

Rural Household Cooking Fuel Choices: A Multidimensional Approach

Indian Journal of Human Development
14(3) 493–506, 2020
© 2020 Institute for
Human Development
Reprints and permissions:
in.sagepub.com/journals-permissions-india
DOI: 10.1177/0973703020975046
journals.sagepub.com/home/jhd



Chander Shekhar Mehra¹ and Niti Saxena²

Abstract

The multidimensional nature of poverty seeps into all other aspects of human development of which one is access to clean energy. Literature highlights the high correlation between human development and unequal access to energy.¹ The term “fuel poverty” is of importance here especially in a developing country context like India where access to clean fuel is a far sight for many despite key governmental initiatives promoting and subsidising access to clean fuel. The present article examines the impact of poverty on the likelihood of a household to shift to clean fuel for cooking, heating, etc. The present article shows that the usual method of estimating poverty is not sufficient enough to explain behavioural change among rural households. In fact, the simple provision of subsidy cannot ensure the complete shift of households to clean cooking fuel option. For this, a multidimensional poverty index concept is more appropriate. The article shows that multidimensional poverty act as a binding constraint in the shift of household from polluting cooking fuel to clean fuel option.

Keywords

Multidimensional poverty, clean fuel, Ujjwala Yojana, multiple imputation, multinomial logit regression, sustainable development goals

Introduction

Indoor air pollution from traditional biomass burning contributes to serious health problems, particularly cancer and respiratory infections (Bentayeb & Simoni, 2013; Chakraborty et al., 2014; Kankaria, et al., 2014). The time required for biomass collection can preclude formal employment outside the household for women (Burke & Dundas, 2015; World Bank and United Nations Development Programme, 2004), and the cost of purchasing biomass can weigh heavily on household budgets where formal biomass markets exist. Moreover, a growing body of literature suggests that incomplete combustion products and black carbon from traditional biomass burning make a significant contribution to climate change

¹ S M Sehgal Foundation, Gurugram, Haryana, India.

² Development Research and Extension, S M Sehgal Foundation, Gurugram, Haryana, India.

Corresponding author:

Chander Shekhar Mehra, S M Sehgal Foundation, Plot No. 34, Sector 44, Institutional Area, Gurugram 122003, Haryana, India.
E-mail: cs.mehra@smsfoundation.org

(Chowdhury et al., 2019; Goldemberg et al., 2018). Smoke inhaled by women from unclean fuel is as harmful as smoking cigarettes. In addition to this, women and children have to go through the drudgery of collecting wood.

In this context, the present article examines the impact of poverty on the likelihood of a household to shift to clean fuel for cooking, heating, etc. The traditional view is that income support or financial subsidy is enough to ensure that people change their behaviour. It is with this perception that Government of India usually devises policies such as the Ujjwala Yojana under which below poverty line (BPL) families are provided with the financial support of INR 1,600 per LPG connection. Since it is usually the female member of the household who does the cooking, the connection is issued in the name of the woman of the household under this scheme. Consumers are also provided interest-free loans for the purchase of an LPG gas stove and LPG refill from oil marketing companies. The determination of whether a household is considered poor is based on the poverty ratio, which is determined by the poverty line.

The scheme is expected to ensure that households in rural areas will take advantage and shift to clean fuel LPG for cooking and other household chores. To assess the impact of the scheme, a household survey was conducted in Samastipur district of Bihar, India, which covered 605 households. The study was financially supported by Mondo Power Ltd. and was conducted by S M Sehgal Foundation. The survey focused on perceptions and preferences for cooking fuel, cooking methodologies, and issues related to cooking. The present article shows that the usual method of estimating poverty is not sufficient enough to explain behavioural change among rural households. For this, a multi-dimensional poverty index (MPI) concept is more appropriate.

Description of the Sample

The study focuses on the perception and preferences of rural households to choose cooking fuel options and the associated constraints. The proposed study also delves into the challenges and opportunities available with the local entrepreneurs working to provide clean cooking options to households in Samastipur district of Bihar.

The sample households covered for the household survey belong to the rural areas of Samastipur district. The survey covered both the households having LPG connection as well as those without LPG connection. For conducting the survey, the field surveyors got in touch with sarpanch/ward member to get an idea of the caste composition of the village as well as the location of different caste and religion tolas within the village. The idea was to ensure as equal representation as possible to all the segments of the society. Both LPG and non-LPG households were covered so that appropriate comparison could be made to understand the factors underlying the choice of cooking fuel by the households.

As per the Census 2011, Office of the Registrar General of India there are 1,129 inhabited villages in the Samastipur district of Bihar. The total number of households in the rural areas of the district is 807,402.

The formula for determining the minimum sample of households is the following:

$$n = \frac{Z_a^2 [p(1-p)]N}{Z_a^2 [p(1-p)] + (N-1)ME_p^2}$$

where

Z_a = Z-score for various levels of confidence

ME_p = margin of error in terms of proportions

N = total population

n = sample size

p = the sample proportion, that is, the expected result from each question in the survey. In Bihar, the LPG cover is about 70 per cent so the value of p is taken as 0.7.

For the present case, we have the following values:

Z_a = for 95 per cent confidence level, the Z -score is 1.96

ME_p = ± 5 per cent or ± 0.05

N = 8,07,402

n = sample size

p = 0.7

$$n = \frac{(1.96)^2 [0.7(1 - 0.7)] 807402}{(1.96)^2 [0.7(1 - 0.7)] + (807402 - 1)(0.05)^2}$$

$$n = \frac{(1.96)^2 (0.21) 807402}{(1.96)^2 (0.21) + (807401)(0.05)^2}$$

$$n = 323$$

For the 807,402 households, the minimum sample size with 95 per cent level of confidence and 5 per cent margin of error is 323. However, for the sake of gaining an in-depth understanding of the issues at hand, it has been decided to cover 600 households.

A mix of multi-stage stratified random sampling and probability proportional to size sampling were used. For the probability proportional to size (PPS) sampling, the population in the villages as the size measure was used. Village population data from Census 2011, Office of the Registrar General of India,

Table I. Demographic Profile of Sample Villages as per the Census 2011, Govt. of India

Sub-Division	Blocks	Sample Villages/Town	Total HH	Total Population
Samastipur	Warishnagar	Sari	696	3,505
		Chak Milki	112	576
		Moglani Chak	566	2,654
		Sadipur	783	3,696
		Kishunpur	1,257	5,962
	Samastipur	Punas	1,106	5,291
		Raghunadhpur Bela	621	2,784
		Musapur	1,204	6,350
		Bejha Dih	830	3,905
		Hakimabad	885	5,110
Rosera	Rosera	Thahar Basarhia	2,024	9,217
		Jahangirpur	4,109	19,898
		Harpur	1,159	5,683
		Mohiuddi Nagar	4,446	21,157
		Chak That	1,812	9,259
	Bithan	Parkauli	150	1,024
		Lad Kapasia	581	2,895
		Ujan	1,158	6,229
		Bardauni	612	2,888
		Jagmohra	1,137	6,031

Source: Census 2011, Government of India.

Table 2. Block Wise Households Covered

Block	Village	Number of HH Covered
Samastipur	Punas	30
	Raghunadhpur Bela	30
	Mushapur	30
	Bejha Dih	30
	Hakimabad	30
Warishnagar	Sari	32
	Chak Milki	30
	Mogiani Chak	30
	Sadipur	30
	Kishunpur	30
Rosera	Thahar Basarhia	31
	Jahangirpur	30
	Harpur	30
	Mohiuddi Nagar	30
	Chak That	30
Bithan	Parkauli	31
	Lad Kapasia	30
	Ujan	31
	Bardauni	30
	Jagmohra	30
Total		605

Source: Household survey.

Government of India, was used for the PPS sampling. The district of Samastipur is divided into four administrative subdivisions by the government: Samastipur, Rosera, Dalsingsarai and Patori. At the first stage, two subdivisions were randomly selected. At the second stage, two blocks from each of the two subdivisions were randomly selected.

At the third stage, five villages from each of the selected blocks were selected using probability proportional to size (PPS) sampling. Table 1 shows the demographic profile of the sample villages in Samastipur as per the Census 2011.

At the fourth stage, 30 households were selected from each of the sampled villages using stratified random sampling. Stratification was done on the basis of scheduled caste (SC) households and non-SC households. Stratification using scheduled tribe (ST) has not been done as ST population constitutes only 0.04 per cent of the total population in the district. Census 2011 provides caste data for only SC and ST. A total of 605 households were covered under the quantitative household survey as shown in Table 2.

Multidimensional Poverty

The poverty ratio in rural areas of Bihar is estimated at 34.1 per cent (Finance Department, Government of Bihar, 2020). The estimation is based on the Tendulkar Committee methodology (Planning Commission, 2009), which is based on the consumption of items like cereal, pulses, milk, edible oil, non-vegetarian items, vegetables, fresh fruits, dry fruits, sugar, salt and spices, other food, intoxicants, fuel, clothing, footwear, education, medical (non-institutional and institutional), entertainment, personal and toiletry goods, other goods, other services, and durables. The committee concluded all India poverty line of INR 446.68 per capita per month in rural areas in 2004–2005.

However, such estimation suffers from the drawback that people just above the poverty line need not be better off than those below the line, similarly, poverty line measures fail to appreciate the fact that people further below the poverty line are poorer than people closer to the line (Ray, 1998; Todero & Smith, 2015). This calls for a different approach to understand the deprivations faced by the households.

To understand the true extent of deprivation faced by the households, an MPI was calculated. The MPI takes into account that there are negative interaction effects when people have multiple deprivations (Todero & Smith, 2015). This was done to better understand the failure of government policies and subsidies in ensuring behavioural change among the households to make a shift to clean fuel options or to improved cookstoves.

In the multidimensional poverty approach, a poor person is identified using the dual cut-off method by (a) the cut-off levels within each of the dimensions and (b), the cut-off of the number of dimensions in which a person must be deprived (below the line) to be deemed multi-dimensionally poor (Todero & Smith, 2015). For the present analysis, the MPI incorporates three dimensions at the household level: health, education, and standard of living.

With respect to health, two indicators have been used: whether any member of the household is suffering from any disability and whether any person in the household is suffering from any disease. The two indicators have been weighted equally. Each of the indicators counts one-sixth towards the maximum possible deprivation in the MPI.

Regarding education, two indicators: whether not even one adult member (18 years and above) has completed five years of education, and whether any children in the 6–17 age group are not studying. Again, each of the indicators counts one-sixth towards the maximum possible deprivation in the MPI.

In terms of standard of living, equal weight is placed on three deprivations (each accounting one-ninth towards the maximum possible): lack of electricity, inadequate flooring, and lack of more than one of five assets: mobile phone, television, radio, bicycle and motorcycle.

Individuals are then identified as multi-dimensionally poor if their household is deprived by a weighted sum of 0.3 or more.

As can be seen in Figure 1, two-thirds (67.4%) of the sample households are multi-dimensionally poor.

This is significantly higher than the poverty ratio estimated by the government of Bihar. The high level of multidimensional poverty means that: just the income subsidy provided by the government to make people shift to clean fuel will not be sufficient for such a shift.

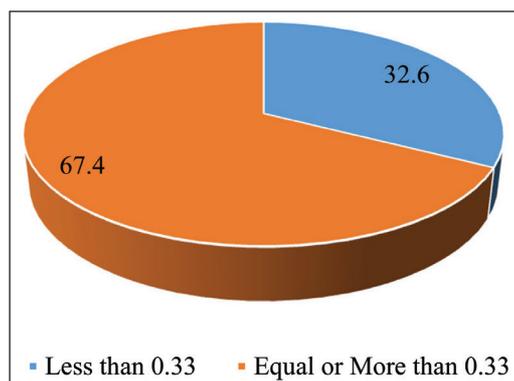


Figure 1. Distribution of HH According to MPI

Source: The authors.

Dealing with Missing Values

In the present case, we are facing the problem of missing values for a single independent variable (IV), which is continuous (*annual income last year*). The problem arises not only due to the inability of the households to recall their income last year but also due to the reluctance on the part of the households to share their actual annual income data. The present section presents the various methods that can be used for dealing with missing data, and also the final model that has been used in the present article for handling the missing observations in the *annual income last year* variable. Out of the total 605 households, data for the *annual income last year* variable was missing in 335 cases.

At the outset, it is assumed that data for the income variable that is missing is actually missing at random (MAR). The assumption means that, for example, if for a single variable W missing data exist, there is a vector of variables X that is always observed then, missing data on W may depend on the vector of variables X , but it does not depend on W itself after adjusting for X

$$\Pr(R_Z = 1 | X, W) = \Pr(R_Z = 1 | X) \quad (1)$$

In order to deal with the problem of missing values, the following methods are available.

Dropping the Variable

One can simply drop the variable with missing data. This is an appropriate option if the variable in question has little effect on the dependent variable. However, the very fact that one is considering an IV for inclusion shows that the IV in question is important and is likely to have an effect on the dependent variable. So for the present case, this option is not feasible.

Dropping Subjects

This option is feasible if missing data is limited to a small number of subjects. Under this, if a subject has missing data on any of the variables being used for the analysis, then that subject is dropped from the final analysis. However, the remaining cases may not be representative of the population. Moreover, this option is not feasible when data is missing on a large number of subjects.

Pairwise Deletion of Missing Data

Under this, a correlation matrix is constructed for each pair of IVs for those cases where values for both the variables exist. In other words, cases, where data is missing for either or both the variables, are excluded from the computation. However, this procedure is applied only when the data are randomly missing. Here again, the problem is that the results obtained cannot be said to be representative of the entire population. Moreover, this technique cannot be applied to logistic regression.

Substituting Value Along with Missing Data Indicator

This procedure involves two steps.

- Step 1:* Add some arbitrary value for all the missing data cases. The value is either 0 or mean of the variable.
- Step 2:* Create and add a dummy variable in the analysis that is coded 1 if in the first step arbitrary value has been plugged in for the missing data, and 0 otherwise.

While this approach ensures all the cases are included in the analysis, the method produces biased estimates of the coefficients (Allison, 2002).

Maximum Likelihood Method

The basic principle behind the maximum likelihood estimation is to choose as estimates those values which would, in the case of being true, ensure that the probability of observing what has in fact been observed is maximised (Allison, 2002). However, given that this method requires specifying a joint probability distribution for all the variables makes it complex. Moreover, this method is suitable for linear and log-linear regression models and not for models like logistic regression and Poisson regression when a large dataset is at hand, requiring the use of statistical software.

Multiple Imputation Method

Under this method, missing values are imputed using an appropriate model that incorporates random variation. Random imputation is desirable as it ensures that the biases that are endemic to deterministic imputation are eliminated. Moreover, multiple imputations also deal with the problem that if one uses imputed data as if it was real data, then the estimated standard errors will be too low, while test statistics will be too high. Under multiple imputations, multiple completed datasets are produced each with different, randomly drawn, imputed values, which tackle the standard error problem. Another advantage of multiple imputations over maximum likelihood method is that multiple imputations can be used with any type of model.

Proposed Imputation Model

For the present analysis, we have used multiple imputation method to deal with the missing data in the annual income variable. The share of observations with missing income data is around 55 per cent out of the total sample of 605 households. Still, multiple imputation can be used for generating estimates of the missing values even when the proportion of missing data is high (Nguyen, et al., 2017; Stuart et al., 2009). In the present case, predictive mean matching (PMM) technique has been used for imputing values for the annual income variable. PMM is a partially parametric method that matches the missing value to the observed value with the closest predicted mean (or linear prediction).

Under this method, at first linear regression has been used to generate estimates of the missing value for the annual income variable. The imputation model also contains the variables *whether the household uses clean fuel, MPI, and type of farmer by landholding*. This has been done because the analytic model that will be used for the analysis (multinomial logit regression) includes these variables. If these variables are not included in the imputation model, then the relationship between the annual income variable (for

which values are being imputed) and the excluded variables will be biased towards 0. There is no general specific rule for determining the sufficient number of imputations. This is because the actual number of imputations necessary for multiple imputations to perform satisfactorily depends not only on the amount of information missing due to non-response but also on the analysis model and the data. The number of imputations done in the present case: 100. This means that for every missing value of annual income variable, 100 non-missing replacements have been imputed. This has been done to make the results more robust.

Table 3. Result of Multiple Imputation Using Predictive Mean Matching

Univariate imputation	Imputations	100		
Predictive mean matching	Added	100		
Imputed: $m = 1$ through $m = 100$	Updated	0		
	Nearest neighbours	5		
Observations per m				
Variable	Complete	Incomplete	Imputed	Total
Annual income last year	270	335	335	605
(complete + incomplete = total; imputed is the minimum across m of the number of filled-in observations)				

Source: The authors.

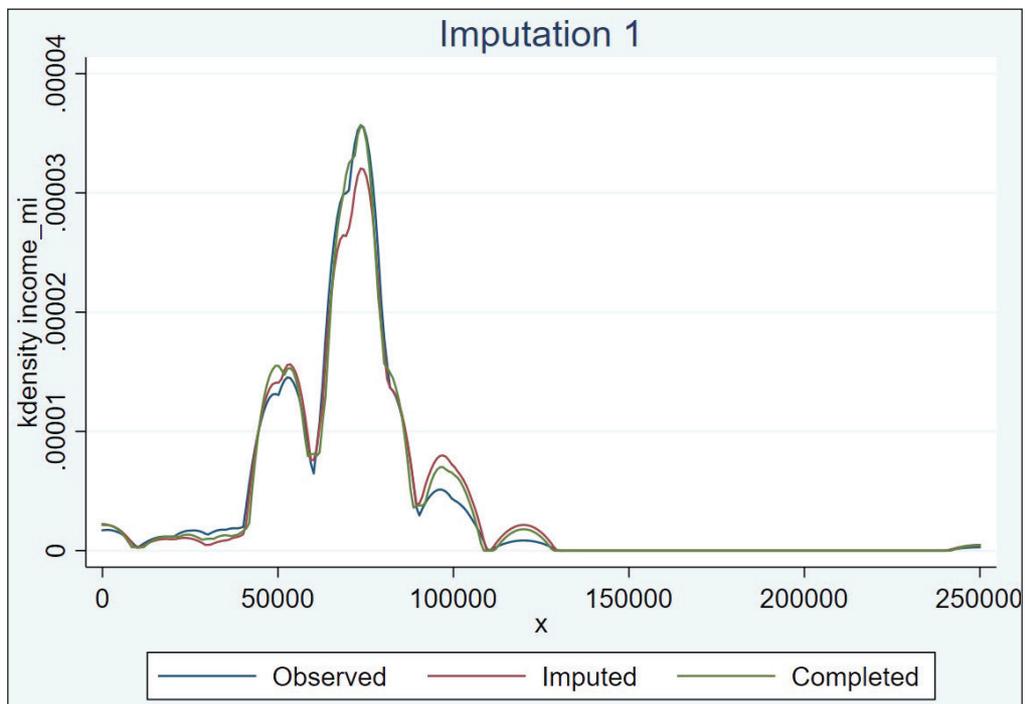


Figure 2. Diagnostic Plot for Annual Income

Source: The authors.

At first, the linear regression was used to generate estimates of the missing value for the annual income variable. After this, five neighbours having similar estimated values were identified. Note here that what we are referring to is the estimated value for the neighbour, and not the observed value of the neighbour. The observed value of the randomly chosen nearest neighbour was then used for imputed value for the case with missing data on the annual income variable. The result of the multiple imputation using PMM is presented in Table 3.

To check whether the imputation model is reasonable, one can use graphical and numerical checks that provide information about the distribution of imputed values. A diagnostic plot for the *annual income variable* is presented in Figure 2.

Figure 2 is a diagnostic plot for the annual income variable to check whether the imputation model used is appropriate or not. As the figure shows, the observed and imputed values are almost similar in distribution. This shows that the imputation model that was used is appropriate and has generated imputed data that can be considered as reasonable and usable for the analysis.

As can be seen from Figure 3, the imputed values lie within the bounds of the observed values. So the imputation model can be termed as appropriate for the analysis.

An alternative approach for checking the imputation model is to use a Kolmogorov–Smirnov (KS) test. The KS test is a non-parametric procedure for testing whether the two samples are from the same

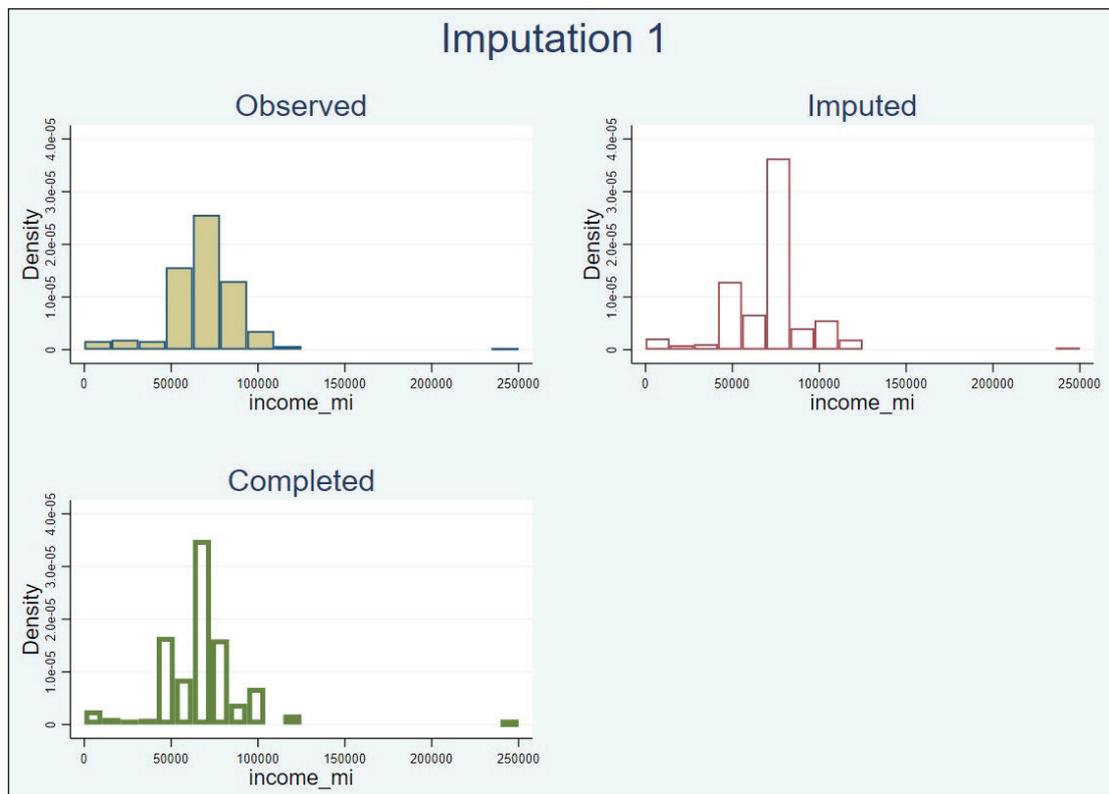


Figure 3. Histogram for Observed, Imputed and Completed Values for Annual Income

Source: The authors.

population (Nguyen et al., 2013). For each imputation model, separate KS tests were performed on each of the 100 imputed datasets, resulting in 100 p -values for each imputed variable. In none of the cases was the p -value less than .05, which shows that the distribution of the imputed values does not significantly differ from the observed data.

Multinomial Logit Regression Model

The multinomial logit regression (MNL) model is a discrete response model that falls under the category of unordered response. In other words, under this model a unit's response or choice depends on individual characteristics of the unit—but not on attributes of the choices (Wooldridge, 2010).

What makes the MNL model attractive is that it does not assume normality, linearity, or homoscedasticity. This feature of MNL is why it is used frequently as compared to discriminant analysis that requires the satisfaction of the mentioned assumptions.

We are considering an MNL model with three outcomes and three IVs. The outcome variable is *whether the household uses clean fuel*. The three possible outcomes are polluted fuel only ($status = 0$), polluted and clean fuel ($status = 1$) and clean fuel only ($status = 2$).

The mode for cooking fuel choice is

$$Prob(Y_i = J) = \frac{e^{\beta_j x_i}}{\sum_{k=0}^J e^{\beta_k x_i}} \quad J = 0, 1, 2$$

The estimated equations provide a set of probabilities for $J + 1$ choices for a decision-maker with characteristics x_i (Greene, 2012).

The explanatory variables are presented in Table 4.

The MNL model simultaneously estimates binary logits among all pairs of the outcome categories. In the present case, we have three categories: Category A = Polluted Fuel Only ($status=0$), category B = polluted and clean fuel ($status = 1$) and category C = clean fuel only ($status = 2$). The MNL is in effect simultaneously estimating three binary models which are the following:

$$\ln \left[\frac{\Pr(A | X)}{\Pr(C | X)} \right] = \beta_{0,A|C} + \beta_{1,A|C^{x1}} + \beta_{2,A|C^{x2}} + \beta_{3,A|C^{x3}}$$

$$\ln \left[\frac{\Pr(B | X)}{\Pr(C | X)} \right] = \beta_{0,B|C} + \beta_{1,B|C^{x1}} + \beta_{2,B|C^{x2}} + \beta_{3,B|C^{x3}}$$

$$\ln \left[\frac{\Pr(A | X)}{\Pr(B | X)} \right] = \beta_{0,A|B} + \beta_{1,A|B^{x1}} + \beta_{2,A|B^{x2}} + \beta_{3,A|B^{x3}}$$

Table 4. Description of the Variables used for Multinomial Logistic Regression

Variable Name	Description
Multi-dimensional poverty index (MPI)	Index (<i>higher value means higher deprivation</i>)
Type of farmer by landholding	Categorical variable (<i>landless = 1, marginal farmer = 2, small farmer = 3, medium farmer = 4, large farmer = 5</i>)
Annual income (<i>last year</i>)	Continuous variable

Source: The authors.

Because response probabilities must sum to unity, there is an implicit constraint on the three logits, which is

$$\ln \left[\frac{\Pr(A|X)}{\Pr(C|X)} \right] - \ln \left[\frac{\Pr(B|X)}{\Pr(C|X)} \right] = \ln \left[\frac{\Pr(A|X)}{\Pr(B|X)} \right]$$

In terms of parameters this means

$$\beta_{k,A|C} - \beta_{k,B|C} = \beta_{k,A|B}$$

The output of the MNL model is presented in Tables 5 and 6.

In the above, Average RVI is average relative variance increase due to non-response, which in the present case relates to annual income variable. It basically shows the average relative increase in the variance of the estimates due to the missing values in the annual income variable. The value of average RVI is averaged over all the coefficients of the model. The value of average RVI in the present case is 0.4899. The values of average RVI is zero when there is no missing data or when missing data has not been imputed.

The largest FMI is the largest fraction of missing information. The importance of this statistic lies in the fact that it gives an idea of whether the number of imputations done is sufficient for the analysis. A rule of thumb is that the number of imputations should be greater than or equal to $(100 \times \text{FMI})$. In the present case, the largest FMI is 0.6570, so the value obtained by multiplying it with 100 is 65.7. The number of imputations done in the present case (100) exceeds the required number of imputations (65.7). So, one can say that the number of imputations done in the present case are sufficient for the analysis.

Coming to the F test statistic. The F statistic with numerator 6 and 3910.6 denominator degrees of freedom is 4.52. The null hypothesis is that all the coefficients of the model are equal to zero. The estimated value of F statistic is larger than the table value, so we reject the null hypothesis. In fact, the significance level of the test is 0.1 per cent, so we reject the null hypothesis at 5 per cent level of significance.

Now coming to the interpretation of the coefficients, Table 6 contains the value of coefficients with *polluted fuel only* being the base category.

From the above, one can say that an increase in the value of MPI reduces the log-odds between having only clean fuel and having only polluted fuel by -3.463 . This shows that multidimensional poverty acts as a binding constraint on the ability of households to move towards clean fuel. Similarly, the log-odds between having polluted and clean fuel, and having only polluted fuel improve as the landholding size of households increases.

Table 5. Result of the Multinomial Logistic Regression

Multiple-imputation estimates	Imputations	100
Multinomial logistic regression	Number of observations	605
	Average RVI	0.4899
	Largest FMI	0.6570
DF adjustment: large sample	DF: min	231.45
	Avg.	3,66,987.20
	Max.	27,22,008.83
Model F test: equal FMI	$F(6, 3910.6)$	4.52
Within VCE type: OIM	Prob. > F	.0001

Source: The authors.

An interesting thing that is emerging from Table 6 is that the coefficient of income level is zero for both categories, despite the fact that income level is generally used for estimating poverty. On the other hand, the coefficient of MPI shows that a household's ability to move from polluting fuel to clean fuel is heavily dependent on the various type of deprivations that are excluded from the simple income method of estimating poverty.

The coefficients that are obtained after running the MNL regression are difficult to interpret as they are relative to the base outcome. Another way to evaluate the effect of covariates is to examine the marginal effect of changing their values on the probability of observing an outcome (Wooldridge, 2010). The average marginal effects have also been estimated. In Table 7, those marginal effects that are significant at 5 per cent level of significance have been shown, as only those matter. Marginal effects measure the change in the probability of households to shift to different categories of the dependent variable (other than the base outcome) due to the change in the IV.

From Table 7, we can see that the MPI and income have a significant impact on the probability to shift to clean fuel only, while the landholding of the household has a significant impact on the probability to shift to polluted and clean fuel.

The marginal effects show that on average, as MPI deprivation increases, the probability of households to shift completely to clean fuel declines by 32 percentage points. Similarly, as the land-holding size increases, the probability of households to shift from polluted fuel to a combination of polluted and clean fuel increases by 8.8 percentage points.

Table 6. Coefficients Obtained after Multinomial Logistic Regression

Whether the Household uses Clean Fuel	Coefficient
Polluted Fuel Only	(base outcome)
Polluted and Clean Fuel	
MPI	-0.207
Type of farmer by landholding	0.420*
Annual income (last year)	0.000
_cons	-0.414
Clean Fuel Only	
MPI	-3.463*
Type of farmer by landholding	0.261
Annual income (last year)	0.000*
_cons	-2.647

Source: The authors.

Note: *Significant at 0.05 level of significance.

Table 7. Average Marginal Effect for Variables with Significant Values

Variables	Average Marginal Effect
Multi-dimensional poverty index (MPI)	
Clean fuel only	-0.325*
Type of farmer by landholding	
Polluted and clean fuel	0.088*
Annual income (last year)	
Clean fuel only	0.000*

Source: The authors.

Note: *Significant at 0.05 level of significance.

An important thing that emerges is that the marginal effect value of annual income variable for the clean fuel only category of the dependent variable, though significant, is zero. For the polluted and clean fuel category, the marginal effect value of annual income variable is not significant.

The above result shows that income level, which is used for estimating poverty, will not help in changing the cooking fuel usage behaviour of individuals towards only clean fuel. If the government wants the people in rural areas to shift completely to clean fuel and completely shun the solid fuel options like firewood, then what is needed is a comprehensive economic policy that deals with the multiple deprivations to which individuals are exposed in rural areas. One must remember that in rural areas, firewood and chips are in most cases available free of the cost compared to clean fuel options like LPG. So just income subsidies are not sufficient to bring behavioural change in rural areas.

Conclusion and Way Forward

The above analysis has shown that subsidies provided by the government to promote the shift to cleaner fuels does not seem to be enough. This is because the traditional poverty ratio method fails to detect the true extent of deprivation. Under the multidimensional poverty method, two-thirds of the households were found to be multidimensional poor. So what is needed is an appreciation of the correct level of deprivation and to plan accordingly. The following measures can go a long way in promoting and creating a conducive environment in which behavioural change can take place that makes use of clean fuel as a new norm.

Direct Financial Assistance

Multidimensional poverty among two-thirds of the sampled households means that the subsidy provided by the government is not sufficient to make the households shift to clean fuel. What is needed is full financial assistance in which the LPG connection and first refill are provided free of cost to the households. Increased coverage of LPG will make the financial debt incurred by the government self-liquidating in the long-run due to increased revenue from LPG usage.

Do Away with Delivery Charges

During the survey, it was found that households were forced to pay delivery charges for LPG refills. The government should ensure that the LPG distributors are barred from charging delivery fees to the consumers. In case distributors face additional financial constraints as a result of servicing far-off places, the government should reimburse their delivery expenses instead of these distributors recovering the expenses from the consumers.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The authors received no financial support for the research, authorship and/or publication of this article.

Note

1. SDG 7 is ensuring access to affordable, reliable, sustainable and modern energy for all. <https://sustainabledevelopment.un.org/content/documents/17480PB8.pdf>

References

- Allison, P. (2002). *Monograph on missing data* (paper # 136). SAGE Publications.
- Bentayeb, M., & Simoni, M. (2013). Indoor air pollution and respiratory health in elderly. *Journal of Environmental Science and Health*, 48, 1783–1789.
- Burke, P. J., & Dundas, G. (2015). Female labor force participation and household dependence on biomass energy: Evidence from national longitudinal data. *World Development*, 67, 424–437.
- Chakraborty, D., Mondal, N. K., & Datta, J. K. (2014). Indoor pollution from solid biomass fuel and rural health damage: A micro-environmental study in rural area of Burdwan, West Bengal. *International Journal of Sustainable Built Environment*, 3(2), 262–271.
- Chowdhury, S., Dey, S., Guttikunda, S., Pillarisetti, & Smith, K. R. (2019). Indian annual ambient air quality standard is achievable by completely mitigating emissions from household sources. *Proceedings of the National Academy of Sciences*, 116(22), 10711–10716.
- Cohen, J., Cohen, P., West. S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Lawrence Erlbaum Associates.
- Finance Department. (2020). *Bihar economic survey 2019–20*. Government of Bihar. <http://finance.bih.nic.in/Reports/Economic-Survey-2020-EN.pdf>
- Goldemberg, J., Martinez-Gomez, J., Sagar, A., & Smith, K. R. (2018). Household air pollution, health, and climate change: Cleaning the air. *Environmental Research Letters*, 13, 030201.
- Greene, W. H. (2012). *Econometric analysis* (5th ed.). Pearson Education.
- Kankaria, A., Nongkynrih, B., & Gupta, S. K. (2014). Indoor air pollution in India: Implications on health and its control. *Indian Journal of Community Medicine: Official Publication of Indian Association of Preventive & Social Medicine*, 39(4), 203–207. <https://doi.org/10.4103/0970-0218.143019>
- Nguyen, C. D., Carlin, J. B., & Lee, K. J. (2013). Diagnosing problems with imputation models using the Kolmogorov-Smirnov test: A simulation study. *BMC Medical Research Methodology*, 13, 144.
- Nguyen, C. D., et al. (2017). Model checking in multiple imputation: An overview and case study. *Emerging Theme in Epidemiology*, 14, 8. <https://doi.org/10.1186/s12982-017-0062-6>
- Planning Commission. (2009). *Report of the expert group to review the methodology for estimation of poverty*. Government of India. http://www.indiaenvironmentportal.org.in/files/rep_pov.pdf
- Ray, D. (1998). *Development economics*. Princeton University Press.
- Stuart, E. A., Azur, M., Frangakis, C., & Leaf, P. (2009). Multiple imputation with large data sets: A case study of the children's mental health initiative. *American Journal of Epidemiology*, 169(9), 2009.
- The International Bank for Reconstruction and Development/The World Bank and United Nations Development Programme. (2004). *The impact of energy on women's lives in rural India*. <http://documents1.worldbank.org/curated/en/600391468771661997/pdf/ESM2760PAPER0Impact0of0Energy.pdf>
- Toderò, M. P., & Smith, S. C. (2015). *Economic development* (12th ed.). Pearson Education.
- Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). The MIT Press.