeneity of our instrument may remain questionable. We will therefore discuss potential bias in our main results when the exogeneity assumption is not satisfied.

Estimation results

Columns 1a and 1b in Table 3.5 present the estimated coefficients and corresponding average marginal effects, respectively, where the model for cooking fuel usage pattern is specified as in Column 1a of Table 4. We find that the coefficient on the residual from the cooking fuel usage pattern equation is not significant at the ten percent level; in other words, there is little evidence that the fuel usage pattern is endogenous in equation (3.4). 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 model for cooking fuel usage patterns is specified as in either Columns 2a or Column 3a of Table 3.4. These findings are in line with most of the literature that reveals the detrimental effect of dirty fuel usage on respiratory diseases, particularly among women in developing countries (e.g., Stabridis and Gameren, 2018).[9] In agreement with the literature, we also observe that older individuals and housewives are more vulnerable to the health risks of HAP (e.g., Cincinelli and Martellini, 2017; Verma and Imelda, 2019). Overall, the results for equation (4.4), along with those for equation (4.1), support the finding that SPEs

[9] However, contradictory shreds of evidence are also available in the literature. For example, Mortimer et al. (2017) show that clean cooking technology does not result in any significant improvement in respiratory health among children in rural Malawi.

are indirectly associated with health status[10] through their associations with cooking fuel usage patterns[11].

<Table 3.5 approximately here>

3.4. Simulation of a hypothetical policy

The results provided in Section 3 show that SPEs are associated with cooking fuel usage patterns and related health status. Using our estimated parameters, it is possible to simulate a hypothetical scenario under which information provision is likely to change SPEs.

According to previous studies, the effect of information provision on clean cooking fuels and/or technologies seems ambiguous. For example, Zahno et al. (2020) show that health messaging increases LPG consumption in India. In contrast, Jeuland et al. (2014b) discuss the limited impact of information provision on the use of clean cooking technology in India.

Unlike the above studies, which evaluate real-world interventions, we present a simulation of an information provision policy as Delavande and Kohler (2016).[12] Consider a case in which individuals are informed of the proportion of sick individuals among those who predominantly use dirty fuels. In other words, we assume that individuals are provided information on the estimated conditional probability of being sick among primary dirty fuel users. In our sample, this value is 397 out of 458, that is 0.87.

[10] To explore the possibility of SPEs being directly associated with self-reported health status, we re-estimated the models in Table 3.5 while excluding the cooking fuel usage patterns from the set of covariates. The results (available upon request) indicate that SPEs are still statistically insignificant, suggesting that they are not directly associated with individuals' health status.

[11] In estimating equation (3.4), we control for SPE variables by assuming that they are exogenous. Due to individual-specific unobserved factors, however, SPE variables may be endogenous as discussed in subsection 3.5.1. This raises the possibility that SPE variables may be "bad controls" (Angrist and Pischke, 2009) in equation (3.4). To address this issue, we re-estimated the models in Table 3.5 by not controlling for SPE variables. The results (available upon request) are broadly similar to those reported in Table 3.5.

[12] Note that it is difficult, if not impossible, to communicate the 'true risk' of dirty fuel usage, since identifying the true risk is quite difficult. If we wish to conduct an honest information campaign, we may need to mention the unclear effect of switching to clean fuels given the importance of ambient air pollution and other confounders.

Following Delavande and Kohler (2016), we assume that hypothetical information provision is successful at educating people, who fully revise their SPEs of becoming sick from dirty fuel usage to the value of 0.87. As a result, 510 (78) individuals with an $SP(s = 1|d)$ of below (above) 0.87 update the elicited value to that level. Under this hypothetical policy, most of the individuals become more aware of the health risks associated with dirty fuel usage.

To conduct the counterfactual, we use the estimated models specified in the first columns of Table 3.4 and Table 3.A2.[13] Using the estimates of equation (2), we calculate the predicted fraction of days of dirty fuel usage. We change the value of $SP(s = 1|d)$ to 0.87 but keep the other variables unchanged. Similarly, from the estimates of equation (4) and the predicted fraction of days of dirty fuel usage, we calculate the predicted probabilities of being sick. A baseline is calculated using the SPE values elicited in our survey.

Table 3.6 presents the policy simulation results. The table reports the summary statistics for the predicted fraction of days of dirty fuel usage in Panel I. On average, the hypothetical policy reduces the fraction of days of dirty fuel usage to 65.3% from the baseline level of 67.8%. Panel II of Table 6 reports the predicted probabilities of being sick. The average predicted proportion of sick individuals reduces to 80.5% relative to the baseline situation of 82.3%. The average change in the predicted probability of becoming sick is significantly different from zero (p <0.01). However, the magnitude of the change is small in size (1.8 percentage points).

The hypothetical policy examined above is expected to complement the current strategy adopted by the Government of India. Despite the PMUY scheme, the majority of households continue to use dirty cooking fuels as either primary or secondary source of cooking fuel

[13] The model shown in Table 3.A2 is a restricted version of that reported in Table 3.5 where the residuals from the fuel usage pattern models are excluded from a set of regressors. Table 3.A2 is presented in the appendix (Appendix 3.A4).

(Gould and Urpelainen, 2018). Our results indicate that an information provision policy for dirty cooking fuels may act as a supplement for government policy, however, the magnitude of the impact may not be large in size.

<Table 3.6 approximately here>

3.5. Discussion about the potential bias in our estimates

3.5.1 Association between SPEs and cooking fuel usage patterns

In examining the association between SPEs and cooking fuel usage patterns in subsection 3.1, we address the issue of reverse causality by using lagged values of SPEs as regressors. However, our results may remain subject to endogeneity bias, as this also occurs as a result of an omitted relevant variable. For example, the degree of health risk tolerance, one of the unobserved individual specific factors, may induce a correlation between an SPE variable and the error term (implicitly present in equation (1)). To better understand this possibility, consider the result presented in Column 2a of Table 3.4. A negative correlation occurs if respondents with higher health risk tolerance tend to underestimate the expected reduction in the likelihood of becoming sick due to clean fuel usage (i.e., $SP(s=1|c) - SP(s=1|d)$) and also tend to use dirty fuels more often in the future. In this case, the estimated coefficient of $SP(s=1|c) - SP(s=1|d)$ will be biased downward. In other words, the true effect of $SP(s=1|c) - SP(s=1|d)$ may be larger in magnitude than reported.

Innate ability may also be an omitted relevant variable. Respondents with higher levels of innate ability may be better at assessing probabilities related to the risk of becoming sick and at the same time more likely to take the necessary steps to adopt clean fuels in the future. In this case, the direction of possible bias is difficult to assess, because it is unclear a priori whether the correlation between the level of innate ability and the SPE variable is positive or negative. Moreover, respondents with higher levels of innate ability may be more likely to adopt improved stoves and to make more efficient and safer use of dirty fuels, thereby using them more often. This possibility makes it even more challenging to assess the direction of potential bias, because the sign of the correlation would become ambiguous even between the level of innate ability and the use of dirty fuels in the future.

Health risk tolerance and innate ability are not an exhaustive list of omitted relevant variables. One might argue that female empowerment is also an unobserved determinant of cooking fuel usage patterns; a correlation may occur between the SPE variable and the error term if females with higher levels of empowerment are better or more confident at properly assessing probabilities related to health risks and have more bargaining power over the types of cooking technologies adopted at home. Overall, these arguments seem to suggest that, due to the possibility that a variety of relevant variables are omitted, (1) our results should be interpreted as correlative rather than causative and that (2) the direction of possible bias in our results is indeterminate.

3.5.2 Association between cooking fuel usage patterns and health

In subsection 3.2, where we examine the association between dirty fuel usage patterns and the likelihood of suffering from disease symptoms, we argue for the exogeneity of our instrument (i.e., an opportunity to access free cooking fuel) based on ethnographic evidence and assuming that a key unobservable factor in equation (3.3) is health unconsciousness. However, the instrument exogeneity assumption may be questioned, as a different unobserved factor for health could be correlated with the instrument. Innate ability, for example, may be an unobserved factor for health in that, respondents with higher levels of innate ability are more likely to take better care of their own health. At the same time, respondents with higher levels of innate ability may be able to find cooking fuels in common lands or open roads more easily independent of their dwelling location and may therefore be more likely to report that this resource is freely available. As such, a correlation may exist between the instrument and the error term (containing innate ability).

Ambient air pollution is another possible unobserved factor that may invalidate the exogeneity of the instrument. The likelihood of suffering from the disease symptoms may be influenced by the level of ambient air pollution. At the same time, the level of ambient air pollution may be associated with the distance between a dwelling and trees or forests, which influences whether the respondent has an opportunity for free firewood. In this way, the instrument may be correlated with the error term.

If the instrument exogeneity assumption is violated, the use of the instrument is unlikely to help mitigate endogeneity bias. Even worse, the resulting bias in our estimates could be larger in magnitude than it is when we do not use the instrument. For this reason and for comparative purposes, we estimate the same models as before while not controlling for possible endogeneity and provide the results. As presented in Table 3.A2[14], the results are broadly similar to those reported in Table 3.5. Our instrumental variable results, therefore, should be interpreted with the caveat that they might be nothing more than correlational.

In subsection 3.3, we tried to control for the endogeneity of the 'cooking fuel usage pattern' variable using an instrument (an opportunity to access free cooking fuel). However, a further complication in interpreting the estimated coefficients may arise from the inclusion of two potentially endogenous SPE variables (individually or their combination) as covariates. In

[14] Various unobserved confounders in the error term may have introduced bias into this analysis. The direction of the bias will depend on the correlation between the endogenous regressor (here the fraction of days of dirty fuel usage) and the error term, which is the question under consideration. As mentioned above, a possible omitted variable may be the degree of health unconsciousness. Health-conscious individuals tend to have better health than their health-unconscious counterparts; the latter may reflect a higher tendency toward dirty fuel usage than the former. Therefore, this variable is likely to be positively correlated with the endogenous regressor. Another omitted factor may be individuals' health vulnerability. Individuals with vulnerable health status may hesitate to use dirty cooking fuels and hence be less likely to become sick from HAP. This factor is therefore likely to be negatively correlated with the endogenous regressor in contrast to health unconsciousness. These arguments suggest that the error term contains multiple unobserved factors correlated with the endogenous regressor in opposite directions. For this reason, it seems to be difficult to determine the direction of the bias.

particular, the presence of the two endogenous SPE variables in the estimation model may introduce a bias in the estimated coefficient of the fraction of days of dirty fuel usage. Here, the direction of the bias is likely to depend on the correlation between the endogenous variables and the error term, as well as the correlation between the endogenous variables and the cooking fuel usage pattern.

It is not unreasonable to consider that the endogeneity of the SPE variables included in equation (3.4) may arise from the simultaneous elicitation of the SPE, cooking fuel usage pattern and health status variables. However, as mentioned earlier, the SPEs were elicited one period earlier than the health status and fuel usage pattern variables to avoid the issues of simultaneity and reverse causality. For this reason, a plausible threat to the exogeneity assumption of the SPE variables is unobserved individual-specific factors. For example, individuals' degree of health unconsciousness, a possible candidate for the omitted relevant variable (implicitly present in the error term of equation (3.3)), may induce a correlation between the SPE variables and the error term.

For a better understanding of the direction of bias in this case, consider the result presented in Column 1a of Table 3.5. There may be a negative association as individuals with higher degrees of health unconsciousness tend to perceive the risk to become sick to be lower than their health-conscious counterparts[15]. On the other hand, as shown in sub-section 3.1, $SP(s = 1|c)$ ($SP(s = 1|d)$) is positively (negatively) associated with the cooking fuel usage pattern. Thus, in this case, the direction of the possible bias is difficult to assess a priori and depends on the relative strengths of the associations[16]. Similarly, as discussed in sub-section 5.1, the presence of other omitted relevant variables such as innate ability and women

[15] For example, the former category of individuals tends to evaluate a low risk of becoming sick from dirty cooking fuels, thus assigning a lower value to $SP(s = 1|d)$, than their health-conscious peers.
[16] The derivation and related discussion of the direction of bias if the omitted relevant variable is the 'degree of health unconsciousness' is presented in detail in the appendix (3.A1) as supplementary material. For the derivation, we borrow the idea from Acemoglu et al. (2001).

empowerment may result in ambiguity in assessing the direction of the association between the SPE variables and error terms. Therefore, we may conclude that the direction of bias due to the presence of endogenous SPE variables in our case may remain indeterminate and depends on the relative strength of the individual components.

3.6. Conclusion

In this study, we analyze a unique dataset from rural Indian households to examine individuals' SPEs of becoming sick with symptoms typically resulting from HAP exposure and their role in cooking fuel usage patterns. We also investigate relationships between individuals' cooking fuel usage patterns vis-à-vis their health status. Our results support the hypothesis that individuals' expectations and cooking fuel usage patterns are significantly associated, although the magnitude might be small. The results also show that the usage of dirty cooking fuel and the likelihood of suffering more from physical symptoms related to HAP are positively and significantly correlated. Based on the estimated coefficients, the policy simulation analysis suggests that in response to information hypothetically provided about the health risks of dirty cooking fuels, individuals will tend to reduce dirty fuel usage and consequently enhancing the associated health status. However, the magnitude of the change has been found to be limited in our analysis.

Our results should be interpreted with caution for various reasons. 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. Third, the results may not be generalizable to rural India as a whole because they are not necessarily an unbiased representation of the population (that is, all individuals in rural India). However, several areas of West Bengal may be similar to our study area in terms of ethno- and sociodemographic features, which are located in the proximity of an urban metropolis. As such, the findings reported here are expected to be valid for such areas.

Finding a valid instrument for SPEs and eliciting information on innate abilities and female empowerment may be a challenge, which we wish to address in our future endeavors. 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. To comprehensively assess the role of SPEs on health, we plan to focus on the long-term health effects of HAP, which may demand the involvement of professional medical teams during the elicitation of responses. We plan to extend our research in this direction in future

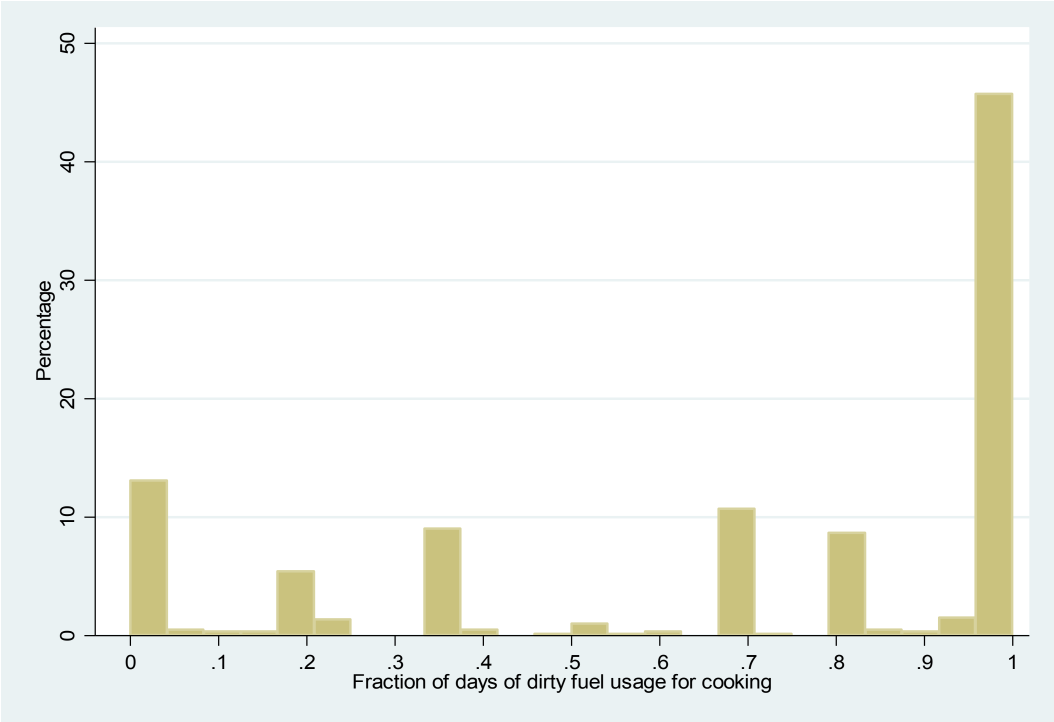

Figure 3.1. Distribution of fraction of days of dirty fuel usage for cooking

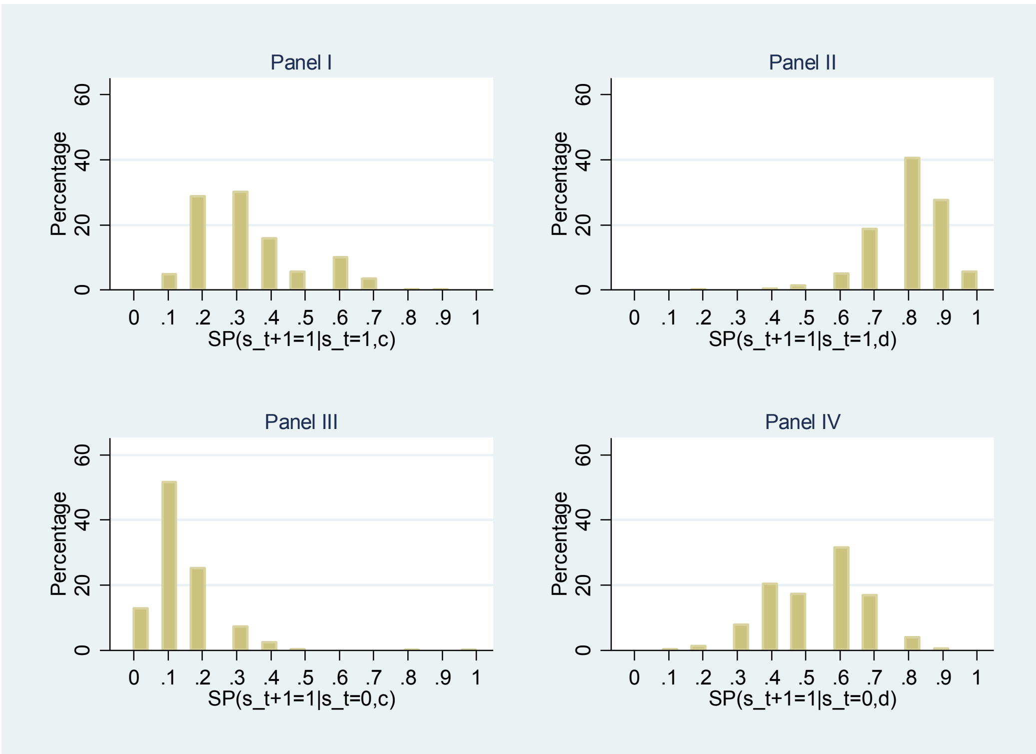

Figure 3.2. Distribution of the elicited SPEs

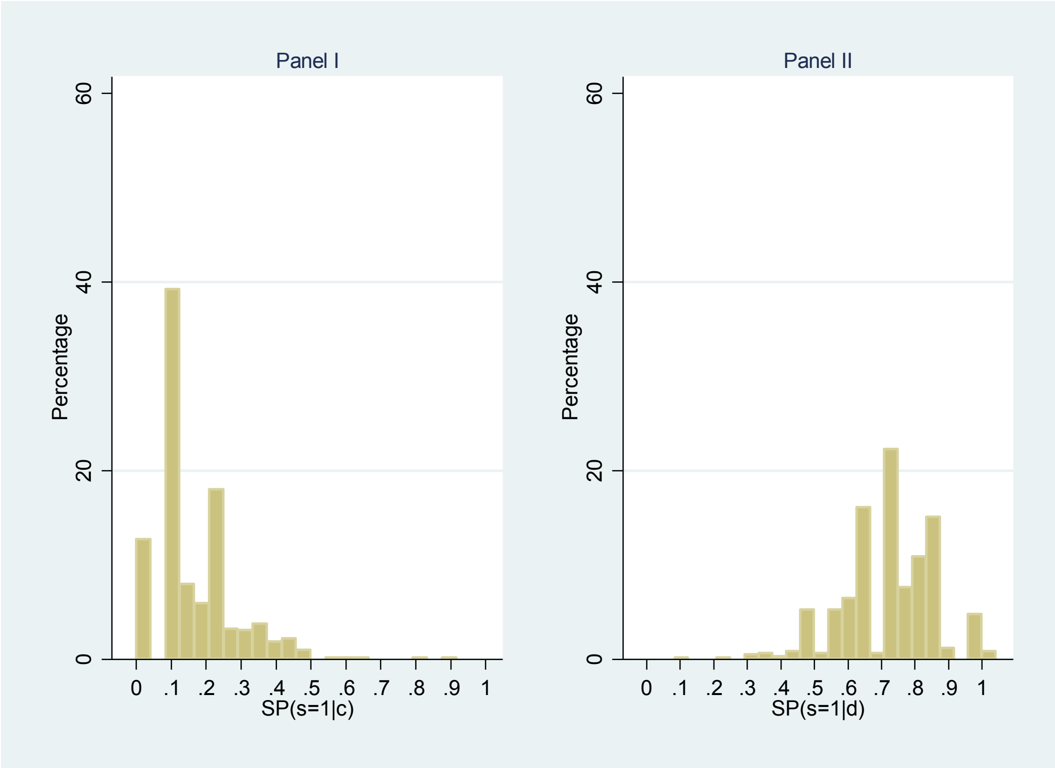

Figure 3.3. Distribution of the equilibrium SPEs

Table 3.1. Descriptive statistics

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				
$SP(s_{t+1} = 1 \mid s_t = 0, c) - SP(s_{t+1} = 1 \mid s_t = 0, d)$	-0.40	0.15	-0.8	0.5
$SP(s_{t+1} = 1 \mid s_t = 1, c) - SP(s_{t+1} = 1 \mid s_t = 1, d)$	-0.47	0.16	-0.8	0.7
$SP(s = 1 \mid c)$	0.17	0.12	0	0.91
$SP(s = 1 \mid d)$	0.73	0.14	0.11	1
$SP(s = 1 \mid c) - SP(s = 1 \mid d)$	-0.56	0.17	-1	0.52
$SP(s = 1 \mid c)/SP(s = 1 \mid d)$	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. The sample size is 557.

Table 3.2. 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 3.3. Estimation results of the cooking fuel usage pattern equation with the elicited SPEs (average marginal effects)

	(1)	(2)	(3)
$SP(s_{t+1} = 1 \vert s_t = 0, c) - SP(s_{t+1} = 1 \vert s_t = 0, d)$	0.161*	0.137	
	(0.095)	(0.092)	
$SP(s_{t+1} = 1 \vert s_t = 1, c) - SP(s_{t+1} = 1 \vert s_t = 1, d)$	-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
χ^2	282.67	284.09	284.59

Note: This table provides estimation results for equation (3.1), 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 3.4. Estimation results for the cooking fuel usage pattern equation with 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	
χ^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 3.5. Estimation results for the health status equation 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)
Residuals from the fuel usage equation	-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	
χ^2	128.77		128.77		132.47	

Note: This table provides estimation results for equation (3.3), 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 3.6. Policy simulation results

	Baseline	Information about the health risks of dirty fuels
Panel I: On fraction of days of dirty fuel usage		
Predicted probability		
25%	0.504 (0.021)	0.481 (0.024)
50%	0.739 (0.015)	0.706 (0.013)
75%`	0.879 (0.011)	0.857 (0.009)
Mean	0.678 (0.01)	0.653 (0.01)
% with reduced dirty fuel usage	87.07%	
% with increased dirty fuel usage	12.93%	
Average change in predicted probability	-0.025 (0.001)	
conditional on reduced dirty fuel usage	-0.030 (0.001)	
conditional on increased dirty fuel usage	0.013 (0.002)	
Panel-II: On probability of being sick		
Predicted probability		
25%	0.745 (0.019)	0.719 (0.019)
50%	0.911 (0.008)	0.897 (0.007)
75%	0.963 (0.003)	0.957 (0.003)
Mean	0.823 (0.008)	0.805 (0.009)
Average change in predicted probability		-0.018 (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

Appendix

Appendix 3.A1

Direction of the bias in the effect of dirty fuel usage when other endogenous variables elicited in the prior period are included

In Equation (3), the variables cooking fuel usage pattern and the SPEs (two in number) may be endogenous. We have tried to control for the endogeneity of the cooking fuel usage pattern using an instrumental variable. However, the potentially endogenous SPE variables may introduce bias in assessing the effect of the cooking fuel usage pattern on the likelihood to become sick. To assess the direction of the bias in such a case, for the sake of simplicity, we assume that the cooking fuel usage pattern and SPE variables are the only control variables in Equation (3). Denoting $SP(s = 1|c)$ as Z_1 and $SP(s = 1|d)$ as Z_2, Equation (3), for *i* = 1, 2, …, n , may be represented as:

$$s_i^* = \gamma_0 + \gamma_1 w_i + \pi_1 Z_{1i} + \pi_2 Z_{2i} + u_i \qquad \text{(i)}$$

By standard argument,

$$plim\, \widehat{\gamma_1} = \gamma_1 + \frac{cov(\widetilde{w}, u)}{var(\widetilde{w})}$$

$$or, plim\, \widehat{\gamma_1} = \gamma_1 - \kappa_1 \frac{cov(Z_1, u)}{var(\widetilde{w})} - \kappa_2 \frac{cov(Z_2, u)}{var(\widetilde{w})} \qquad \text{(ii)}$$

where, κ_1, κ_2 and $\widetilde{w}$ are respectively the coefficients of Z_1, Z_2 and the residual of the auxiliary equation $w = \kappa_1 Z_1 + \kappa_2 Z_2 + \widetilde{w}$. Therefore the direction of the bias will depend on the signs of $\kappa_1, \kappa_2, cov(Z_1, u)$ and $cov(Z_2, u)$.

Defining $var(Z_1) = \sigma_{11}, var(Z_2) = \sigma_{22}$, $\sigma_{w1} = cov(Z_1, w), \sigma_{w2} = cov(Z_2, w)$ and $cov(Z_1, Z_2) = \sigma_{12}$, we can find the estimates of the coefficients of the auxiliary equation as:

$$\begin{bmatrix} \kappa_1 \\ \kappa_2 \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{w1} \\ \sigma_{w2} \end{bmatrix}$$

yielding

$$\kappa_1 = \frac{\sigma_{22}\sigma_{w1} - \sigma_{12}\sigma_{w2}}{\sigma_{11}\sigma_{22} - \sigma_{12}^2}$$

$$\kappa_2 = \frac{\sigma_{11}\sigma_{w2} - \sigma_{12}\sigma_{w1}}{\sigma_{11}\sigma_{22} - \sigma_{12}^2}$$

From subsection 3.1, we get that Z_1 and dirty fuel usage have a positive association (see Table 4) as individuals with higher Z_1 values tend to become more averse to using clean cooking fuel and hence are likely to increase dirty fuel usage. Therefore, it is likely that $cov(w, Z_1) > 0$. Analogously, from Table 4, we may conclude that $cov(w, Z_2) < 0$. Furthermore, in the data, we find that Z_1 and Z_2 are weakly associated with a correlation coefficient 0.083, suggesting that $\sigma_{12} > 0$. Under these conditions, we have $\kappa_1 > 0$ and $\kappa_2 < 0$.

Since the SPE variables are elicited one period prior to the elicitation of the health status and fuel usage pattern, the possible source of endogeneity is the unobserved individual-specific factors implicitly contained in the error term of Equation (i). As discussed in sub-section 5.2, consider the individual unobserved factor as the degree of health unconsciousness defined by δ_i. In presence of such an unobserved factor, Equation (i) may be re-written as:

$$s_i^* = \gamma_0 + \gamma_1 w_i + \pi_1 Z_{1i} + \pi_2 Z_{2i} + (\tau\delta_i + \varepsilon_i) \qquad \text{(iii)}$$

where τ stands for the coefficient of the degree of health unconsciousness with respect to the propensity to become sick. We also assume that $cov(Z_1, \varepsilon) = 0$ and $cov(Z_2, \varepsilon) = 0$. Health unconscious individuals (large δ) tend to have a higher propensity to become sick than their health- conscious counterparts. Therefore, it is likely that $\tau > 0$.

Intuitively, it is likely that health unconscious (large δ) individuals evaluate the risk of becoming sick to be lower than their health conscious (small δ) counterparts. For instance, health unconscious individuals tend to evaluate a low risk of becoming sick from dirty cooking fuels than their health-conscious peers, thus expressing lower values of Z_2. Thus, it may be expected that $cov(Z_2, \delta) < 0$. Under the assumption that $cov(Z_2, \varepsilon) = 0$, we find that $cov(Z_2, u) < 0$. Analogously, $cov(Z_1, u) < 0$ as $cov(Z_1, \delta) < 0$.

Recalling Equation (ii),

$$plim\, \widehat{\gamma_1} = \gamma_1 - \kappa_1 \frac{cov(Z_1, u)}{var(\widetilde{w})} - \kappa_2 \frac{cov(Z_2, u)}{var(\widetilde{w})} = \gamma_1 - \alpha$$

where α is the magnitude of the bias. We obtain that $\kappa_1 > 0$ and $\kappa_2 < 0$. Intuitively, $cov(Z_1, u) < 0$ and $cov(Z_2, u) < 0$. Therefore, in Equation (ii), $\kappa_1 \frac{cov(Z_1,u)}{var(\widetilde{w})} < 0$ but $\kappa_2 \frac{cov(Z_2,u)}{var(\widetilde{w})} > 0$. Since the directions of the two terms in the bias component (α) are in opposite direction, we may conclude that the direction of the bias may remain indeterminate a priori and will depend upon the relative strengths of the individual terms.

Appendix 3.A2

Selected Part of the Questionnaire Used for the Field Survey

<table>
<tr><th colspan="6">Cooking fuel usage related information</th></tr>
<tr><td rowspan="7">1</td><td colspan="2" rowspan="7">In the <u>last 30 days before last month</u>, how many days did you use the following fuel for cooking?(Please mention for each fuel type. You can mention 0 if you have not used that variety of fuel)</td><td colspan="3">Fuel type</td><td>Days</td></tr>
<tr><td colspan="3">Electricity</td><td></td></tr>
<tr><td colspan="3">LPG</td><td></td></tr>
<tr><td colspan="3">Kerosene</td><td></td></tr>
<tr><td colspan="3">Coal/Charcoal</td><td></td></tr>
<tr><td colspan="3">Solid fuels like cow dung cakes, straw</td><td></td></tr>
<tr><td colspan="3">Firewood</td><td></td></tr>
<tr><td></td><td colspan="2"></td><td colspan="3">Others (please specify)</td><td></td></tr>
<tr><td>2</td><td colspan="2">Do you have some opportunity to get/collect cooking fuel for free?</td><td colspan="3">Yes</td><td>No</td></tr>
<tr><th colspan="6">Health related information</th></tr>
<tr><td rowspan="3">3</td><td colspan="2" rowspan="3">Did you experience these problems mentioned below in the last 30 days?</td><td>Dry cough</td><td>Yes</td><td colspan="2">No</td></tr>
<tr><td>Sore/runny eyes</td><td>Yes</td><td colspan="2">No</td></tr>
<tr><td>Difficulty in breathing</td><td>Yes</td><td colspan="2">No</td></tr>
<tr><th colspan="6">Subjective Probability related Information</th></tr>
<tr><td colspan="6">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.”.</td></tr>
<tr><td>4</td><td colspan="5">Consider a hypothetical individual who is identical to you. Imagine that there are options of primary fuel for cooking. The current health condition is mentioned below. In each such situation, how likely is it that she will remain (become) sick in the next 3 months if she uses the following fuels?</td></tr>
<tr><td colspan="6">To the enumerator: Please ask only a likelihood of <u>Sick</u>.</td></tr>
<tr><td colspan="2">Description of health status</td><td colspan="2">Case-I: She is <u>Healthy</u></td><td colspan="2">Case-II: She is <u>Sick</u></td></tr>
<tr><td colspan="2">Fuel used for cooking in all the 30 days in the last month</td><td>LPG/ Kerosene/ Electricity</td><td>Firewood/ Cow dung cakes/ Coal</td><td>LPG/ Kerosene/ Electricity</td><td>Firewood/ Cow dung cakes/ Coal</td></tr>
<tr><td>a</td><td>Sick</td><td></td><td></td><td></td><td></td></tr>
<tr><td>b</td><td>Healthy</td><td></td><td></td><td></td><td></td></tr>
</table>

Appendix 3.A3

Table 3.A1. Estimation results for the cooking fuel usage pattern equation with equilibrium SPEs (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**	
		(0.093)	
$SP(s=1\|c)/SP(s=1\|d)$			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.

Appendix 3.A4

Table 3.A2. Estimation results for the health status equation with the equilibrium SPEs (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 5 in that the residuals from the cooking fuel usage pattern equation 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)

Chapter 4. Economics of clean air: Valuation of reduced health risks from Household Air Pollution - A study of rural Indian households

4.1. Introduction

Household Air Pollution (HAP, hereafter) caused primarily by the incomplete combustion of dirty cooking fuels coupled with inefficient cooking practices, is a salient environmental and health risk particularly in rural areas of developing countries. Estimates of the burden of HAP in India alone show approximately 1.04 million premature deaths as well as 31.4 million disability adjusted life years (Balakrishnan et al., 2014).

Health risks associated with HAP can be adequately prevented through exposure reduction via the usage (adoption) of modern cooking fuels (technologies). However, regardless of the expected health benefits from such behaviour, several factors such as, liquidity constraints (e.g., Bensch et al., 2015) and a failure to perceive the seriousness of such health risks (Mobarak et al., 2012) may pose significant barriers to the usage (adoption) decision. Thus, it transpires that the implementation of suitable intervention policies to mitigate the HAP-related health risks may face numerous logistic challenges that in turn may reduce its effectiveness. To ensure the effectiveness of such interventions, understanding the attitude and/or preference of the potential beneficiaries towards the mitigation of such health risks often reflected through their perceived private health benefits from the reduced exposure to HAP, becomes necessary (Shannon et al., 2019).

Perceived private health benefits of the individuals may be evaluated by estimating the valuation of reduction in health risks related to HAP as has been attempted in this study. In

---

The version of the chapter submitted for interim reporting of PhD is published as a discussion paper in the WINPEC Discussion Paper series: Chattopadhyay, M. (2021). *Economics of clean air: Valuation of reduced health risks from Household Air Pollution-A study of rural Indian households* (No. 2119). This is a single authored chapter.

particular, we attempt to assess the individuals' valuation of reduced health risk from HAP exclusively, derived from a hypothetical improvement in household air quality using a stated preference approach.

Since the valuation of reduced health risk is non-amenable to market valuation, we have employed the contingent valuation method (CVM, hereafter). CVM is particularly useful for estimating the economic valuation of non-market goods owing to its greater flexibility in creating specific markets with proposed improvement (Andersson et al., 2016). Of late, the application of CVM to estimate the economic valuation of environmental resources and/or health risks related to ambient air pollution, smog mitigation and nuclear power among others, have garnered attention (e.g.: Du and Mendelsohn, 2011; Sun and Zhu, 2014; Sun et al., 2016).

Compared to the wealth of literature related to environmental valuation , research on the valuation of economic cost of HAP is relatively small (Jeuland et al., 2015). To the best of our knowledge, only the study by Shannon et al. (2019) have implemented the CVM approach to estimate the valuation of HAP-related health risks in tandem with risks from water contamination, in rural India. In this chapter, we have tried to extend the literature related to the valuation of environmental health risks from HAP exclusively using CVM. In particular, we exploit the double bounded dichotomous choice (DBDC, hereafter) approach under CVM.

Although DBDC method is often preferred in literature to estimate the WTP because of its relative statistical efficiency[1] (Hanemann et al., 1991), its criticisms are widespread. DBDC responses may suffer from anomalies such as starting point bias (Gelo and Koch, 2015).

[1] To estimate the individuals' WTP, elicitation of bid responses using a single dichotomous question (single bounded dichotomous choice, SBDC approach) has been preferred over other approaches for being incentive compatible (Carson et al, 1998). However, critics argue that since individual preferences are developed through repetition and practice, the estimated WTP obtained by SBDC approach may suffer from uncertainty due to lack of well-defined preference (Plott et al., 2005). Thus, to incorporate the dynamic aspect of individual preferences, the DBDC approach is introduced, where the elicitation of bid response is supplemented with a follow-up dichotomous choice question. This improves the statistical efficiency of the estimated WTP by exploiting the combination of responses from both the rounds (Cameron and Quiggin, 1994).

Starting point bias in DBDC model arises when individuals uncertain about the true value of the good, tend to attribute the initial bid to be the true value. Therefore, as an extension of estimating the individuals' WTP for reduction in HAP in this study. we attempt to explore the presence and source of starting point bias in our DBDC model.

We analyse a unique contingent valuation dataset from 557 survey respondents in rural West Bengal, India. The annual mean WTP from the double bounded dichotomous approach is estimated to be INR 886.59 (~USD 14.30) that accounts for approximately 1.06% of annual household income. The ratio of the estimated mean annual WTP to average household expenditure lies in a comparable range with previous literature. Furthermore, the results demonstrate the presence of anchoring effect validating the existence of starting point bias in our DBDC model.

Given the evidence of the influence of individual-specific covariates in the individuals' WTP decision, we further attempt to explore the within-sample heterogeneity of the estimated mean WTP based on contextually relevant judiciously selected covariates. Our analysis of the within-sample heterogeneity of estimated mean WTP further indicates that any variation in individual-specific covariates may result in sufficient fluctuation in estimated mean WTP within the sample. This analysis may be particularly useful in designing effective policies for smooth implementation of interventions targeted towards HAP mitigation.

The contribution of this study is threefold. First, it attempts to provide a direction to understand the stated preference for the reduction in health risks related to HAP exclusively by assessing the WTP values. This is possibly needed to understand the individual preferences and attitude for reduction in HAP for successful implementation[2]. Second, the analysis of

[2] Studies have elaborated that despite the implementation of several interventions targeted for low-income households in developing countries, few have delivered desired results. This is mainly due to the tendency of the households to reduce the sustained usage of modern cooking technology due to lack of maintenance or reversion

within-sample heterogeneity of the estimated mean annual WTP further enables us to recommend policies like generating public awareness about HAP risk and targeting potential beneficiaries based on observable characteristics. Such policy is expected to ensure smooth implementation and enables one to assess the effectiveness of intervention programs to reduce HAP. Finally, as a minor contribution to literature on DBDC models, this study attempts to explore the presence and potential cause of starting point bias in DBDC model in the context reduction in HAP in developing countries.

The remainder of the chapter is organized as follows. Section 2 describes the survey methodology and variables considered for the study along with their summary statistics. It also elaborates on the methodology to elicit the bid responses. The next section presents our empirical models and estimation results. Section 4 presents the analysis of the within-sample heterogeneity of the estimated mean WTP based on contextually relevant covariates and its policy implications. The chapter concludes by discussing policy implications and directions of future research in section 5.

4.2. Introduction to dataset

4.2.1. Survey design

This study uses the data collected during December 2017 – January 2018. Out of the 596 individuals interviewed via door-to-door interview method in the first round, 588 respondents completed the survey. For the analysis, we exclude from our sample the respondents who have no spouse or provide no information on the spouse, thus, our effective sample size reduces to

---

to former behaviour once, the promotion period is over (Hanna et al., 2016). Therefore, for successful and sustained implementation, it is necessary to understand the attitude and preference of the potential beneficiaries about reduction in HAP-related health risks before such interventions,

557. Table 4.1 presents the descriptive statistics of the variables used in the study which is explained in detail in the following section.

<Table 4.1 approximately here>

4.2.2. Description of the variables and their summary statistics

Self-reported health status

The association of HAP with various respiratory and vision-related diseases is well-established in literature (Smith and Pillarisetti, 2017). To account for this, following, Hanna et al. (2016) and based on our initial analysis of pilot study data, we have selected the three most common symptoms namely, dry cough, sore or runny eyes and difficulty in breathing in the final questionnaire .

The self-reported health status of the respondent refers to whether the respondents have experience at least one of three above mentioned minor yet common symptoms caused by HAP in the last 30 days. These physical symptoms are indeed, found to be prevalent among the respondents; around 76% of them have experienced at least one of the above-mentioned physical symptoms in the last 30 days (see Table 4.1).

Cooking fuel usage pattern

In this study, the cooking fuel usage pattern of the household is an important covariate and is represented by the fraction of days the dirty cooking fuel has been used for cooking in a 30-day period. We have computed the fraction of days of dirty fuel usage using the information on the number of days the respondent used coal/charcoal, solid biomass fuels, and firewood[3] in 30 days prior to the previous month.

The fraction of days of dirty fuel usage is found to be 0.68 on an average, suggesting the higher prevalence of dirty fuel usage in rural India (see Table 4.1). At a cursory glance, this may suggest that despite the significant burden of HAP related symptoms, respondent households tend to show a relatively high-risk behaviour related to cooking fuel usage. This also supports the necessity of such valuation assessment.

Perception of health risks related to HAP

The approach to incorporate individuals' risk perception in the form of verbal scales (like Likert scale) to represent subjective likelihood is often criticised due to non-verifiability of the assessment and difficulty with inter-personal comparability of the subjective risk (Anglewicz and Kohler, 2009). Therefore, we include the individuals' perceived subjective health risk related to HAP in the form of probabilistic expectations on a scale of zero to ten elicited through interactive elicitation method using visual aids.

We presume that the respondents' perceived risk of suffering from above-mentioned HAP related symptoms in the next 30 days depends only on their current health status and fuel usage. In other words, the risk of being sick[4] is assumed to follow a first-order Markov process conditional on cooking fuel usage (Ross, 1996). Based on this assumption, we have elicited the respondents' perceived likelihood of becoming sick from HAP-related physical symptoms in

[3] Although WHO (2018) classified kerosene to be dirty cooking fuel, we have conducted our survey in 2017-18, before this classification was published. In our study, we have consistently followed the nomenclature referred by Duflo et al. (2008) among others, where kerosene is classified as clean cooking fuel. We have explained in detail that our results do not change much even if we had considered kerosene to be dirty cooking fuels in Chapter 3.
[4] By "sick," we refer only to the situation of having suffered from at least one of the three HAP-related physical symptoms—dry cough, sore or runny eyes, and difficulties in breathing in a 30-day period.

the next 30 days from dirty fuel usage for two alternative situations, namely, currently being *sick* and *not sick*[5].

Using these two elicited perceptions conditional on dirty fuel usage, we calculate the equilibrium distribution of the Markov process denoted as $SP[s = 1|d]$[6]. Under the assumption of first-order Markov dependence, this represents the perception about the long-term fraction of periods during which the respondent would be sick provided that she uses dirty fuels. In other words, $SP[s = 1|d]$ may be interpreted as the perceived health risk from dirty fuel usage. Likewise, we derive the other equilibrium distribution of the Markov process conditional upon clean fuel usage denoted by $SP[s = 1|c]$. We include the difference in the perceptions of health risks from dirty and clean fuel usage (i.e., $SP[s = 1|d] - SP[s = 1|c]$) in our analysis. This difference, being non-negative, may be interpreted as the individuals' perceived increase in health risk from using dirty cooking fuels instead of clean cooking fuels.

Although the respondents have exhibited relatively low risk-averting behaviour, they seem to perceive an association between dirty fuel usage and deterioration of their health. The mean difference of 0.57 between $SP[s = 1|d]$ and $SP[s = 1|c]$ suggests that individuals on an average perceive that the dirty fuel usage is 57 percentage points more likely to degrade their health than the clean fuel usage.

Methodology to elicit the bid responses & their description

As our survey respondents are individuals from a rural area, it has been presumed that they are not much familiar with the sophisticated preventive measures from HAP. Therefore, to facilitate their understanding, the enumerators referred to the preventive device to be something

[5] A detailed description of elicitation of individuals' subjective risk perception about HAP-related health risks in probabilistic form through interactive elicitation method using visual aids is presented in Chattopadhyay et al. (2021)

[6] We have discussed the derivation and description of $SP[s = 1|d]$ in detail in subsection 3.2.2 in Chapter 3 of this dissertation.

similar to electric chimneys or exhaust fans that will reduce the incidence and extent of smoke in the cooking area, as an example.

To control for any hypothetical bias arising from over-stating (under-stating) WTP from the true value, we have tried to present the scenario as realistic as possible. Before the elicitation, the enumerator team elaborated the benefit that the respondent may accrue and the cost they may have to incur for using the hypothetical preventive device[7]. For elicitation of bid responses, the following question was asked to the respondents:

"Are you willing to pay [initial bid] per year for using this preventive device?"

Informal interview during the pilot test helped us to choose three levels of initial bid: INR 100, INR 500 and INR 750. To avoid the problem of initial bid bias, we have randomly assigned these initial bids to the respondents. Following the standard norms of the DBDC approach, the follow-up bid has been doubled (halved) if the respondents give affirmative (negative) responses for the initial bid.

Characteristics of the bid responses

Since the bids are randomly assigned, following Imbens and Rubin (2015), we try to ensure the balance among the covariates in the assignment mechanism. Table 4.2 indicates that the number of individuals assigned to each initial bid level is more or less the same. Figure 4.1 presents the histograms along with the kernel density estimates of the estimated propensity score for each bid category[8] resulting from such random allocation of the survey units in three categories.

[7] The instructions and survey instruments used to elicit the bid responses are presented in the appendix (4.A1).
[8] The estimated propensity scores are obtained through the multinomial discrete choice logistic regression model, as a natural generalization of the method suggested by Imbens and Rubin (2015), p 286-287, to accommodate the three bid categories as the dependent variable.

<Figure 4.1 approximately here>

A visual inspection of Figure 4.1 suggests that the estimated propensity scores in the three categories are more or less similar and are lying in the range of 0.1 to 0.6 with the mode in the range of 0.3 to 0.4. Thus, it may be concluded that the propensity score matching has been ensured in the three categories resulting in covariate balance among the three bid groups. Such covariate balance among the bid groups ensures that the groups are otherwise similar, and thus the bid responses are not biased in favour of any particular group.

Table 4.2 presents the distribution of proportion of acceptance across various bid levels. We find a negative relationship between bid prices and likelihood to accept the bid; as the levels of bid increases, the proportion of acceptance to pay the bid level decreases. For example, out of 197 individuals who are assigned the initial bid of INR 100, approximately 68% of them have expressed willingness to pay both the initial and the follow-up bid. But the share of individuals having affirmative responses in both rounds steadily declines to 8.65% (out of 185) when the initial bid assigned is INR 750.

<Table 4.2 approximately here>

Other covariates

In addition to the respondents' health status, cooking practice and perception of health risks, individual and household-specific factors may affect the respondents' valuation of reduced health risks. Therefore, in our model, we control for a set of factors including number of cooks (surrogate of household size), total monthly household expenditure (surrogate for household income), age, respondents' years of schooling, dummy for holding decision-making authority in the household (respondent holds the household decision-making authority), dummies for the occupation of the spouse (spouse works in informal sector and that in the agricultural sector), dummy for the location of the cooking area (cooking area located inside the dwelling area), dummy for ventilation (cooking area has ventilation facility) and dummy for the ownership of television.

4.3.Estimation models and results

The theoretical underpinning of the CVM essentially follows from the cost-benefit analysis. A rational individual is willing to pay the proposed bid for any environmental resource and/or risk reducing goods (services) if and only if his or her utility from the resources and/or risk reduction is at least equivalent to the utility without them. The equivalent analogous condition states that a rational individual will agree to pay the proposed bid if and only if the willingness to pay for such goods (services) is at least equal to the proposed bid (Donfuet et al., 2014). It may be noted that the estimation of the mean WTP is dependent on the distributional assumption of the stochastic component of the utility function. In what follows, we present our DBDC model.

4.3.1. DBDC approach

Estimation model

To estimate the individuals' valuation of reduced health risks from reducing HAP exposure and thereby improving indoor air quality, we try to estimate a DBDC model. As a benchmark, we start with the most generalised DBDC model allowing the individuals' WTP to vary over the two rounds. Let the latent WTP (expressed in the logarithmic form[9]) of individual *i* in round k (w_{ik}^*) be a linear function of her prior experiences of HAP-related symptoms (s_i), cooking fuel usage pattern ($cook_i$), individual and household characteristics (z_i), and the perception of

[9] The lognormal specification of WTP distribution appears to fit the skewed pattern of survey responses in a better way (Herriges and Shogren, 1996). It also allows us to ignore the negative WTP by restricting the distribution in the interval$(0, +\infty)$.

health risks ($risk_i$) where $k = \{1,2\}$ representing initial and follow-up rounds respectively. In particular, the individual's WTP given the observed characteristics may be specified as:

$$\begin{bmatrix} w_{i1}^* \\ w_{i2}^* \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_i' \\ \boldsymbol{X}_i' \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} u_{i1} \\ u_{i2} \end{bmatrix}, \tag{4.1}$$

where $X_i = [1 \quad \boldsymbol{z_i} \quad cook_i \quad s_i \quad risk_i]$and $\boldsymbol{u_i} = \begin{bmatrix} u_{i1} \\ u_{i2} \end{bmatrix}$ is the idiosyncratic error term uncorrelated with X_i. We further assume, with the usual notation $\begin{bmatrix} u_{i1} \\ u_{i2} \end{bmatrix} \sim N_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}\right)$. Since the individual's WTP is latent, it is estimated using the observed bid responses. Assuming $\begin{bmatrix} Y_{i1} \\ Y_{i2} \end{bmatrix}$ to be individual i's bid response, let $\mathbb{I}(\cdot)$ be an indicator function that links the individual's latent WTP to bid response in the following way:

$$Y_{ik} = \mathbb{I}(w_{ik}^* \geq b_{ik}), \tag{4.2}$$

The indicator function $\mathbb{I}(\cdot)$ takes the value 1 if $w_{ik}^* \geq b_{ik}$ where, b_{ik} is the bid value of individual i for round k, $k = 1,2$.

However, Cameron and Quiggin (1994) argue that in any DBDC survey, if the response to follow-up bid is elicited immediately after the initial contingent valuation question, it may be assumed that the response to the follow-up bid is drawn from the same distribution as that for the initial bid. In other words, the latent WTP of an individual in both the rounds are identical ($w_{i1}^* = w_{i2}^*$) and any variation in the model is caused by the randomness of the error terms. Furthermore, Cameron and Quiggin (1994) argue in favour of the model where the latent WTP is identical in both the rounds, but the error terms have a non-unitary correlation. Under these cross-equation parametric restrictions, the model specified in equations (4.1) reduces to the following restricted bivariate probit model.

$$\begin{bmatrix} Y_{i1} \\ Y_{i2} \end{bmatrix} = \mathbb{I}\left(\begin{bmatrix} \boldsymbol{X_i'\beta} + u_{i1} \\ \boldsymbol{X_i'\beta} + u_{i2} \end{bmatrix} \geq \begin{bmatrix} b_{i1} \\ b_{i2} \end{bmatrix}\right), \tag{4.3}$$

with $\begin{bmatrix} u_{i1} \\ u_{i2} \end{bmatrix} \sim N_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix}\right)$. Equation (4.1) as well as equation (4.3) suggest for the joint estimation of parameters through the maximum likelihood estimation method.

Estimation results

<Table 4.3 approximately here>

First we try to estimate the model without imposing the restrictions that the parameters are equal in both the equations. In other words, we try to estimate the set of equation shown in (4.1) jointly using an unrestricted bivariate probit model. Following the standard norms in literature (e.g., Du and Mendelsohn, 2011) we report only the estimation results to the initial bids in Table 4.3[10]. We start with the individual- and household-specific covariates (column 1) and sequentially incorporate the covariates related to health (column 2) and risk perception (columns 3) in the estimation model.

In line with the previous literatures, the results show a negative price effect that remains uniform across all model specifications ($p < 0.01$) Several other robust patterns in the estimation results are also observed once we increase the number of controls.

We find that individuals with higher household income tend to have a higher likelihood to accept the bid. The estimated coefficient of household expenditure (surrogate for income) remains consistent across the three model specifications for the unrestricted model, as can be seen in Table 4.3. Furthermore, individuals' experience of being sick with HAP related symptoms is positively and significantly associated with her probability of paying the proposed bids. In addition, households having a higher fraction of days of dirty fuel usage have a lower tendency to pay the proposed bid. This may indicate that individuals' cooking practices may have some role in individuals' valuation of the reduced health risks from HAP. Besides,

[10] We present the results of the initial bid and final bid of the unrestricted bivariate probit model in Appendix 4.A1

individuals holding household decision-making authority have a higher tendency to pay the proposed bid. Interestingly, younger individuals tend to pay the proposed bid more in comparison to the older individuals[11]. Interestingly the correlation coefficients of the two responses for the three model specifications are statistically significant (see Table 4.3).

Researchers have proposed that if the error terms of the two bid responses are sufficiently correlated, then efficiency gain may be obtained by restricting the means and variances to be same across the two equations, provided the restrictions are supported by data (e.g.:Jeanty et al., 2007). Therefore, before the implementation of the restriction $w_{i1}^* = w_{i2}^*$ in the estimation model, we need to test whether this restriction is statistically valid in our data. Following Jeanty (2007), we test the null hypothesis that the coefficients of the covariates are equal in the set of equation in (4.1) using the Wald test. The test statistics for the three models are obtained as 7.50 (p-value: 0.8046), 7.76 (p-value: 0.9091) and 12.57(p-value: 0.6356) for models 1, 2 and 3 respectively. This suggests that the null hypothesis that the coefficients of the covariates are equal in the initial round and follow up round presented in equation (4.1) may not be rejected at 10% level and the restricted models are preferable.

We impose the cross-parametric restrictions and present the results of the restricted bivariate probit in columns 4, 5 and 6 in Table 4.3. Similar to the estimation using the unrestricted model, we sequentially introduce the socio-demographic, health, and perception related variables in our estimation model.

As observed from Table 4.3, the results of the restricted models are qualitatively similar to those in the unrestricted model. We find that the negative association of bid responses with the likelihood to accept the bid to hold true (p <0.01) in the restricted model as well. In addition, we find that households with higher income tend to have a higher likelihood to accept the bid

---

[11] In conformation to literature (e.g.: Du and Mendelsohn, 2011), the estimated coefficients of the initial round response of the unrestricted DBDC models is qualitatively and quantitatively similar to that from the SBDC model (presented in Appendix 4.A2)

(p<0.1). The estimated coefficients of household expenditure although lower than that in the unrestricted model, are stable across the three model specifications in the restricted model. Furthermore, individuals' past experience of sickness related to HAP, cooking fuel usage patterns, age, holding decision-making authority are significantly associated with the likelihood to accept the bid. Interestingly, individuals' risk perception of HAP is statistically significant in the restricted model (p<0.01). Individuals having higher perception of HAP-related risks tend to accept the bid more.

4.3.2. Starting point bias in DBDC model

Estimation model

Despite the advantages, the criticisms of DBDC are widespread. For example, empirical evidence suggests that WTP in two rounds of DBDC approach are often driven by the heterogeneity in preference, which results in divergence in WTP in both rounds (Herriges and Shogren, 1996). The key explanation behind this divergence is starting point bias where individuals' response in follow up round is dependent on that of the initial round. Starting point bias may be broadly categorised into anchoring effect and shift effect; both of which arise when individuals are uncertain about the true value of the non-market good and perceive the initial bid to be the true value.

Anchoring effect arises when uncertain individuals update their follow-up WTP in a Bayesian perspective, conditioned on their prior beliefs of WTP and initial bid (Herriges and Shogren, 1996). Mathematically, under the anchoring effect, the WTP in follow-up round, w_{i2}^{*} can be represented as: $w_{i2}^{*} = \gamma b_{1i} + (1-\gamma) w_{i1}^{*}$, where γ is the anchoring effect parameter with $0 < \gamma < 1$, assumed to be constant across all individuals; w_{i1}^{*} is the WTP in initial round and b_{1i} is the initial bid for representative individual *i*. Alternatively, under the shift effect, individuals consider the increasing (decreasing) follow-up bid to be an unfair request to pay an additional amount for the same good (indicator of an under-quality good) resulting in a tendency to understate the WTP in the follow-up round (Alberini et al., 1997). Thus, under the shift effect, the WTP in the follow-up round is specified as $w_{i2}^{*} = w_{i1}^{*} + \delta$, where δ is the shift effect parameter with $\delta < 0$. In the simultaneous presence of anchoring and shift effects, the follow-up WTP may be expressed as $w_{i2}^{*} = \gamma b_{i1} + (1-\gamma)w_{i1}^{*} + \delta$ (Gelo and Koch, 2015).

To investigate the potential sources of starting point bias, modeling the data in the panel format becomes particularly useful. Since we have two responses for each respondent, it is possible to represent the data in the panel structure defined in the following way:

$$w_{it}^{*} = \boldsymbol{X_i'\beta} + u_{it}, \qquad (4.4)$$

where, the unobserved error term $u_{it} = (\alpha_i + r_{it})$ captures the individual-specific (random) effect α_i and idiosyncratic effect, r_{it} $\forall\, t = 1,2$, representing initial and follow-up rounds respectively.

The panel data structure specification allows for the inclusion of shift effect and anchoring effect in the model. The shift effect is introduced in the model as an indicator variable denoted by $(t-1)$ $\forall\, t = 1,2$; it takes the value 1 if the response is from the second round $(t = 2)$ and 0 if the response is from the first round $(t = 1)$(Alberini et al., 1997). Alternatively, the anchoring effect is introduced in the model as $(t-1)b_{it}$ that captures the possibility that response in the follow-up question depends on the initial bid (Gelo and Koch, 2015). Therefore, in the presence of shift effect and anchoring effect, equation (4.4) may be reorganized in the following way:

$$w_{it}^{*} = \boldsymbol{X_i'\beta} + \delta(t-1) + \gamma(t-1)b_{it} + u_{it}, \qquad (4.5)$$

where, δ is the shift effect parameter, and γ is the anchoring effect parameter. If an individual's observed sequence of bid response is defined as Y_{it}, then she will be willing to pay the bid if,

$w_{it}^* \geq b_{it}\ \forall\ t = 1,2$. Under the assumption that u_{it} is normally distributed, we can estimate equation (4.4) using a random effect probit model.

Estimation results

<Table 4.4 approximately here>

Columns 1 to 4 in Table 4.4 presents the estimation results of the random effect probit model defined in equation (4.5). We start assuming the absence of any starting point bias (column 1) and alternatively include shift effect (column 2) and anchoring effect (column 3) in the model. Finally, to ensure whether the shift (anchoring) effect is not capturing any other effect inappropriately, the simultaneous presence of both the effects is also considered in the estimation (column 4).

We observe that the anchoring effect is present within the DBDC model, but its effect is marginal (p <0.1) (see columns 3 and 4 of Table 4.4). This suggests that the initial bid may have influenced the individuals' decision in the follow-up round. Estimation results from the model that potentially include both the sources of starting point bias, suggest that in absolute terms, the value of the effect is approximately 0.217 (see column 4). In other words, the individuals tend to refer 21.7% of their WTP to the initial bid while responding to the follow-up question. We also find that the coefficient of the shift effect is not statistically significant at 10% level. In other words, there is little evidence that shift effect is the source of starting point bias in our model. In addition, we find that the estimated coefficients of the other covariates in Table 4.4 are robust and lie in a more or less in a comparable range with those in the restricted DBDC model presented in columns 4 to 6 in Table 4.3.

4.3.3. Mean WTP estimates from the DBDC model & related discussions

Jeanty et al. (2007) have argued that it is preferable to use the restricted models for deducing inference on the WTP estimates given that the data supports the restrictions from a statistical standpoint. As discussed in sub-section 4.3.1, we cannot reject the null hypothesis that the restrictions are valid at 10% levels for models 1, 2 and 3 in Table 4.3 and hence it is better to use the restricted models for estimating the WTP. Therefore, we use the restricted model – model 6 of Table 4.3 to estimate our mean WTP as that has the highest log likelihood value among the three restricted models and also, contains the full set of covariates.

Given the log-normal specification of the model, the estimated mean WTP can be computed as $\hat{E}(w^*) = \exp\left(\frac{\bar{X}'\hat{\beta}}{\hat{\theta}} + 0.5\hat{\sigma}^2\right)$ where, $\hat{\beta}, \hat{\theta}$ are the estimated parameters of covariates and bids respectively; $\bar{X}$ is the average of other covariates based on the data (Du and Mendelsohn, 2011). Apart from providing the point estimates of mean WTP, we also report its confidence intervals using the Monte Carlo simulation method developed by Krinsky and Robb (1986)[12] by employing the parametric bootstrap procedure.

We estimate the mean annual WTP from the restricted DBDC model to be INR 886.59 with the 95% confidence interval between INR 697.44 – 1310.93[13]. This estimated mean annual WTP accounts for approximately 1.06% of annual household expenditure of the sample households. For a better understanding, we present the distribution of the estimated mean

[12] The point estimates along with the 95% confidence intervals by Krinsky- Robb method of mean WTP may be computed using the user-written command *wtpcikr* (10000 replications) in statistical software STATA (Jeanty, 2007).

[13] As a cursory check, we obtain the estimated mean WTP from the unrestricted DBDC model as INR 731.68 (95% CI between: INR 589.08 and INR 1012.93) using the estimated coefficients of column 3 of Table 4.3. It is lower than the estimated mean WTP obtained from the restricted model as indicated in Jeanty et al. (2007). Furthermore, we compare the estimated mean WTP obtained from unrestricted DBDC model to that in SBDC model (using coefficients of column 3 in Table 4A.2). As indicated in literature (e.g.: Gelo and Koch, 2015; Donfuet et al., 2014), the estimated mean WTP from the unrestricted DBDC model is significantly lower than that in the SBDC model (INR 734.91; CI: 589.25 – 1025.13). We summarise these findings in the appendix (Table 4A.3)

annual WTP in Figure 4.2[14]. A visual inspection of Figure 4.2 suggests that the distribution of the estimated mean WTP is positively skewed one; majority of the respondents express a lower WTP for the improved indoor quality from HAP reduction. This suggests that respondents in our sample perceive a low valuation of the improved air quality from HAP reduction. One plausible reason behind this may be that respondents in our sample may not perceive the risk from HAP to be a serious one from their prior experience and regular exposure of it owing to habitual and age-over practices over generations[15].

<Figure 4.2 approximately here>

As a comparison[16], Shannon et al. (2019) have estimated the mean monthly WTP for preventive device for HAP reduction to be in the range of USD 1.09 - 1.68[17]. This accounts for approximately 1 – 2% of the monthly household income in their stated preference study of rural Indian households. It is expected that different samples, survey time as well as CV scenario may affect the different mean values of WTP but the ratio of WTP to household income is expected to lie in a comparable range (Sun et al., 2016). Although, we have conducted the study in a different part of India and the monthly household expenditure in our sample was marginally more (~$120)[18] compared to that in Shannon et al. (2019), the ratio of mean WTP to income of both the studies lies in a moderately relative range. The findings from these stated preference studies may suggest that individuals in rural India have consistently

---

[14] We have generated the 10000 bootstrapped estimates of the mean annual WTP for the over-all sample using Monte Carlo simulation method based on the estimated coefficients in model 6 in Table 4.3.

[15] Our findings conform with previous studies. For example, Mobarak et al. (2012) have observed similar findings in rural Bangladesh where female members of the household consider the health risks from HAP to be less serious than other common health hazards arising from contaminated water or food poisoning.

[16] To compare our estimated valuation for reduced health risks from HAP with that in existing literatures, we express the estimated mean WTP in *monthly* terms. Based on our analysis, the estimated mean WTP per month is approximately INR (886.59/12)=73.88 (USD 1.17) which is, on an average, around 1% of the respondents' monthly household income.

[17] Shannon et al. (2019) have implemented the CV study in 2013 in rural Rajasthan (located in western India and have different geographical and topographical features compared to our survey site) where the average income of the sample households was USD 100.

[18] We have implemented the CV survey in 2017-2018. During the period of 2013-2017, India has witnessed a growth in GDP from 6.39 in 2013 to 7.17 in 2017 (World Bank, 2019). As a result of this economic development, the average income of the people of rural India is expected to increase as may have been reflected here.

attributed a low valuation to the reduction in HAP-related health risks which is robust across time and study area.

It is not unreasonable to assume that the estimated WTP may appear to be negligible in nominal terms. However, McPhail (1993) has suggested that that improvement in access and/or provision of basic amenities like piped water in developing countries may be considered to be affordable if it is approximately 5% of the household income. Drawing reference to this frequently quoted measure of affordability for basic amenities in developing countries or the 'five-percent rule' (Shannon et al., 2019) we may conclude that, although the estimated mean WTP appears to be negligible in nominal terms, it may be non-trivial in the context of low-income economies particularly in rural areas.

It will be interesting to compare the individuals' estimated WTP to their actual expenditure for cooking fuels. On average, the households in our sample reported paying INR 5267.52 (~ USD 85) on average to meet their annual cooking fuel consumption demand that accounts for approximately 6.33% of their average annual household income. The estimated mean WTP in our study (~ USD 14.30; 1.06% of average annual household income) is sufficiently lower than the actual expenditure on cooking fuels in our sample. Though this finding may appear surprising, it is consistent with literatures on valuation of environmental goods like access to safe drinking water in developing countries (e.g.: Beaumais et al., 2014; Wang et al., 2010) where the WTP is lower than the actual expenditure the households are incurring.

Literatures on economic valuation of any risk reducing goods and/or services often suggest that estimated WTP of the individuals is often lower than the actual fixed cost they have to pay for adopting a new technology to mitigate risk, particularly in low-income countries (e.g.: Mugisha et al., 2012; Scarpa and Willis, 2010) . For example, Kabyanga et al. (2018) have observed that households' maximum WTP for flexible biogas ballon digester unit is approximately 10 times lower than the actual investment the respondents have to make in

Uganda. Our study for hypothetical preventive measure intended for HAP reduction also suggests similar trend. As discussed earlier in Chapter 1, our sample households have to incur a fixed cost of INR 5000 (~ USD 82; ~6.01% of average annual household income) to switch to LPG that will reduce the HAP, which is sufficiently higher than our estimated annual WTP (~ USD 14.30; 1.06% of average annual household income).

Summarising the above findings, we may conclude that our sample households are willing to pay around 1% of their average annual household expenditure to reduce HAP-related health risks and it is non-negligible in the context of a low-income economy. Furthermore, in conformation with literature on valuation of other environmental or risk-reducing goods, our results suggest that the estimated WTP is lower than the actual fuel expenditure as well as, the initial investment to switch to cleaner cooking fuels. The following two reasons can be attributed to substantiate the latter findings. First, in comparison to average rural Indian households, our sample households are relatively low-income ones[19]. Second, the households are unsure about how to prioritize the reductions in HAP-related health risks and hence, may not be willing to pay for reducing HAP beyond a certain threshold value.

4.4. Heterogeneity of the estimated mean annual WTP and its policy implications

Individuals' WTP may be commodity specific and also depends on space and time (Sun et al., 2016). Since for a given commodity, the individuals are quite likely to develop their valuation based on their individual specific characteristics, the estimated mean WTP is expected to be sensitive to these individual-specific attributes. This gives us a rationale to explore the within-sample heterogeneity of estimated mean WTP, given our evidence that some of the relevant and indicative covariates influence the individuals' WTP decision.

[19] During 2017-18, the average annual income of a rural Indian household was INR 8059 (NABARD, 2018)

We focus on three contextual attributes related to individuals' WTP for mitigating health risks from HAP exposure – self-reported health status, cooking fuel usage pattern and perception of increased health risk from dirty fuel usage. Unlike the first variable, the latter two is continuous in the interval [0,1]. For this reason, for the latter two variables, we focus on the estimated mean annual WTP evaluated at the median and that at the two endpoints.

For this analysis, we use the estimated coefficients of the restricted bivariate probit model results presented in column 6 of Table 4.3 for reasons mentioned before. We also present the kernel density of the bootstrapped estimates of the mean across the variation in covariates in Figure 4.3. The possible heterogeneity in the estimated mean within the sample, along with the 95% confidence intervals under various scenarios is presented in Table 4.5.

<Figure 4.3 approximately here>

Panel I of Figure 4.3 presents the heterogeneity in estimated mean with respect to self-reported health status. Although the distribution of the two groups *sick* and *not sick* is positively skewed, they somewhat differ in shape in terms of spread. The estimated mean of the latter group (INR 514.38) is much lower than that of the former (INR 1053.47) (see Table 4.5); the *sick* individuals, on an average are willing to pay a higher annual premium[20] for the preventive device than their *not sick* counterparts, with a larger confidence interval.

The within-sample variation in the estimated mean with respect to cooking fuel usage is presented in Panel II of Figure 4.3. The kernel densities of the quantities at all the levels of

[20] Incidentally, it is interesting to compare the estimated mean WTP of the *sick* individuals with their 'actual' expenditure to reduce the most recent event of the above-mentioned physical symptoms, which can be considered as their defensive expenditure. In monthly terms, the estimated mean WTP for the sick individuals is INR 87.79 while, from the data it is obtained that, they spend INR 57.25 on average per month to treat their recent events of sickness from aforementioned symptoms. Our result conforms with the economic theory that individuals WTP is likely to exceed their defensive expenditure (e.g.: Atreya et al., 2011; Albeini and Krupnick, 2000); individuals tend to include the defensive expenditure as well as intangible costs such as suffering from the symptoms while bidding for the preventive measures. However, the magnitude of the ratio between defensive expenditure to the elicited WTP (1.53), although more than unity, is slightly low in our study compared to Alberini and Krupnick (2000). This may be due to economic, cultural, and institutional differences between India and other countries and also, our emphasis on few common but minor physical symptoms associated with HAP.

fraction of days of dirty fuel usage is positively skewed. Although there is some overlap among the densities, it is evident from the figure that exclusive clean (dirty) cooking fuel users have the highest (lowest) WTP for the preventive measure despite possibly requiring it the least (most). To be specific, the estimated mean annual WTP of the former group (INR 1468.38) exceeds that of the latter group (INR 696.60) by around two times. Panel II further indicates that, the exclusive dirty fuel users consistently have the lowest perceived private health benefit from reduced health risk related to HAP which results to the lowest spread among the three groups.

Given the evidence that individuals' self-reported health status and cooking fuel usage may result in within-sample variation in estimated mean, we attempt the investigate the joint impact of these variables on the heterogeneity of the quantity. For this purpose, we classify the categories based on the two groups of self-reported health status (*sick* and *not sick*) and that of fuel users (*exclusive clean fuel users* and *exclusive dirty fuel users*[21]). We plot the distributions of the bootstrapped estimates of the quantity corresponding to these four groups in panel III of Figure 4.2. This figure shows that the group *exclusive clean fuel users and sick* has the highest valuation of reduced health risks related to HAP among the four groups, while an exactly opposite pattern is observed for the group *exclusive dirty fuel users and not sick*. The individuals in the group *exclusive dirty fuel users and not sick* seem to consistently express the lowest perceived private health benefits from reduction in health risks related to HAP, resulting in the lowest dispersion in the distribution. Panel III further reveals that the distribution of the remaining two groups is more or less overlapping, indicating similar perception of private health benefits from reduced health risks related to HAP.

[21] In the sample, around 12% (44%) households are exclusive clean (dirty) fuel users.

Finally, we present the within-sample heterogeneity of estimated mean with respect to different levels of perceptions of increased health risk from dirty fuel usage in Panel IV of Figure 4.3. The figure shows that the groups assigning maximum value of the perception is likely to have the highest valuation of the reduced health risks related to HAP. This is quite evident given our finding that perception positively influences the individuals' likelihood to pay the bid as discussed in previous sub-section. To be specific, the individuals assigning the maximum value of the perceived increase in health risk has an estimated annual mean WTP as INR 841.22 which is approximately twice that of the group expressing minimum value to the perceived risk.

<Table 4.5 approximately here>

Summarizing the observations of the within-sample heterogeneity of the estimated mean WTP, we may conclude that it is sensitive to both health and non-health factors. The respondents in the group expressing the highest WTP, have a larger spread on their perceived private health benefit from reduced health risks related to HAP. Concurrently, respondents in other groups, who are more conservative in nature with lower WTP, hold a more consistent view.

This finding may appear to be in contrast to the findings of previous literature (Sun and Zhu,2014) in the context of WTP for avoidance of nuclear power. However, it is to be noted that the risk perception or threat related to nuclear power is enormous in magnitude and seldom encountered in reality. Concurrently, the individuals in developing countries particularly in rural areas perceive the risk from HAP to be a trivial one, not to speak of any threat whatsoever, because of their prior experience with prolonged and regular exposure of it owing to habitual and age-old practices over generations[22]. Furthermore, we have elicited the bid responses by

[22] We do not disagree with the findings of Sun and Zhu (2014). Rather, this study may indicate that lower perceived private health benefit is likely to prevail over familiar facilities such as clean air and water which are regularly

mentioning about common symptoms related to HAP, of which they are often ignorant in the short run[23]. Therefore, it will not be unreasonable to argue that individuals belonging to the group expressing higher WTP may have their perceived private health benefit deviated further from each other in the context of reduced health risks from HAP. Thus, some individuals of the group have expressed a higher valuation of reduced health risks than others. This may result in the heterogeneity of the perception of private health benefit within the groups, resulting in the larger spread.

Above evidence leads to at least two important directions for policy design. First, since the respondents' perception of increased health risks from dirty fuel usage has a significant positive effect on their WTP for reduced health risks related to HAP, a plausible policy may be suggested as follows. Government at the local level may launch programs to educate the people about the possible health hazards (both short term and long term) related to HAP as well as, the urgency to adopt/use modern cooking fuels (technology). It is likely that with an increased knowledge, the individuals' concerns about HAP related issues will increase. This is expected to enhance the chance of success of the current or future intervention programs[24].

Second, following Yokoo et al (2020), the results may enable one to improve the efficiency of awareness generation and/or information provision programs by addressing the target group

---

used. The observation by Takama et al (2012) further substantiate this finding; low-income households in Ethiopia have a higher WTP to reduce the risk associated with explosion of cook stoves which is an unfamiliar and rare kind of health hazard in comparison to the much familiar hazard of burning. Thus, the perceived private health benefit is likely to differ and expected to be higher for facilities which are rather uncommon in everyday use like hazards from nuclear power station or toxic chemical reactor. As a result, framing of policies for facilities which are in everyday use with a relatively low perception of hazard, although are more important and widespread, seems to be difficult and may take considerable amount of time and persuasion.

[23] In general, it is difficult to elicit the risk perception or valuation of risks for serious physical symptoms related to HAP like COPD or ALRI, as they occur less frequently (Yokoo et al., 2020).

[24] It is to be noted that generation of public awareness particularly in the context of interventions targeted to HAP reduction is of much relevance for ensuring sustained success of such programs. For example, despite the initiative by the Government of India to provide free LPG connections to households lying below poverty line, a large section of the target households continues to use dirty cooking fuels either as their primary or as the secondary sources of cooking fuel (Gould and Urpelainen, 2018). This may reflect the necessity of a concurrent awareness generation about the health risks related to HAP along with the intervention policies to reduce such HAP.

identified by observables. For example, the individuals belonging to the group *exclusive dirty fuel users and not sick* show a consistent but more conservative attitude towards HAP problems with the lowest valuation for reduction in health risks related to HAP. Any environmental policy in general and awareness generation and/or information provision policy towards reduction of HAP in particular, may be found to be more effective if such an understanding can be targeted and altered.

4.5. Conclusion

Health risk related to HAP is a salient feature of the households in developing countries particularly in the rural areas. We analyze data from a unique contingent valuation survey in rural India to estimate the individuals' valuation of reduced health risks derived from hypothetical improvement in indoor air quality using stated preference method. In particular, we estimate their WTP for reduction in HAP related health risks using DBDC approach. As a minor extension, we also attempt to investigate the presence and potential source of starting point bias in our DBDC model.

The potential impact of our results on the literature related to the valuation of environmental health risks related to HAP in the context of a developing economy, is worth noting. We estimate the mean annual WTP for mitigating the health risks related to HAP to be INR 886.59 (~ USD 14.30) which is approximately 1.06% of the annual household income. Although we have conducted the study in a different time and with a different sample, the ratio of our estimated WTP to average household expenditure is in a comparable range with previous literature. In addition, our results suggest the presence of anchoring effect thereby validating the presence of starting point bias in our DBDC model. Furthermore, sufficient within-sample heterogeneity of the estimated mean annual WTP with respect to judiciously selected covariates is also observed. This enables us to recommend policies like generating public awareness about HAP risk and targeting potential beneficiaries based on observable characteristics. Such policy is expected to ensure smooth implementation and enables one to assess the effectiveness of intervention programs to reduce HAP.

Our results should be interpreted with caution. First, our analysis of perceived private health benefits focuses on common but minor physical symptoms related to HAP which is often ignored by individuals in the short run. Consequently, the estimated mean WTP may yield an appropriate lower (upper) bound of the valuation. Second, it is to be noted that we have concentrated exclusively on the health risks related to HAP in this study. However, simultaneous presence of multiple health risks are rules rather than exceptions, particularly in developing economies. and this may affect the individuals' WTP for a specific risk. More detailed analyses on multiple health risks (for example, health risks related to HAP coupled with health risks from improper sanitization for women or consumption of contaminated water) are required, including relevant sensitivity analysis. Third, the results may not be generalizable for the entire rural India because they are not necessarily an unbiased representation of the population (that is, all the individuals in rural India). However, there may be several areas in West Bengal similar to our study area in terms of ethno-socio-demographic features, which are located in the proximity of an urban metropolis. As such, it is expected that the findings here will be valid for those areas.

We have confined our attention to the perceived health benefit from the viewpoint of the respondent, and this may be an oversimplification. For a holistic analysis of the individuals' valuation of reduced health risks related to HAP, we need to include the household burden of diseases, especially that of the kids. Finally, for a comprehensive analysis of individuals' demand for the preventive device, we need to take up future studies that will also identify which attributes of this preventive measure is given priority by the potential beneficiaries apart from estimating the valuation. This may demand the necessity of a stated preference study. We would like to extend our research in these directions in future.

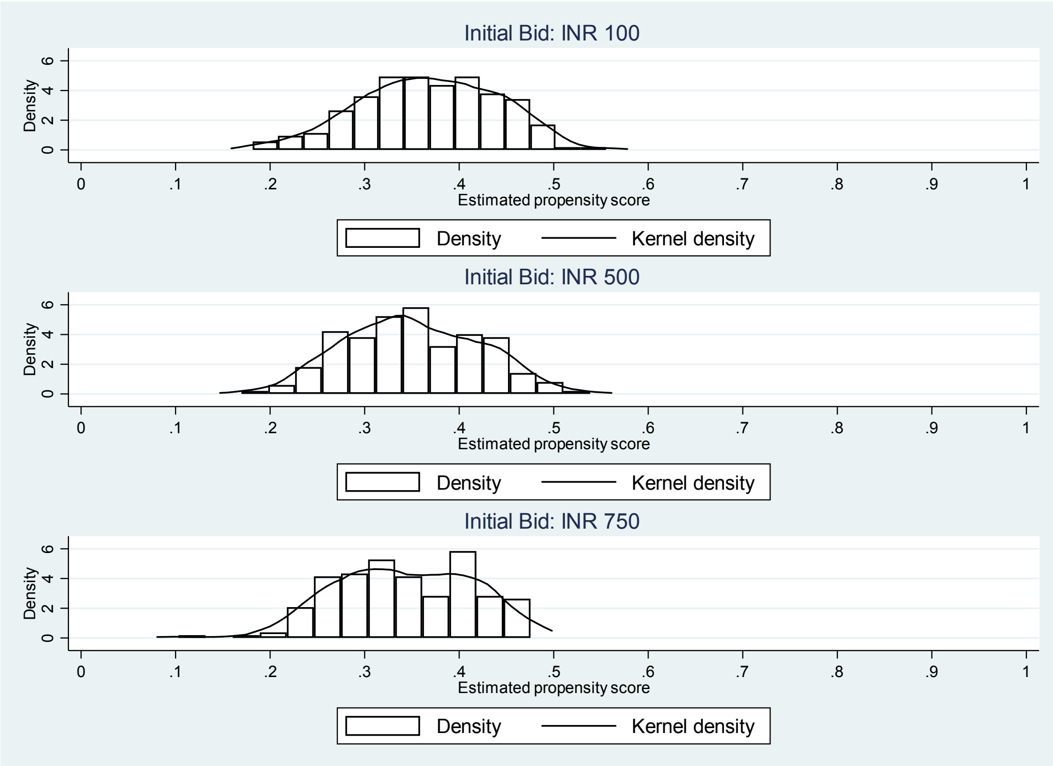

Figure 4.1. Histogram and Kernel density of the estimated propensity scores across three bid groups

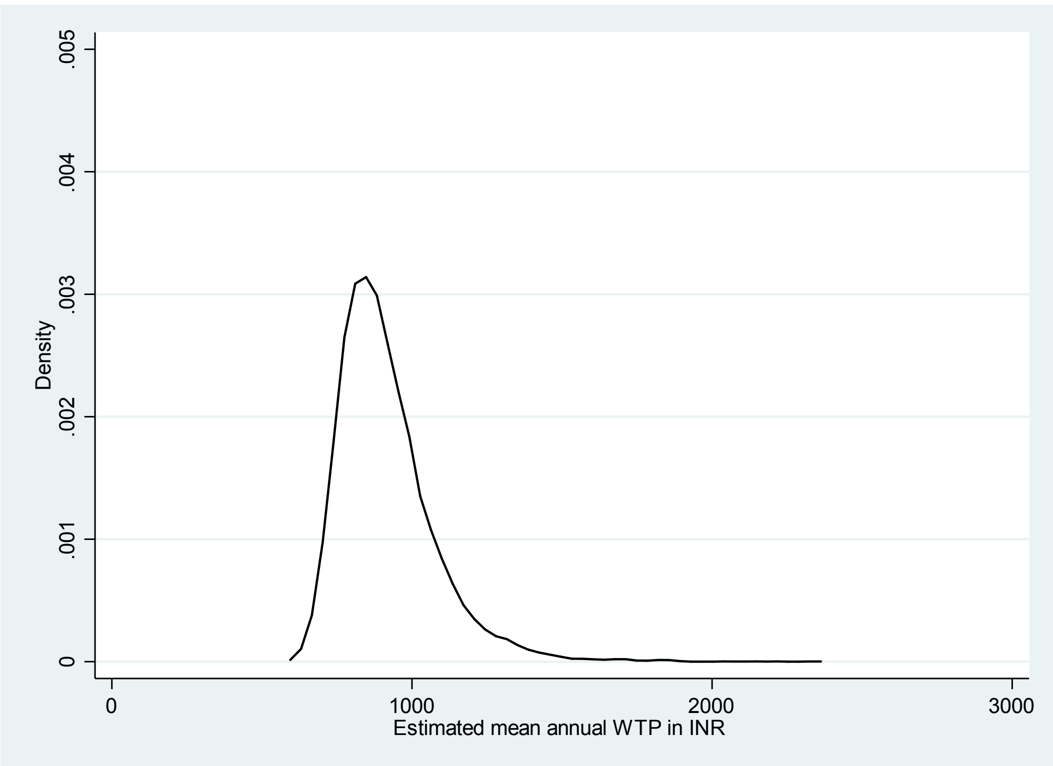

Figure 4.2 Distribution of the estimated mean annual WTP in INR

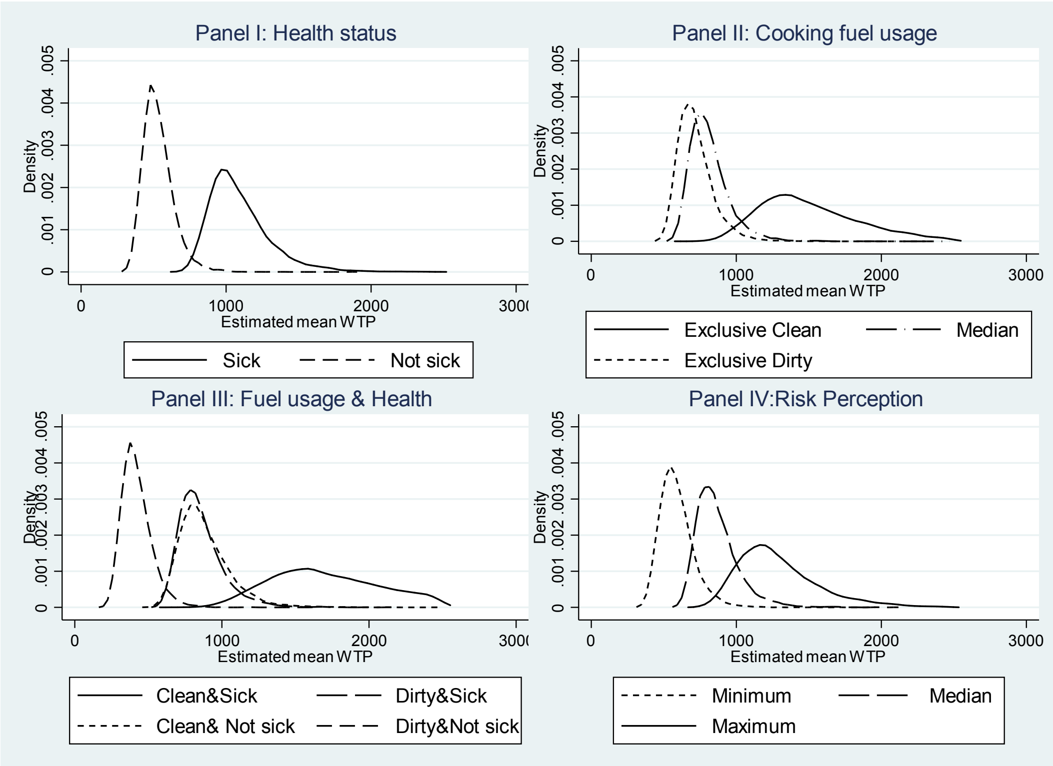

Figure 4.3. Kernel density of the bootstrapped estimates of mean WTP

Table 4.1. Descriptive statistics

Variable	Mean	SD	Min	Max
Health-related variable				
Sick in last 30 days with at least one physical symptom (binary)	0.76	0.43	0	1
Cooking practice-related variable				
Fraction of days of dirty fuel usage in last 30 days	0.68	0.38	0	1
Risk perception-related variable				
$SP(s\|d) - SP(s\|c)$	0.57	0.22	0.14	1
Other 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
Holds household decision-making authority (binary)	0.06	0.24	0	1
Spouse works in informal sector (binary)	0.3	0.46	0	1
Spouse works in agricultural sector (binary)	0.27	0.44	0	1
Expenditure (in INR 1,000)	7.51	3.74	2.3	55
Kitchen located inside dwelling area (binary)	0.16	0.36	0	1
Access to ventilation in cooking area (binary)	0.97	0.16	0	1
Owns television (binary)	0.86	0.35	0	1

Note: In risk perception-related variables, $SP(.|.)$, s denotes the likelihood of being sick from at least one of the physical symptoms (dry cough, sore or runny eyes and difficulties in breathing) and c(d) represents clean(dirty) cooking fuel usage. The sample size in 557.

Table 4.2: Distribution of the bid responses

Lower Follow-up Bid	Initial Bid	Higher Follow-up Bid	Yes-Yes	Yes-No	No-Yes	No-No	*N*
50	100	200	68.53%	17.77%	7.11%	6.60%	197
250	500	1000	10.29%	29.14%	38.86%	21.71%	175
375	750	1500	8.65%	20.00%	37.30%	34.05%	185

Note: This table presents the distribution of the bid responses across the respondents. Each value of the bid is expressed in Indian National Rupee (INR) where INR 62= USD 1 (the average monthly exchange rate in December 2017- January 2018 when the survey was conducted). N represents the number of respondents who were assigned that level of bid random. The total sample size is 557.

Table 4.3. Results of the DBDC model

	Unrestricted model			Restricted model		
	(1)	(2)	(3)	(4)	(5)	(6)
Bid(log)	-0.860***	-0.866***	-0.866***	-0.804***	-0.805***	-0.800***
	(0.074)	(0.074)	(0.073)	(0.072)	(0.072)	(0.071)
Number of cooks	-0.073	-0.066	-0.066	-0.061	-0.053	-0.051
	(0.141)	(0.142)	(0.143)	(0.121)	(0.121)	(0.117)
Age	-0.012*	-0.013**	-0.013**	-0.011**	-0.012**	-0.011**
	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)
Years of schooling	-0.001	-0.002	-0.002	-0.004	-0.005	-0.005
	(0.017)	(0.017)	(0.017)	(0.013)	(0.013)	(0.013)
Decision-maker	0.529**	0.509*	0.505*	0.494**	0.466**	0.456**
	(0.269)	(0.272)	(0.272)	(0.199)	(0.201)	(0.2)
Spouse works in informal sector	0.025	0.082	0.079	0.036	0.087	0.084
	(0.149)	(0.151)	(0.152)	(0.112)	(0.113)	(0.113)
Spouse works in agricultural sector	-0.219	-0.188	-0.192	-0.111	-0.076	-0.084
	(0.151)	(0.152)	(0.152)	(0.114)	(0.113)	(0.113)
Kitchen located inside	-0.118	-0.120	-0.125	-0.027	-0.025	-0.0413
	(0.159)	(0.160)	(0.160)	(0.125)	(0.124)	(0.122)
Ventilation	-0.109	-0.243	-0.246	-0.003	-0.124	-0.130
	(0.328)	(0.343)	(0.338)	(0.283)	(0.284)	(0.269)
Owns television	0.044	0.035	0.026	-0.008	-0.021	-0.049
	(0.184)	(0.186)	(0.186)	(0.131)	(0.132)	(0.131)
Household expenditure	0.041*	0.04*	0.04*	0.03*	0.029*	0.029*
	(0.022)	(0.022)	(0.022)	(0.016)	(0.016)	(0.016)
Fraction of days of dirty fuel usage	-0.0327	-0.538**	-0.530**	-0.107	-0.622***	-0.596***
	(0.178)	(0.257)	(0.257)	(0.124)	(0.184)	(0.182)
Sick		0.601***	0.59***		0.605***	0.573***
		(0.215)	(0.214)		(0.156)	(0.154)
$SP(s\|d) - SP(s\|c)$			0.217			0.623***
			(0.27)			(0.19)
ρ	0.298**	0.258**	0.237*	0.395***	0.358***	0.322***
	(0.121)	(0.121)	(0.122)	(0.112)	(0.110)	(0.108)
Log likelihood	-647.7	-639.8	-632.5	-651.8	-644.0	-639.4
χ^2	169.2	185.7	202.4	152.3	165.5	174.9
χ^2 test for ρ	6.063	4.591	3.773	12.55	10.66	8.974

Note: This table provides the estimation results for the DBDC model presented in equation (4.1), where the dependent variable is the bid response (=1 if the respondent is willing to pay the proposed bid) for the initial round. Columns 1 to 3 presents the estimation result of the unrestricted model while columns 4 to 6 presents the estimation results of the restricted model. The sample size is 557. ***, ** and * indicate statistical significance at the one, five and ten per cent levels respectively. Robust standard errors are presented in parenthesis. The constant terms are not reported for the sake of space.

Table 4.4. Result of model addressing starting point bias

	(1)	(2)	(3)	(4)
Bid (log)	-0.964***	-0.958***	-0.959***	-0.972***
	(0.131)	(0.13)	(0.128)	(0.118)
Shift effect parameter		0.14		-1.143
		(0.089)		(0.718)
Anchoring effect parameter			0.027*	0.217*
			(0.015)	(0.121)
Number of cooks	-0.061	-0.062	-0.061	-0.059
	(0.133)	(0.133)	(0.132)	(0.122)
Age	-0.013**	-0.013**	-0.013**	-0.012**
	(0.005)	(0.005)	(0.005)	(0.005)
Years of schooling	-0.006	-0.006	-0.006	-0.005
	(0.015)	(0.015)	(0.015)	(0.014)
Decision-maker	0.549**	0.551**	0.545**	0.501**
	(0.237)	(0.238)	(0.235)	(0.219)
Spouse works in informal sector	0.101	0.100	0.1	0.097
	(0.132)	(0.132)	(0.131)	(0.122)
Spouse works in agricultural sector	-0.101	-0.104	-0.103	-0.089
	(0.139)	(0.139)	(0.137)	(0.128)
Household expenditure	0.035*	0.035*	0.034*	0.032*
	(0.018)	(0.018)	(0.018)	(0.017)
Kitchen located inside	-0.05	-0.052	-0.052	-0.046
	(0.149)	(0.149)	(0.148)	(0.137)
Ventilation	-0.157	-0.151	-0.148	-0.134
	(0.337)	(0.338)	(0.334)	(0.310)
Owns television	-0.06	-0.064	-0.064	-0.06
	(0.157)	(0.158)	(0.156)	(0.145)
Fraction of days of dirty fuel usage	-0.719***	-0.718***	-0.712***	-0.665***
	(0.234)	(0.234)	(0.232)	(0.216)
Sick	0.691***	0.690***	0.684***	0.636***
	(0.193)	(0.193)	(0.191)	(0.178)
$SP(s\|d) - SP(s\|c)$	0.751***	0.755***	0.750***	0.712***
	(0.250)	(0.251)	(0.248)	(0.231)
$ln\sigma^2$	-0.794*	-0.790*	-0.842*	-1.289*
	(0.475)	(0.471)	(0.483)	(0.696)
Log likelihood	-639.4	-638.1	-637.8	-636.6
χ^2	64.04	66.32	68.49	83.26

Note: This table provides the estimation results for equation (4.5), where the dependent variable is the bid response (=1 if the respondent is willing to pay the proposed bid). The sample size is 557. ***, ** and * indicate statistical significance at the one, five and ten per cent levels respectively. Robust standard errors are presented in parenthesis. The constant terms are not reported for the sake of space.

Table 4.5. Within-sample heterogeneity analysis of estimated mean WTP

	Estimated Mean	95% Confidence Interval Lower Bound	Upper Bound
Across categories of health status			
Sick	1053.47	798.00	1631.01
Not sick	514.38	373.23	779.68
Across different levels of fuel usage			
Exclusive clean fuel usage	1468.38	978.14	2595.33
Median	788.99	622.28	1155.73
Exclusive dirty fuel usage	696.6	539.48	1021.03
Across different levels of fuel usage & health status			
Exclusive clean fuel user & sick	1745.30	1084.07	3314.65
Exclusive dirty fuel user & sick	827.98	638.71	1239.35
Exclusive clean fuel user & not sick	852.17	633.48	1305.32
Exclusive dirty fuel user & not sick	404.28	265.96	662.88
Across different levels of perceptions			
Minimum value	447.53	333.86	703.46
Median value	569.71	411.79	866.29
Maximum value	841.22	663.88	1235.37

Note: The mean WTP per year across different categories of individual-specific factors is estimated using the results of the restricted DBDC model presented in column (6) of Table 4.3. The estimated mean WTP per year is expressed in INR where USD 1=INR 62 (average of the average monthly exchange rate in December 2017 and January 2018). The confidence intervals of corresponding mean WTPs are computed using Krinsky and Robb method (number of replications 10000).

Appendix

Appendix 4.A1

Instructions and question to elicit the bid responses

<table>
<tr><th colspan="5">Willingness to Pay</th></tr>
<tr><td colspan="5">We understand that the issue of smoke coming from the burning of cooking fuels while you cook is quite serious for the household health, especially yours. We request you to think of a situation where some public program has been implemented for the public interest in all the villages under this village council. In this program, some kind of preventive device similar to an electric chimney or exhaust fan is installed in the cooking area of the house at a minimal cost or for free. The expected benefit from the preventive device is the following: the incidence and extent of the smoke during cooking will be greatly reduced. This, in turn, will effectively reduce your chances of suffering from related physical symptoms like dry cough, sore or runny eyes, difficulties in breathing. Moreover, this will also improve the indoor air quality of the household such that your kids or other family members will also have a lesser chance of suffering from the above-mentioned diseases. However, please note that the program fund will not be sufficient to finance the usage or maintenance cost for the device. Once installed you need to pay [initial bid] per year for using the device. We shall now request you to kindly consider your household budget constraint and other financial obligations before answering this question</td></tr>
<tr><td colspan="5">To the enumerator: Please ask the <u>value according to the group id</u> randomly assigned to you</td></tr>
<tr><td rowspan="4">I</td><td>Group ID</td><td>A</td><td>B</td><td>C</td></tr>
<tr><td rowspan="3">Are you willing to pay this amount per year for the preventive device?</td><td>Rs. 500</td><td>Rs. 750</td><td>Rs. 100</td></tr>
<tr><td>Yes</td><td>Yes</td><td>Yes</td></tr>
<tr><td>No</td><td>No</td><td>No</td></tr>
<tr><td colspan="5">To the enumerator: If YES then <u>proceed to ii</u>; if NO then <u>proceed to iii</u></td></tr>
<tr><td rowspan="4">Ii</td><td>Group ID</td><td>A</td><td>B</td><td>C</td></tr>
<tr><td rowspan="3">Are you willing to pay this amount per year for the preventive device?</td><td>Rs.1000</td><td>Rs. 1500</td><td>Rs. 200</td></tr>
<tr><td>Yes</td><td>Yes</td><td>Yes</td></tr>
<tr><td>No</td><td>No</td><td>No</td></tr>
<tr><td rowspan="4">Iii</td><td>Group ID</td><td>A</td><td>B</td><td>C</td></tr>
<tr><td rowspan="3">Are you willing to pay this amount per year for the preventive device?</td><td>Rs.250</td><td>Rs. 375</td><td>Rs. 50</td></tr>
<tr><td>Yes</td><td>Yes</td><td>Yes</td></tr>
<tr><td>No</td><td>No</td><td>No</td></tr>
</table>

Appendix 4.A2

Table 4.A1. Results of the unrestricted DBDC model (both rounds)

	Initial round response			Follow-up round response		
	[1a]	[2a]	[3a]	[1b]	[2b]	[3b]
Initial Bid(log)	-0.86***	-0.866***	-0.866***			
	(0.074)	(0.074)	(0.073)			
Final Bid (log)				-0.676***	-0.672***	-0.676***
				(0.104)	(0.106)	(0.108)
Number of cooks	-0.073	-0.066	-0.066	-0.064	-0.056	-0.049
	(0.141)	(0.142)	(0.143)	(0.148)	(0.148)	(0.144)
Age	-0.012*	-0.013**	-0.013**	-0.009	-0.011*	-0.009
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Years of schooling	-0.001	-0.002	-0.002	-0.007	-0.008	-0.008
	(0.017)	(0.017)	(0.017)	(0.016)	(0.016)	(0.016)
Decision-maker	0.529**	0.509*	0.505*	0.438*	0.403	0.403
	(0.269)	(0.272)	(0.272)	(0.243)	(0.248)	(0.248)
Spouse works in informal sector	0.025	0.082	0.08	0.053	0.01	0.099
	(0.149)	(0.151)	(0.152)	(0.138)	(0.139)	(0.14)
Spouse works in agricultural sector	-0.219	-0.188	-0.192	-0.015	0.0201	0.012
	(0.151)	(0.152)	(0.152)	(0.143)	(0.144)	(0.146)
Household expenditure	0.041*	0.04*	0.04*	0.021	0.02	0.019
	(0.022)	(0.022)	(0.022)	(0.019)	(0.019)	(0.019)
Kitchen located inside	-0.118	-0.12	-0.125	0.046	0.05	0.026
	(0.159)	(0.16)	(0.160)	(0.154)	(0.154)	(0.151)
Ventilation	-0.109	-0.243	-0.246	0.102	-0.014	-0.033
	(0.328)	(0.343)	(0.338)	(0.357)	(0.356)	(0.343)
Owns television	0.044	0.035	0.026	-0.063	-0.08	-0.131
	(0.184)	(0.186)	(0.186)	(0.163)	(0.165)	(0.164)
Fraction of days of dirty fuel usage	-0.033	-0.538**	-0.53**	-0.168	-0.676***	-0.647***
	(0.178)	(0.257)	(0.257)	(0.160)	(0.233)	(0.233)
Sick		0.601***	0.590***		0.596***	0.554***
		(0.215)	(0.214)		(0.191)	(0.191)
$SP(s\|d) - SP(s\|c)$			0.217			1.001***
			(0.270)			(0.259)
ρ	0.298**	0.258**	0.237*			
	(0.121)	(0.121)	(0.122)			
Log likelihood	-647.7	-639.8	-632.5			
χ^2	169.2	185.7	202.4			
χ^2 for ρ	6.063	4.591	3.773			

Note: This table provides the estimation results for the unrestricted bivariate probit model defined in equations (4.1) and (4.2) l, where the dependent variable is the bid response (=1 if the respondent is willing to pay the proposed bid) for the initial round (in columns indicated by "a") and follow-up round (in columns indicated by "b"). The sample size is 557. ***, ** and * indicate statistical significance at the one, five and ten per cent levels respectively. Robust standard errors are presented in parenthesis. The constant terms are not reported for the sake of space.

Appendix 4.A3

Table 4.A2. Results of the SBDC model

	[1]	[2]	[3]
Initial Bid (log)	-0.849***	-0.858***	-0.859***
	(0.073)	(0.073)	(0.073)
Number of cooks	-0.071	-0.064	-0.064
	(0.143)	(0.144)	(0.144)
Age	-0.012**	-0.014**	-0.013**
	(0.006)	(0.006)	(0.006)
Years of schooling	-0.001	-0.002	-0.001
	(0.017)	(0.017)	(0.017)
Decision-maker	0.543**	0.524*	0.517*
	(0.269)	(0.27)	(0.27)
Spouse works in informal sector	0.011	0.071	0.071
	(0.150)	(0.152)	(0.152)
Spouse works in agricultural sector	-0.222	-0.190	-0.193
	(0.151)	(0.152)	(0.152)
Household expenditure	0.041*	0.04*	0.04*
	(0.023)	(0.022)	(0.023)
Kitchen located inside	-0.109	-0.114	-0.119
	(0.157)	(0.159)	(0.158)
Ventilation	-0.111	-0.246	-0.247
	(0.327)	(0.341)	(0.337)
Owns television	0.07	0.056	0.044
	(0.185)	(0.185)	(0.186)
Fraction of days of dirty fuel usage	-0.029	-0.537**	-0.525**
	(0.178)	(0.255)	(0.254)
Sick		0.604***	0.589***
		(0.211)	(0.211)
$SP(s\|d) - SP(s\|c)$			0.238
			(0.273)
Log likelihood	-299.0	-294.5	-294.1
Pseudo R^2	0.224	0.236	0.237
χ^2	148.2	157.2	158.1

Note: This table provides the estimation results for (4.1) ignoring the second round response, where the dependent variable is the bid response (=1 if the respondent is willing to pay the proposed bid). The sample size is 557. ***, ** and * indicate statistical significance at the one, five and ten per cent levels respectively. Robust standard errors are presented in parenthesis. The constant terms are not reported for the sake of space.

Appendix 4.A4

Table 4.A3. Estimates and confidence intervals of mean WTP through various approaches

Entity	Estimate	95% Confidence Interval	
		Lower Bound	Upper Bound
$\mu_{DBDC,UR}$	731.68	589.08	1012.93
μ_{SBDC}	734.91	589.25	1025.13
$\mu_{SBDC} - \mu_{DBDC,UR}$	3.23 (0.033)		

Note: The annual mean WTP for the preventive device under unrestricted (UR) DBDC (SBDC) method is estimated using the results presented in column 3 of Table 4.3 (column 3 of Table 4.A4). The values are expressed in INR where USD 1=INR 62 (average monthly exchange rate during December 2017 - January 2018). The confidence intervals of corresponding mean WTPs are computed using Krinsky and Robb method (number of replications 10000). The standard error of the difference of mean WTP in SBDC and DBDC method is computed using bootstrap (number of replications 500).

Chapter 5. Conclusion

5.1. Summary of the dissertation

Household air pollution (HAP, hereafter) is a salient environmental and health risk particularly in rural areas of developing countries. Despite the severe health hazards associated with HAP, the use of dirty cooking fuels continues unabated in low-income economies, particularly in rural areas. In this background , this dissertation comprising of three empirical papers, focuses on the economics of HAP with an emphasis on the choice and usage of cooking fuels in rural India.

In the second chapter of the dissertation, we try to explore the role of access to information disseminated through internet on individuals' cooking fuel choice. Analysing a unique survey data of 565 rural Indian households, we find that access to internet has a negative and significant association with the likelihood to choose dirty cooking fuels. In particular, the propensity score matching approach suggests that households with access to internet are approximately 24 percentage points less likely to choose dirty cooking fuels compared to matched control group (households without access to internet).

In the third chapter of the dissertation thesis, we attempt to examine the individuals' subjective probabilistic expectations about health risks when using different types of fuel and their role in cooking fuel usage patterns. We also explore how these patterns, in turn, are associated with health status. Using data collected from 557 rural Indian households, we find that subjective probabilistic expectations of becoming sick from dirty fuel usage are negatively and significantly associated with the fraction of days of dirty fuel usage in households. Concurrently, dirty fuel usage and self-reported health status of the individual being sick are also significantly correlated. We then conduct a policy simulation of information provision regarding the health risks of dirty 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, these changes may be small in size.

In the fourth chapter of the dissertation, we attempt to assess the individuals' valuation of reduced health risk from HAP exclusively, derived from a hypothetical improvement in household air quality using stated preference method. In particular, we try to estimate the individuals' willingness to pay for reduction (WTP) in health risks related to HAP using a double bounded dichotomous choice (DBDC) approach. Concurrently, as an extension of estimating the individuals' WTP for reduction in HAP in this study, we attempt to explore the presence and source of starting point bias in our DBDC model. Using a unique contingent survey of 557 respondents in rural India, we estimated the mean annual WTP for the reduction in HAP to be INR 886.59 (~ USD 14.30) which accounts for approximately 1.06% of the annual household expenditure. Furthermore, our analysis suggests the presence of anchoring effect that validates the presence of starting point bias in our DBDC model. We also find that the estimated mean WTP is sensitive to several health and non-health factors.

The findings of the dissertation should be interpreted with caution as the results may not be generalizable for the entire rural India because they are not necessarily an unbiased representation of the population (that is, all the individuals in rural India). However, there may be several areas in West Bengal similar to our study area in terms of ethno-socio-demographic features, which are located in the proximity of an urban metropolis. As such, it is expected that the findings here will be valid for those areas.

In this dissertation, we attempt to present a holistic overview of the issues associated with mitigation of HAP as well as the mechanism behind choice making in the context of cooking fuel in rural India. Although the dissertation contains three empirical papers addressing different aspects of cooking fuel choice and/or usage, the three chapters are interlinked which we shall discuss now.

As mentioned earlier, one potential reason behind the continued unabated use of dirty cooking fuels particularly in rural areas of developing countries may be the knowledge gap. As a starting point of the thesis, we attempt to explore the role of information access disseminated through internet on cooking fuel choice of the individuals. Households with access to information are likely to have a higher perception of expectation of becoming susceptible to health hazards related to HAP and hence less likely to choose dirty cooking fuels. This may suggest that, if access to information has some role on cooking fuel choice and/or usage, subjective perception, or the subjective expectations regarding the health hazards from HAP may act as a mediator. This idea forms the basis of the third chapter of our dissertation. Specifically, in the third chapter, we attempt to segregate the impact of the individuals' SPEs of becoming sick from HAP related symptoms from that of internet access on the fuel usage pattern and focus on the former. Although in the third chapter we find that apart from access to internet is being significantly associated with cooking fuel usage pattern, individuals' subjective expectations have some direct influence on their fuel usage. Therefore, the third chapter of this dissertation may be considered to be an extension of the second chapter. Although in the third chapter we have successfully explored the role of individuals' SPEs on fuel usage pattern, there remains a limitation: the role of individuals' preferences on fuel usage pattern was not addressed. In the context of understanding the choice mechanism, it is necessary to explore the role of both preference and expectations as different combinations of expectations and preferences may lead to the same observed choice (e.g.: Manski, 2004). Furthermore, individuals' preferences may be influenced by individuals' subjective expectations however, information access and/or provision may have an ambiguous direct impact on preferences (Epstein and Peters, 2009). Thus, to have a comprehensive view on the issues of HAP mitigation and clean fuel adoption, it becomes necessary to study the role of individuals' preferences related to HAP reduction. We attempt to fill this gap by exploring the individuals' preferences for HAP mitigation in the fourth chapter of the dissertation. In particular, we try to understand how much the individuals are valuing the perceived private health benefit using a stated preference method. Although, we fail to find any evidence of significant direct impact of information access on individuals' valuation or WTP[1], we find that individuals' perceived risk from dirty fuel usage significantly affects their valuation.[2] Thus the fourth chapter is expected to complement the findings of the second and third chapters and help in better understanding of the choice mechanism behind HAP mitigation in rural India.

5.2. Policy Implications of our results

The policy implications of our results are threefold and are expected to be complementary to each other. First, improvement of the access to information or information provision particularly about dirty cooking fuels may be considered as a feasible policy for implementation. Such policies are expected to generate better awareness about health hazards related to HAP particularly in rural areas and their implementation is likely to reduce the usage and/or likelihood of choice of dirty cooking fuels. Second, the efficiency of the aforementioned information access and/or provision policy that results in awareness generation may be increased by addressing the target group identified by observables. For example, some individuals holding particular type of traits may perceive a low health risks related to HAP thus, have a higher tendency to use dirty cooking fuels. Any environmental policy in general and awareness generation about HAP in particular, may be found to be more effective if such an attitude can be targeted and altered. Finally, provision of education particularly to the female members is likely to be another long-term policy prescription that will help to reduce the issue

[1] In the fourth chapter of the dissertation, we have used data from the second (2017-2018) round of the survey where we have not collected information regarding the access to internet. As a surrogate of access to information disseminated through internet, we use the variable 'ownership of television' in the study.

[2] However, we may not concretely conclude that access to information has no role on individuals' preference because access to information can indirectly affect the WTP through affecting risk perception.

of HAP. Since the effect of information provision may vary across levels of education, individuals with lower levels of education may not utilize the information effectively even if information is provided or the access to information is improved. Therefore, another alternative policy may be designed where the households having low levels of education but access to information for better awareness generation is targeted first such that the HAP issue is gradually mitigated.

It is interesting to note that our proposed policies particularly the improving internet access and education provision, are in line with the current policies undertaken by the Government of India to reduce information gap as well as, improve female literacy. Among the various policies implemented by the Government of India, two major flagship programs – provision of internet access to all individuals as well as promotion of female literacy, have the most relevance to our results and related policy propositions. First, the Digital India initiative implemented by the Ministry of Electronics and Information Technology (MeitY) under Government of India since July 2015, aims to provide high-speed internet services to all citizens particularly in rural areas and improve digital literacy (MeitY, 2019). As an initial budget, the Government of India has invested INR 25 billion (USD 340 million) for this program (MeitY, 2019). This program is expected to establish and expand core ICT infrastructure that not only will increase the scope of access to information but also improve delivery of services and promote skill development for individuals as well. Second, in order to reduce female foeticide and improve female literacy in India, Ministry of Women and Child Development in association with Ministry of Education and Ministry of Health and Family Welfare, has launched the *Beti Bachao, Beti Padhao* (Save the girl child, Educate the girl child) program since January 2015. With an initial investment of INR 1 billion (USD 13 million), this scheme has a threefold objective – prevent female foeticide, ensure survival, and protection and education of the girl child (Ministry of Women and Child Development, 2019). In the context of HAP reduction, these two policies undertaken by the government get support from our findings in this dissertation. These two programs jointly are expected to reduce the information access gap related to health hazards from HAP as well as, make the beneficiaries more aware about the potential health risks from HAP.

However, based on our findings, we would like to propose few suggestions to improve the efficiency of the existing programs particularly in the context of HAP reduction in rural India. It is often seen that these two aforementioned programs are being implemented disjointly and hence, there is an efficiency loss. For maximum efficiency, provision of education and improving internet access should be done simultaneously. This is because, any information provided may not be utilised effectively by an uneducated or poorly educated individuals. Therefore, on the basis of our findings, we may suggest that these two programs should work hand-in-hand for generating public awareness against the health hazards associated with HAP. Generation of public awareness particularly in the context of interventions targeted to HAP reduction is of much relevance for ensuring sustained success of such programs. Another plausible suggestion to increase the efficiency of both these programs in generating awareness is by identifying the individuals based on observables who have the most conservative attitude regarding HAP as target households to whom the interventions would be addressed first. Although this may involve additional monetary and time cost, the efficacy of the programs is expected to be significantly enhanced.

5.3. Direction of Future Research

In this work, we have attempted to explore the role of various socio-economic, demographic and belief related attributes on individuals' behaviour to reduce HAP. However, there still remain few unobserved factors such as innate ability and women empowerment. People with higher levels of innate ability may be better at assessing health risks related to

HAP, and hence, they are more likely to use clean cooking fuels or hold a higher valuation for reduced health risks from HAP. Similarly, female empowerment may be related to the types of cooking fuels adopted at home. In this analysis, we have tried to incorporate few indicators of women empowerment such as whether women hold the decision-making authority in the households and economic independence of women. However, several other indicators of women empowerment such as educational freedom, economic contribution, perceived status within the household and health may still remain unobserved. Similarly, innate ability of individuals in general and female members in particular, may be measured through several physical and psychological instruments. This is not attempted in this study owing to various social and logistic constraints. Since females are one of the worst sufferers of HAP, we wish to extend our analysis in this direction by exploring the gender role in reducing HAP issues by exploring the role of these aforementioned variables. This may demand the involvement of additional survey instruments as well as psychological tests which we wish to take up as our future endeavour.

One of the major contributions of this study in the literature of fuel choice is to explore the role of individuals' subjective probabilistic expectations of HAP-related health risks on their cooking fuel usage pattern. Although we have discussed the possible sources of endogeneity of the individuals' subjective expectations of HAP-related health risks and the directions of the resulting bias, we wish to extend our study by identifying the key instrument(s) for such variables. One candidate in this regard may be developing a scale of measurement of individuals' concern about their health status in general. For example, individuals having higher scores for their concerns about health in general are likely to assign a higher (lower) value to the subjective probabilistic expectation of becoming sick from dirty (clean) fuel usage thus satisfying instrument relevance condition. Concurrently, since the reference statements to develop this scale will be determined by the researcher, it is likely to be unaffected by both observable and unobservable individual-specific covariates thus satisfying the instrument exogeneity condition. To develop such scale, we need to be careful in choosing the statements in the context of attitude-behaviour relationship or social-psychological theory of attitude formation with respect to individuals health and wellbeing. This may demand extensive survey where the subjective probabilistic expectations are elicited through interactive elicitation method using visual aids and also, eliciting response from such a scale. We wish to extend our research in this direction in future.

Despite the wide application of individuals' probabilistic expectations in other fields of economics such as behavioural economics and health economics, its application in environmental economics is quite rare. As our future research avenue, we wish to extend the application of individuals probabilistic expectations related to environmental events on their environmental-related behaviour. Given the impending crisis of climate change, it is important to analyse the role of individuals' perception of risks related to climate change reflected through their apprehension of occurrence of the event. In particular, we would like to explore the role of individuals' subjective expectations of risks related to climate change (e.g., heat stress, flood resulting from rising water levels) on their behaviour to adopt related preventive measures. To explore such relationship, we need to carefully choose our study area in the regions that are already affected or likely to be affected in near future by climate change. This exercise is expected to enable the policymakers to understand the individuals risk perception about climate change and thus design suitable policies.

As discussed earlier, understanding public attitudes as reflected through their perception is necessary for the successful implementation of policies. Individuals' perceptions measured through their probabilistic expectations may be developed from their past experiences, available information, cultural and economic beliefs, political inclination as well as, societal and environmental concern. A possible extension of this research may be to explore how these factors may directly influence individuals' perception which in turn, affects their attitude towards adoption of environmental policies related to climate change. For example, individuals with a high environmental and social consciousness as well as pro-governmental attitude is likely to prefer a governmental policy and hence is likely to support the same. On the other hand, individuals with either low environmental or social conscious or with anti-government attitude are likely to oppose such a policy. This line of research is relevant for both developed and developing countries although in different perspectives. For developed countries, the policy to be studied may include several policies to reduce carbon emission while for developing countries, the policies may involve those related to environmental pollutions. However, since developing countries are also now implementing similar sophisticated measures for reducing carbon emission, understanding public attitude through such a study may be a requirement. We have a plan to extend the research in this direction in future.

References:

Acemoglu, D., Johnson, S., and Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American economic review*, *91*(5), 1369-1401.

Alberini, A., and Krupnick, A. (2000). Cost-of-illness and willingness-to-pay estimates of the benefits of improved air quality: evidence from Taiwan. *Land Economics*, 37-53.

Alberini, A., Kanninen, B., and Carson, R. T. (1997). Modeling response incentive effects in dichotomous choice contingent valuation data. *Land economics*, 309-324.

Allendorf, K. (2013). Schemas of marital change: From arranged marriages to eloping for love. *Journal of Marriage and Family*,*75*(2), 453-469.

American Lung Association (2011). Toxic air: the case for cleaning up coal-fired power plants. *American Lung Association, Washington, DC*.

Andersson, H., Hole, A. R., and Svensson, M. (2016). Valuation of small and multiple health risks: A critical analysis of SP data applied to food and water safety. *Journal of Environmental Economics and Management*, *75*, 41-53.

Anglewicz, P., and Kohler, H. P. (2009). Overestimating HIV infection: The construction and accuracy of subjective probabilities of HIV infection in rural Malawi. *Demographic research*, *20*(6), 65.-96.

Angrist, J. D., and Pischke, J. S. (2008). *Mostly harmless econometrics*. Princeton university press.

Atreya, K., Johnsen, F. H., & Sitaula, B. K. (2012). Health and environmental costs of pesticide use in vegetable farming in Nepal. *Environment, Development and Sustainability*, *14*(4), 477-493.

Balakrishnan, K., Cohen, A. and Smith, K.R. (2014). Addressing the burden of disease attributable to air pollution in India: the need to integrate across household and ambient air pollution exposures. *Environmental Health Perspective*, *122*(1), A6-A7.

Balakrishnan, K., Dey, S., Gupta, T., Dhaliwal, R. S., Brauer, M., Cohen, A. J., ... & Dandona, L. (2019). The impact of air pollution on deaths, disease burden, and life expectancy across the states of India: the Global Burden of Disease Study 2017. *The Lancet Planetary Health*, *3*(1), e26-e39.

Balakrishnan, K., Ghosh, S., Ganguli, B., Sambandam, S., Bruce, N., Barnes, D. F., & Smith, K. R. (2013). State and national household concentrations of PM 2.5 from solid cookfuel use: results from measurements and modeling in India for estimation of the global burden of disease. *Environmental Health*, *12*(1), 1-14.

Baranov, V., Bennett, D., and 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.

Beaumais, O., Briand, A., Millock, K., & Nauges, C. (2014). What are households willing to pay for better tap water quality? A cross-country valuation study. FEEM Working Paper No. 24.2014

Becker, S. O., and Caliendo, M. (2007). Sensitivity analysis for average treatment effects. *The Stata Journal*, *7*(1), 71-83.

Bensch, G., and Peters, J. (2020). One-Off Subsidies and Long-Run Adoption—Experimental Evidence on Improved Cooking Stoves in Senegal. *American Journal of Agricultural Economics*, *102*(1), 72-90.

Bensch, G., Grimm, M., and Peters, J. (2015). Why do households forego high returns from technology adoption? Evidence from improved cooking stoves in Burkina Faso. *Journal of Economic Behavior & Organization*, *116*, 187-205.

Bonjour, S., Adair-Rohani, H., Wolf, J., Bruce, N. G., Mehta, S., Prüss-Ustün, A., ... & Smith, K. R. (2013). Solid fuel use for household cooking: country and regional estimates for 1980–2010. *Environmental health perspectives*, *121*(7), 784-790.

Brown, J., Hamoudi, A., Jeuland, M., and 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.

Buddhakulsomsiri, J., & Parthanadee, P. (2008). Stratified random sampling for estimating billing accuracy in health care systems. *Health Care Management Science*, *11*(1), 41-54.

Cameron, T. A., and Quiggin, J. (1994). Estimation using contingent valuation data from a" dichotomous choice with follow-up" questionnaire. *Journal of environmental economics and management*, *27*(3), 218-234.

Carson, R., Flores, N. E., and Hanemann, W. M. (1998). Sequencing and valuing public goods. *Journal of Environmental Economics and Management*, *36*(3), 314-323.

CHATTOPADHYAY, M., ARIMURA, T. H., KATAYAMA, H., SAKUDO, M., & YOKOO, H. F. (2017). Cooking Fuel Choices—Analysis of Socio-economic and Demographic Factors in Rural India—. *ENVIRONMENTAL SCIENCE*, *30*(2), 131-140.

Chattopadhyay, M., Arimura, T. H., Katayama, H., Sakudo, M., & Yokoo, H. F. (2021). Subjective probabilistic expectations, household air pollution, and health: Evidence from cooking fuel use patterns in West Bengal, India. *Resource and Energy Economics*, *66*, 101262.

Cincinelli, A. and Martellini, T. (2017). Indoor air quality and health. *International Journal of Environmental Research and Public Health, 14*(11), 1286.

Correa, T., Pavez, I., and Contreras, J. (2017). Beyond access: A relational and resource-based model of household Internet adoption in isolated communities. *Telecommunications Policy*, *41*(9), 757-768.

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.

Dasgupta, S. (2004). *Indoor air quality for poor families: new evidence from Bangladesh* (Vol. 3393). World Bank Publications.

Delavande, A. (2014). Probabilistic expectations in developing countries. *Annual Review of Economics, 6*(1), 1-20.

Delavande, A., and Kohler, H. P. (2009). Subjective expectations in the context of HIV/AIDS in Malawi. *Demographic research*, *20*, 817.-874.

Delavande, A., and Kohler, H. P. (2016). HIV/AIDS-related expectations and risky sexual behaviour in Malawi. *The Review of Economic Studies*, *83*(1), 118-164.

Dendup, N., and Arimura, T. H. (2019). Information leverage: The adoption of clean cooking fuel in Bhutan. *Energy policy*, *125*, 181-195.

Donfouet, H. P. P., Jeanty, P. W., and Mahieu, P. A. (2014). Dealing with internal inconsistency in double-bounded dichotomous choice: an application to community-based health insurance. *Empirical Economics*, *46*(1), 317-328.

Du, X., and Mendelsohn, R. O. (2011). Estimating the value of the reduction in air pollution during the Beijing Olympics. *Environment and Development Economics*, *16*(6), 735-749.

Duflo, E., Greenstone, M., & Hanna, R., 2008. Indoor air pollution, health and economic well-being. *SAPI EN. S. Surveys and Perspectives Integrating Environment and Society*, 1(1), 7-16

Epstein, R. M., & Peters, E. (2009). Beyond information: exploring patients' preferences. *Journal of the American Medical Association*, *302*(2), 195-197.

Farsi, M., Filippini, M. and Pachauri, S., 2007. Fuel choices in urban Indian households. *Environment and Development Economics*, *12*(6), 757-774.

Fisher, R. A. (1921). On the" Probable Error" of a Coefficient of Correlation Deduced from a Small Sample. *Metron*, 1, 3-32.

Franklin, P., Tan, M., Hemy, N., & Hall, G. L. (2019). Maternal exposure to indoor air pollution and birth outcomes. *International journal of environmental research and public health*, *16*(8), 1364.

Gelo, D., and Koch, S. F. (2015). Contingent valuation of community forestry programs in Ethiopia: Controlling for preference anomalies in double-bounded CVM. *Ecological Economics*, *114*, 79-89.

Godlonton, S., and Thornton, R. L. (2013). Learning from others' HIV testing: Updating beliefs and responding to risk. *American Economic Review*, *103*(3), 439-44.

Gould, C. F., & Urpelainen, J. (2018). LPG as a clean cooking fuel: Adoption, use, and impact in rural India. *Energy Policy*, *122*, 395-408.

Grieshop, A. P., Marshall, J. D., & Kandlikar, M. (2011). Health and climate benefits of cookstove replacement options. *Energy Policy*, *39*(12), 7530-7542.

Guo, S. (2015). Shaping social work science: What should quantitative researchers do?. *Research on Social Work Practice*, *25*(3), 370-381.

Guo, S., & Fraser, M. W. (2014). *Propensity score analysis: Statistical methods and applications*. SAGE publications.

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.

Hanemann, M., Loomis, J., and Kanninen, B. (1991). Statistical efficiency of double-bounded dichotomous choice contingent valuation. *American journal of agricultural economics*, *73*(4), 1255-1263.

Hanna, R., Duflo, E., and 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.

Heltberg, R. (2005). Factors determining household fuel choice in Guatemala. *Environment and development economics*, *10*(3), 337-361.

Herriges, J. A., and Shogren, J. F. (1996). Starting point bias in dichotomous choice valuation with follow-up questioning. *Journal of environmental economics and management*, *30*(1), 112-131.

Imbens G.W., Rubin D.B., (2015) Causal inference in statistics, social, and biomedical sciences. Cambridge University Press, USA.

Indian Oil Corporation. (2021). *Previous Price of Non-subsidised 14.2 Kg Indane Gas*. https://iocl.com/indane-14Kg-nonsubsid-previous-price (Accessed on August 31, 2021)

International Energy Agency. (2017). *Energy Access Outlook 2017: From Poverty to Prosperity*. https://www.oecd.org/publications/energy-access-outlook-2017-9789264285569-en.htm (accessed 04 February 2018).

Internet and Mobile Association of India (2015). Mobile Internet in India. Retrieved from https://cms.iamai.in/Content/ResearchPapers/5c0455c9-3c38-4d33-a6ab-3e834bee346d.pdf (accessed on April 14, 2020)

Jeanty, P. W., Haab, T. C., & Hitzhusen, F. J. (2007). *Willingness to pay for biodiesel in diesel engines: a stochastic double bounded contingent valuation survey* (No. 381-2016-22252).

Jeanty, P.W. (2007). *Constructing Krinsky and Robb Confidence Interval for Mean and Median WTP Using Stata*. http://repec.org/nasug2007/pwj_nasug07.pdf (accessed 16 June 2019).

Jeuland, M., Bhojvaid, V., Kar, A., Lewis, J., Patange, O., Pattanayak, S. K., ... and Ramanathan, V. (2014a). Preferences for improved cook stoves: Evidence from North Indian villages. *Duke Environmental and Energy Economics Working Paper Series No. EE*, 14-07.

Jeuland, M., Pattanayak, S. K., and Bluffstone, R. (2015). The economics of household air pollution. *Annual Review of Resource Economics*, *7*(1), 81-108.

Jeuland, M., Pattanayak, S. K., and Tan Soo, J. S. (2014b). Preference heterogeneity and adoption of environmental health improvements: Evidence from a cookstove promotion experiment. *Available at SSRN 2490530.*

Kabyanga, M., Balana, B. B., Mugisha, J., Walekhwa, P. N., Smith, J., & Glenk, K. (2018). Are smallholder farmers willing to pay for a flexible balloon biogas digester? Evidence from a case study in Uganda. *Energy for Sustainable Development*, *43*, 123-129.

Kar, A., Pachauri, S., Bailis, R., & Zerriffi, H. (2019). Using sales data to assess cooking gas adoption and the impact of India’s Ujjwala programme in rural Karnataka. *Nature Energy*, *4*(9), 806-814.

Khalil, U., and Mookerjee, S. (2019). Patrilocal residence and women's social status: Evidence from South Asia. *Economic Development and Cultural Change*, *67*(2), 401-438.

Khan, M. N., Islam, M. M., Islam, M. R., & Rahman, M. M. (2017). Household air pollution from cooking and risk of adverse health and birth outcomes in Bangladesh: a nationwide population-based study. *Environmental Health*, *16*(1), 1-8.

Kirk, D. S., and Sampson, R. J. (2013). Juvenile arrest and collateral educational damage in the transition to adulthood. *Sociology of education*, *86*(1), 36-62.

Krinsky, I., and Robb, A. L. (1986). On approximating the statistical properties of elasticities. *The Review of Economics and Statistics*, 715-719.

Langbein, J., Peters, J., & Vance, C. (2017). Outdoor cooking prevalence in developing countries and its implication for clean cooking policies. *Environmental Research Letters*, *12*(11), 115008.

Levine, D. I., Beltramo, T., Blalock, G., Cotterman, C., and Simons, A. M. (2018). What impedes efficient adoption of products? Evidence from randomized sales offers for fuel-efficient cookstoves in Uganda. *Journal of the European Economic Association*, *16*(6), 1850-1880.

Lewis, J. J., and Pattanayak, S. K. (2012). Who adopts improved fuels and cookstoves? A systematic review. *Environmental health perspectives*, *120*(5), 637-645.

Malla, S., and Timilsina, G. R. (2014). *Household cooking fuel choice and adoption of improved cookstoves in developing countries: a review*. The World Bank.

Manski, C. F. (2004). Measuring expectations. *Econometrica*, *72*(5), 1329-1376.

Mantel, N., and Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the national cancer institute*, *22*(4), 719-748.

McConnell, C., and Mupuwaliywa, M. (2016). *Keeping girls in school: Situation analysis for Malawi*. World Bank.

McFadden, D. (1975). *Economic applications of psychological choice models*. Institute of Transportation and Traffic Engineering, University of California.

McPhail, A. A. (1993). The "five percent rule" for improved water service: Can households afford more?. *World Development*, *21*(6), 963-973.

Ministry of Electronics and Information Technology (2019). India's Trillion-Dollar Digital Opportunity. https://www.meity.gov.in/writereaddata/files/india_trillion-dollar_digital_opportunity.pdf (accessed November 10, 2021)

Ministry of Panchayati Raj. (2020). Available via https://egramswaraj.gov.in/ (accessed on May 04, 2020)

Ministry of Petroleum and Natural Gas. (2016). *Pradhan Mantri Ujjwala Yojana*. http://petroleum.nic.in/sites/default/files/PMUY_SchemeE.pdf (accessed 12 March 2020).

Ministry of Women and Child Development. (2019). Beti Bachao Beti Padhao Scheme Implementation Guidelines. https://wcd.nic.in/sites/default/files/Guideline.pdf (accessed November 12, 2021)

Mobarak, A. M., Dwivedi, P., Bailis, R., Hildemann, L., and Miller, G. (2012). Low demand for nontraditional cookstove technologies. *Proceedings of the National Academy of Sciences*, *109*(27), 10815-10820.

Mortimer, K., Ndamala, C.B., Naunje, A.W., Malava, J., Katundu, C., Weston, W., Havens, D., Pope, D., Bruce, N.G., Nyirenda, M. and Wang, D. (2017). A cleaner burning biomass-fuelled cookstove intervention to prevent pneumonia in children under 5 years old in rural Malawi (the Cooking and Pneumonia Study): a cluster randomised controlled trial. *The Lancet, 389*(10065), 167-175.

Mugisha, J., Ajer, B., & Elepu, G. (2012). Contribution of Uganda cooperative alliance to farmers' adoption of improved agricultural technologies. *Journal of agriculture and social sciences*, *8*(1).

National Bank for Agriculture and Rural Development (2018). NABARD All India Rural Financial Inclusion Survey 2016-17. https://www.nabard.org/auth/writereaddata/tender/1608180417NABARD-Repo-16_Web_P.pdf (accessed October 30, 2021)

National Sample Survey Organisation. (2015). *Energy Sources of Indian Households for Cooking and Lighting, 2011-12* http://mospi.nic.in/sites/default/files/publication_reports/nss_report_567.pdf (accessed on January 10, 2016)

Olea, J. L. M., & Pflueger, C. (2013). A robust test for weak instruments. *Journal of Business & Economic Statistics*, *31*(3), 358-369.

Pachauri, S., and Jiang, L. (2008). The household energy transition in India and China. *Energy policy*, *36*(11), 4022-4035.

Papke, L. E., and 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.

Pattanayak, S. K., Jeuland, M., Lewis, J. J., Usmani, F., Brooks, N., Bhojvaid, V., ... and Ramanathan, V. (2019). Experimental evidence on promotion of electric and improved biomass cookstoves. *Proceedings of the National Academy of Sciences*, *116*(27), 13282-13287.

Paul, P. K., Bhuimali, A., and Raj, K. (2017). Digital Divide and Internet Saathi: A Social Computing and Informatics Practice Project for Empowering Indian Women. *International Journal of Applied Science and Engineering*, *5*(2), 55-63.

Perianayagam, A., and Srinivas, G. (2012). Explaining the skewed child sex ratio in rural India: revisiting the landholding-patriarchy hypothesis. *Economic and Political Weekly*, *47*(42), 85-94.

Plott, C. R., and Zeiler, K. (2005). The willingness to pay-willingness to accept gap, the" endowment effect," subject misconceptions, and experimental procedures for eliciting valuations. *American Economic Review*, *95*(3), 530-545.

Press Information Bureau (Government of India). (2019). *Nearly 1.04 crore LPG consumers voluntarily surrender their LPG subsidy under 'Give It Up' campaign.* https://pib.gov.in/Pressreleaseshare.aspx?PRID=1563847 (accessed on March 28, 2020)

Reserve Bank of India. (2012). Handbook of Statistics on Indian Economy. https://www.rbi.org.in/scripts/PublicationsView.aspx?id=16603 (accessed on March 11, 2020)

Ritchie, H., and Roser, M. (2020). *Indoor Air Pollution in Our World in Data.* https://ourworldindata.org/indoor-air-pollution. Accessed on 01 January 2020

Rohra, H., and Taneja, A. (2016). Indoor air quality scenario in India—an outline of household fuel combustion. *Atmospheric Environment*, *129*, 243-255.

Rosenbaum, P. R. (2002). Overt bias in observational studies. In *Observational studies* (pp. 71-104). Springer, New York, NY.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41-55.

Ross, S. M. (1996). Stochastic Processes. John Wiley & Sons. *New York*.

Scarpa, R., & Willis, K. (2010). Willingness-to-pay for renewable energy: Primary and discretionary choice of British households' for micro-generation technologies. *Energy Economics*, *32*(1), 129-136.

Shannon, A. K., Usmani, F., Pattanayak, S. K., and Jeuland, M. (2019). The price of purity: Willingness to pay for air and water purification technologies in Rajasthan, India. *Environmental and Resource Economics*, *73*(4), 1073-1100.

Singh, S., & Bhandari, M. (2012). Money management and control in the Indian joint family across generations. *The Sociological Review*, *60*(1), 46-67.

Smith KR and Pillarisetti A. (2017). Household Air Pollution from Solid Cookfuels and Its Effects on Health. In: Mock CN, Nugent R, Kobusingye O, Smith KR (Eds.), *Injury Prevention and Environmental Health*, vol. 7. World Bank: Washington, DC, pp. 133-152.

Stabridis, O., and van Gameren, E. (2018). Exposure to firewood: Consequences for health and labor force participation in Mexico. *World development*, *107*, 382-395.

Sun, C., and Zhu, X. (2014). Evaluating the public perceptions of nuclear power in China: Evidence from a contingent valuation survey. *Energy Policy*, *69*, 397-405.

Sun, C., Yuan, X., and Yao, X. (2016). Social acceptance towards the air pollution in China: evidence from public's willingness to pay for smog mitigation. *Energy Policy*, *92*, 313-324.

Takama, T., Tsephel, S., and Johnson, F. X. (2012). Evaluating the relative strength of product-specific factors in fuel switching and stove choice decisions in Ethiopia. A discrete choice model of household preferences for clean cooking alternatives. *Energy Economics*, *34*(6), 1763-1773.

Terza, J. V., Basu, A., and Rathouz, P. J. (2008). Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. *Journal of health economics*, *27*(3), 531-543.

Verma, A.P. and Imelda, I. (2019). *Clean Energy Access: Gender Disparity, Health, and Labor Supply*. https://e-archivo.uc3m.es/handle/10016/29397 (accessed 04 August 2020)

Wang, H., Xie, J., & Li, H. (2010). Water pricing with household surveys: A study of acceptability and willingness to pay in Chongqing, China. *China Economic Review*, *21*(1), 136-149.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

World Bank, (2019). *GDP growth (annual %)- India*. https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG?locations=IN (accessed 26 May 2020)

World Health Organisation (2018). *Opportunities for Transition to Clean Household Energy*. https://www.who.int/airpollution/publications/transition-to-clean-household-energy-Ethiopia/en/ (accessed 10 March 2020).

World Health Organization. (2007). *Indoor Air Pollution: National Burden of Disease Estimates*. https://www.who.int/airpollution/publications/nationalburden/en/ (accessed 29 July 2017).

Yokoo, H. F., Arimura, T. H., Chattopadhyay, M., & Katayama, H. (2020). *Subjective risk belief function in the field: Evidence from cooking fuel choices and health in India* (No. 2003). Research Institute for Environmental Economics and Management, Waseda University.(under review in *Journal of Development Economics)*

Zahno, M., Michaelowa, K., Dasgupta, P., and Sachdeva, I. (2020). Health awareness and the transition towards clean cooking fuels: Evidence from Rajasthan. *PloS one*, *15*(4), e0231931.

Zhang, J., and Smith, K. R. (2007). Household air pollution from coal and biomass fuels in China: measurements, health impacts, and interventions. *Environmental health perspectives*, *115*(6), 848-85